SCHOLARLY PUBLICATIONS

A CURRENT AWARENESS BULLETIN
OF RESEARCH OUTPUT

@DTU

(91st Edition)

JULY 2020

BY: CENTRAL LIBRARY

DELHI TECHNOLOGICAL UNIVERSITY
(FORMERLY *DELHI COLLEGE OF ENGINEERING)*
GOVT. OF N.C.T. OF DELHI
SHAHBAD DAULATPUR, MAIN BAWANA ROAD
DELHI 110042

PREFACE

This is the **Ninety first** Issue of Current Awareness Bulletin started by Delhi Technological University, Central Library. The aim of the bulletin is to compile, preserve and disseminate information published by the faculty, students and alumni for mutual benefits. The bulletin also aims to propagate the intellectual contribution of Delhi Technological University (DTU) as a whole to the academia.

The bulletin contains information resources available in the internet in the form of articles, reports, presentations published in international journals, websites, etc. by the faculty and students of DTU. The publications of faculty and student which are not covered in this bulletin may be because of the reason that the full text either was not accessible or could not be searched by the search engine used by the library for this purpose.

The learned faculty and students are requested to provide their uncovered publications to the library either through email or in CD, etc. to make the bulletin more comprehensive.

This issue contains the information published during **July, 2020**. The arrangement of the contents is alphabetical. The full text of the article which is either subscribed by the university or available in the web is provided in this bulletin.

Central Library

CONTENTS

1. A Comparative Study of Image Dehazing Algorithms, ***3.Anil Singh Parihar***, ***8.Yash Kumar Gupta***, ***8.Yash Singodia***, ***8.Vibhu Singh*** and ***6.Kavinder Singh***, Computer, DTU

2. A Detailed Study on Generative Adversarial Networks, ***3.Shailender Kumar*** and ***7.Sumit Dhawan***, Computer, DTU

3. Adaptive Droop based Control for Active Power Sharing in Autonomous Microgrid for Improved Transient Performance, Surya Prakash, ***8.1.Vaibhav Nougain***, and Sukumar Mishra, Electrical, DTU

4. An Embedding-based Deep Learning Approach for Movie Recommendation, ***3.Ram Murti Rawat***, ***8.Vikrant Tomar*** and ***8.Vinay Kumar***, Computer, DTU

5. Analysis of demand risks for the Indian automotive sector in globally competitive environment, ***3.Girish Kumar***, Rajesh Kr. Singh, ***8.1.Rishabh Jain***, ***8.Raman Kain*** and ***8.Naveen***, Mechanical, DTU

6. Analysis of Total Quality Management (TQM) Barriers on its Performance using TOC, ***6.Rakesh Kumar*** and ***3.R.S. Mishra***, Mechanical, DTU

7. Anomalous Object Tracking in Distributed Camera Network, ***8.MD Shahbaz Khan*** and ***3.Indu Sreedevi***, DTU

8. Arduino GSM based Power Theft Detection and Energy Metering System, ***8.Vivek Kumar Jaiswal***, ***6.Hritik Kumar Singh*** and ***3.Kuldeep Singh***, Electrical, DTU

9. Best criteria selection based PROMETHEE II method, ***6.Anjali Singh***, ***3.Anjana Gupta*** and Aparna Mehra, Mathematics, DTU

10. Breast cancer burden in Africa: evidence from GLOBOCAN 2018, ***3.Rajesh Sharma***, USME, DTU

11. Capsmf: a novel product recommender system using deep learning based text analysis model, ***3.Rahul Katarya[1]*** and ***7.1.Yamini Arora[2]***, Computer[1] and IT[2], DTU

12. Citation-Based Scientific Paper Summarization Using Game Theory, ***3.Anil Singh Parihar***, ***6.Anjali Jain*** and ***8.1.Archit Gupta***, Computer, DTU

13. Comparative Analysis of PWM Techniques and SVPWM operated V/f control of Induction Motor having different power ratings, ***8.Divyam***, ***8.1.Ayushi Saxena***, ***6.Bharat Singh*** and ***3.J. N. Rai***, Electrical, DTU

14. Comparative Analysis of Route Planning Algorithms on Road Networks, ***3.R. K. Yadav***, ***8.Giriraj Kishor***, ***8.Himanshu*** and ***8.Kishan Kashyap***, Computer, DTU

15. Comparative Study of Static VAR Compensation Techniques – Thyristor Switched Reactor and Thyristor Switched Capacitor, ***8.Divyam***, ***8.1.Ayushi Saxena***, ***6.Bharat Singh*** and ***3.J. N. Rai***, Electrical, DTU

16. Comparative study on air quality status in Indian and Chinese cities before and during the COVID-19 lockdown period, ***8.Aviral Agarwal***, Aman Kaushik, ***8.Sankalp Kumar*** and ***3.Rajeev Kumar Mishra***, Environmental, DTU

17. Control and Design of Seven-Level Cascaded Multilevel Converter for Transformerless Large-Scale Photovoltaic Integration, ***6.Jyoti Kulkarni***, ***3.Narendra Kumar*** and Bhim Singh, Electrical, DTU

18. COVID 19 Global Pandemic: Impact on Management of Supply Chain, ***6.Rakesh Kumar*** and ***3.R.S. Mishra***, Mechanical, DTU

19. Deep Learning Approach for Arbitrary Image Style Fusion and Transformation using SANET model, ***7.Pratibha Rathi***, ***7.Pranav Adarsh*** & ***3.Dr. Manoj Kumar***, Computer, DTU

20. Dementia Detection Using Machine Learning by Stacking Models, ***3.Ram Murti Rawat***, ***8.Mohammad Akram***, ***8.Mithil*** and ***8.Setti Suneel Pradeep***, Computer, DTU

21. Descriptive epidemiology of incidence and mortality of primary liver cancer in 185 countries: evidence from GLOBOCAN 2018, ***3.Rajesh Sharma***, USME, DTU

22. Design and analysis of a dual-polarization multiband oval ring patch antenna with L-stubs and folded meander line for C-band/X-band/ Kuband/ K-band communications, ***6.Karteek Viswanadha*** and ***3.N.S.Raghava***, Electronics, DTU

23. Detecting Intrusions and Attacks in the Network Traffic using Anomaly based Techniques, ***3.Vinod Kumar***, ***8.Vinay Choudhary***, ***8.Vivek Sahrawat*** and ***8.Vinay Kumar***, Computer, DTU

24. Developing a framework for evaluating sustainability index for logistics service providers: graph theory matrix approach, ***6.Anchal Gupta*** and Rajesh Kumar Singh, DTU

25. Disease Detection in Coffee Plants Using Convolutional Neural Network, ***3.Manoj Kumar***, ***8.Pranav Gupta***, ***8.Puneet Madhav*** and ***8.Sachin***, Computer, DTU

26. Dual-Modal Emotion Recognition Using Discriminant Correlation Analysis, ***3.Dr. Rajiv Kapoor***, ***8.Suhail Saifi***, ***8.Tarun Kapoor***, ***8.Utkarsh Bisaria*** and ***8.Nikhil Singh***, Electronics, DTU

27. Effect of Distributed Damping on the Dynamic Behavior of Flexible Chassis of Heavy Road Vehicle Under Standardized Random Road Responses, Ashish Gupta and ***3.Vikas Rastogi***, Mechanical, DTU

28. Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, ***6.Husain Mehdi*** and ***3.R.S. Mishra***, Mech, DTU

29. Effect of Mesh Size on Soil-Structure Interaction in Finite Element Analysis, ***7.Nasasira Derrick*** and ***3.Amit Kumar Srivastava***, Civil, DTU

30. Efficient strip to slot waveguide mode converter using fast quasi-adiabatic approach, Nikhil Dhingra, Geetika Jain Saxena, ***3.Kamal Kishor*** and Enakshi Khular Sharma, Physics, DTU

31. Energy Efficient Transmission Trends Towards Future Green Cognitive Radio Networks (5G): Progress, Taxonomy and Open Challenges, ***6.Akanksha Srivastava***, Mani Shekhar Gupta and Gurjit Kaur, Electronics, DTU

32. Ensemble learning-based approach for crack detection using CNN, ***8.Vibhu Kailkhura***, ***8.S. Aravindh***, ***8.Shashank Shekhar Jha*** and ***3.N. Jayanthi***, Electronics, DTU

33. Environment Monitoring System for Agricultural Application using IoT and Predicting Crop Yield using Various Data Mining Techniques, ***8.Gautam Gupta***, ***8.Rashi Setia***, ***8.Archana Meena*** and ***3.Bhavnesh Jaint***, Electrical, DTU

34. Estimation of COVID-19 Prevalence under Closed Population using Capture-Recapture approach, ***6.Rakesh Kumar*** and ***3.R.S.Mishra***, Mechanical, DTU

35. Experimental studies of vapour compression refrigeration system with eco-Friendly primary refrigerant and brine mixed with nano particles as secondary refrigerant, Uma Shankar Prasad, ***3.R.S. Mishra*** and R.K. Das, Mechanical, DTU

36. Experimental Validation of Fully Informed Particle Swarm Optimization Tuned Multi-Loop L-PID Controllers for Stabilization of Gantry Crane System, ***3.Sudarshan K Valluru***, ***8.Daksh Dobhal***, ***8.Kumar Kartikeya***, ***8.Manpreet Kaur***, ***3.Madhusudan Singh*** and ***8.Arnav Goel***, DTU

37. Filter Based Scalable Blockchain for Domestic Internet of Things, ***8.Aditya Shyam Bazari***, ***8.Aditya Singh***, ***8.Abdul Ahad Khan*** and ***3.Rajni Jindal***, Computer, DTU

38. FLEXIBLE SOLAR POWER TREE - A CONCEPT, Mukesh Kumar, ***3.J P Kesari*** and Lokesh Rana, Mechanical, DTU

39. Forecasting the Probability of Cost Overrun Risk of Indian Construction Projects using Fuzzy Model, Savita Sharma, ***3.Pradeep K. Goyal*** and R.C. Chhipa, Civil, DTU

40. Fuzzy Logic based Automatic Generation Control, ***8.MD Shadan Salik***, ***8.Shaurya Priyal***, ***8.1.Neha Khokher*** and ***3.DR. Narendra Kumar***, Electrical, DTU

41. Generating Visible Spectrum Images from Thermal Infrared using Conditional Generative Adversarial Networks, ***8.Neeraj Bhat***, ***8.Navneet Saggu***, ***8.Pragati*** and ***3.Sanjay Kumar***, Computer, DTU

42. GloVeInit at SemEval-2020 Task 1: Using GloVe Vector Initialization for Unsupervised Lexical Semantic Change Detection, ***8.Vaibhav Jain***, Computer and Mathematics, DTU

43. Gratitude, social intelligence, and leadership among university teachers: mediation and moderation analysis, ***3.Naval Garg***, Anuradha Jain and B.K. Punia, USME, DTU

44. Implementation Strategy for Resource Optimization of FPGA-Based Adaptive Finite Control Set-MPC using XSG for a VSI System, Vijay Kumar Singh, ***7.1.Ravi Nath Tripathi***, and Tsuyoshi Hanamoto, Electrical, DTU

45. Modelling and optimisation of magnetic abrasive finishing process based on a non-orthogonal array with ANN-GA approach, ***6.S. Ahmad*** , ***3.R. M. Singari*** and ***3.R.S. Mishra***, Mechanical, DTU

46. Modelling Logistic Growth Model for COVID-19 Pandemic in India, ***3.Minni Jain***, ***8.Prajwal Kumar Bhati***, ***8.Pranav Kataria*** and ***8.Rohit Kumar***, CSE, DTU

47. Morphological and Swelling Potential Evaluation of Moringa oleifera Gum/Poly(vinyl alcohol) Hydrogels as a Superabsorbent, Suhail Ahmad, Kaiser Manzoor, ***3.Roli Purwar*** and Saiqa Ikram, Chemistry, DTU

48. Multi-dimensional aggregation: a viable solution for interval data, ***3.Shailender Kumar***, Computer, DTU

49. Node Classification in Complex Networks using Network Embedding Techniques, ***3.Sanjay Kumar***, ***8.Sanidhya Chaudhary***, ***8.Saksham Kumar*** and ***8.Raj Kumar Yadav***, Computer, DTU

50. Nonlinear dynamics of a time-delayed epidemic model with two explicit aware classes, saturated incidences, and treatment, ***6.Kanica Goel***, Abhishek Kumar and ***3.Nilam***, Mathematics, DTU

51. Numerical heat transfer analysis & predicting thermal performance of fins for a novel heat exchanger using machine learning, ***7.Gaurav Krishnayatra***, ***7.Sulekh Tokas*** and ***3.Rajesh Kumar***, Mechanical, DTU

52. On Black–Scholes option pricing model with stochastic volatility: an information theoretic approach, ***6.Luckshay Batra*** and ***3.H. C. Taneja***, Mathematics, DTU

53. ON CERTAIN GENERALIZATIONS OF S, ***3.S. SIVAPRASAD KUMAR*** AND ***6.KAMALJEET GANGANIA***, Mathematics, DTU

54. On the Outage Probability and BER of a DF-Based Multi-Hop System Over and □ □ -Extreme Fading Channels, Tau Raphael Rasethuntsa, ***8.Manpreet Kaur*** and Sandeep Kumar, Computer, DTU

55. Performance analysis of beamforming aided uncorrelated mm-Wave MIMO system for IEEE 802.15.3c standard, Anand Agrawal, Vimal Kumar Mishra, Ashish Gupta and ***6.1.Bajrang Bansal***, Electronics, DTU

56. Performance of Masonry Infill Framed Structures During Earthquake: A State-Of-The-Art Review, Priyanka, ***3.Alok Verma*** and Shobha Ram, Civil, DTU

57. PMS-Sharing : Framework for Automatically Authenticating users in a Group to Allow Sharing Storage, Nitesh Suthar and ***3.Rajesh Kr. Yadav***, Computer, DTU

58. Predicting Link Sign in Online Social Networks based on Social Psychology Theory and PageRank, ***3.Sanjay Kumar***, ***8.Sudhanshu Shekhar Singh*** and ***8.Vipul Gupta***, Computer, DTU

59. Real Time Experimental control and Design of FOP□□ and L-PID Controllers tuned by Invasiveweed optimization for Trajectory Control of TRAS, ***7.1.Hem Prabha*** and ***8.Rajul Kumar***, Electrical, DTU

60. Realization of Regula-Falsi Iteration based Double Precision Floating Point Division, ***8.Riya Jain*** and ***3.Neeta Pandey***, Electronics, DTU

61. Real-Time Implementation of the Optimal Predictor and Optimal Filter: Accuracy Versus Latency, SYED ASEEM UL ISLAM, ***8.1.ANKIT GOEL*** and DENNIS S. BERNSTEIN, Mechanical, DTU

62. Selection of sustainable transport system: a case study, ***8.Abhishek Singh***, Amulya Gurtu and Rajesh Kumar Singh, Civil and Environment, DTU

63. SIX SIGMA APPLICATION IN KETCHUP INDUSTRY FOR DAMAGED GLASS BOTTLES: A CASE STUDY, ***8.1.YATHARTH SHARMA***, ***8.SARMAD HASSAN*** and ***8.RAJAT KUMAR***, Mechanical, DTU

64. Sodium hydroxide effect on the mechanical properties of flyash-slag based geopolymer concrete, ***6.Manvendra Verma*** and ***3.Nirendra Dev***, Civil, DTU

65. Spatial Landslide Risk Assessment at Phuentsholing, Bhutan, Abhirup Dikshit, ***3.Raju Sarkar***, Biswajeet Pradhan, Saroj Acharya and Abdullah M. Alamri, Civil, DTU

66. SUBORDINATION AND RADIUS PROBLEMS FOR CERTAIN STARLIKE FUNCTIONS, ***3.S. SIVAPRASAD KUMAR*** AND ***6.KAMALJEET GANGANIA***, Mathematics, DTU

67. Suppression of Lower Order Harmonics for the full modulation range for a 2-level inverter fed IM drive with a switched capacitive filter technique forming a 42-sided Voltage Space Vector Structure, ***7.1.Rahul Dewani***, Rakesh R, K. Gopakumar, Umanand Loganathan Dariusz Zielinski and Leopoldo G. Franquelo, Electronics, DTU

68. The prospective application of a graphene/MoS2 heterostructure in Si-HIT solar cells for higher efficiency, Chandra Kamal Borah, ***3.Pawan K. Tyagi*** and Sanjeev Kumar, Physics, DTU

69. Theoretical analysis for transmission of Gaussian and sine time irradiance of electromagnetic beam in collisional dusty plasmas, ***6.Ruchi Sharma*** and ***3.Suresh C. Sharma***, Applied Physics, DTU

70. Theoretical Framework of Collaboration between MSME and Private Engineering Colleges to Co-exist during COVID-19, ***3.Seema Singh***, Humanities, DTU

71. Towards Social & Engaging Peer Learning: Predicting Backchanneling and Disengagement in Children, ***8.Mononito Goswami***, ***8.Minkush Manuja*** and ***8.Maitree Leekha***, Computer, DTU

72. Tribological characterisation of thermal sprayed CrC alloyed coating – A review, ***6.Navendu Gupta***, Shailesh Kumar Singh and Shailesh Mani Pandey, Mechanical, DTU

73. Ubiquitin biology in neurodegenerative disorders: From impairment to therapeutic strategies, ***6.Dhiraj Kumar***, ***6.Rashmi K. Ambasta*** and ***3.Pravir Kumar***, Biotechnological, DTU

74. Women Empowerment in India: A Critical Analysis, ***3.Seema Singh*** and ***6.Antra Singh***, Economics, DTU

1. *Chancellor*

2. *Pro Vice Chancellor*	***2.1.*** *Ex Pro Vice Chancellor*
3. *Faculty*	***3.1.*** *Ex Faculty*
4. *Teaching-cum-Research Fellow*	***4.1.****Asst. Librarian*
5. *Alumni*	
6. *Research Scholar*	***6.1.*** *Ex Research Scholar*
7. *PG Scholar*	***7.1.*** *Ex PG Scholar*
8. *Undergraduate Student*	***8.1.*** *Ex Undergraduate Student*

SCHOLARLY PUBLICATIONS

A CURRENT AWARENESS BULLETIN
OF RESEARCH OUTPUT

@DTU

(91st Edition)

JULY 2020

BY: CENTRAL LIBRARY

DELHI TECHNOLOGICAL UNIVERSITY
(FORMERLY *DELHI COLLEGE OF ENGINEERING)*
GOVT. OF N.C.T. OF DELHI
SHAHBAD DAULATPUR, MAIN BAWANA ROAD
DELHI 110042

PREFACE

This is the **Ninety first** Issue of Current Awareness Bulletin started by Delhi Technological University, Central Library. The aim of the bulletin is to compile, preserve and disseminate information published by the faculty, students and alumni for mutual benefits. The bulletin also aims to propagate the intellectual contribution of Delhi Technological University (DTU) as a whole to the academia.

The bulletin contains information resources available in the internet in the form of articles, reports, presentations published in international journals, websites, etc. by the faculty and students of DTU. The publications of faculty and student which are not covered in this bulletin may be because of the reason that the full text either was not accessible or could not be searched by the search engine used by the library for this purpose.

The learned faculty and students are requested to provide their uncovered publications to the library either through email or in CD, etc. to make the bulletin more comprehensive.

This issue contains the information published during **July, 2020**. The arrangement of the contents is alphabetical. The full text of the article which is either subscribed by the university or available in the web is provided in this bulletin.

Central Library

CONTENTS

1. A Comparative Study of Image Dehazing Algorithms, ***3.Anil Singh Parihar***, ***8.Yash Kumar Gupta***, ***8.Yash Singodia***, ***8.Vibhu Singh*** and ***6.Kavinder Singh***, Computer, DTU

2. A Detailed Study on Generative Adversarial Networks, ***3.Shailender Kumar*** and ***7.Sumit Dhawan***, Computer, DTU

3. Adaptive Droop based Control for Active Power Sharing in Autonomous Microgrid for Improved Transient Performance, Surya Prakash, ***8.1.Vaibhav Nougain***, and Sukumar Mishra, Electrical, DTU

4. An Embedding-based Deep Learning Approach for Movie Recommendation, ***3.Ram Murti Rawat***, ***8.Vikrant Tomar*** and ***8.Vinay Kumar***, Computer, DTU

5. Analysis of demand risks for the Indian automotive sector in globally competitive environment, ***3.Girish Kumar***, Rajesh Kr. Singh, ***8.1.Rishabh Jain***, ***8.Raman Kain*** and ***8.Naveen***, Mechanical, DTU

6. Analysis of Total Quality Management (TQM) Barriers on its Performance using TOC, ***6.Rakesh Kumar*** and ***3.R.S. Mishra***, Mechanical, DTU

7. Anomalous Object Tracking in Distributed Camera Network, ***8.MD Shahbaz Khan*** and ***3.Indu Sreedevi***, DTU

8. Arduino GSM based Power Theft Detection and Energy Metering System, ***8.Vivek Kumar Jaiswal***, ***6.Hritik Kumar Singh*** and ***3.Kuldeep Singh***, Electrical, DTU

9. Best criteria selection based PROMETHEE II method, ***6.Anjali Singh***, ***3.Anjana Gupta*** and Aparna Mehra, Mathematics, DTU

10. Breast cancer burden in Africa: evidence from GLOBOCAN 2018, ***3.Rajesh Sharma***, USME, DTU

11. Capsmf: a novel product recommender system using deep learning based text analysis model, ***3.Rahul Katarya[1]*** and ***7.1.Yamini Arora[2]***, Computer[1] and IT[2], DTU

12. Citation-Based Scientific Paper Summarization Using Game Theory, ***3.Anil Singh Parihar***, ***6.Anjali Jain*** and ***8.1.Archit Gupta***, Computer, DTU

13. Comparative Analysis of PWM Techniques and SVPWM operated V/f control of Induction Motor having different power ratings, ***8.Divyam***, ***8.1.Ayushi Saxena***, ***6.Bharat Singh*** and ***3.J. N. Rai***, Electrical, DTU

14. Comparative Analysis of Route Planning Algorithms on Road Networks, ***3.R. K. Yadav***, ***8.Giriraj Kishor***, ***8.Himanshu*** and ***8.Kishan Kashyap***, Computer, DTU

15. Comparative Study of Static VAR Compensation Techniques – Thyristor Switched Reactor and Thyristor Switched Capacitor, ***8.Divyam***, ***8.1.Ayushi Saxena***, ***6.Bharat Singh*** and ***3.J. N. Rai***, Electrical, DTU

16. Comparative study on air quality status in Indian and Chinese cities before and during the COVID-19 lockdown period, ***8.Aviral Agarwal***, Aman Kaushik, ***8.Sankalp Kumar*** and ***3.Rajeev Kumar Mishra***, Environmental, DTU

17. Control and Design of Seven-Level Cascaded Multilevel Converter for Transformerless Large-Scale Photovoltaic Integration, ***6.Jyoti Kulkarni***, ***3.Narendra Kumar*** and Bhim Singh, Electrical, DTU

18. COVID 19 Global Pandemic: Impact on Management of Supply Chain, ***6.Rakesh Kumar*** and ***3.R.S. Mishra***, Mechanical, DTU

19. Deep Learning Approach for Arbitrary Image Style Fusion and Transformation using SANET model, ***7.Pratibha Rathi***, ***7.Pranav Adarsh*** & ***3.Dr. Manoj Kumar***, Computer, DTU

20. Dementia Detection Using Machine Learning by Stacking Models, ***3.Ram Murti Rawat***, ***8.Mohammad Akram***, ***8.Mithil*** and ***8.Setti Suneel Pradeep***, Computer, DTU

21. Descriptive epidemiology of incidence and mortality of primary liver cancer in 185 countries: evidence from GLOBOCAN 2018, ***3.Rajesh Sharma***, USME, DTU

22. Design and analysis of a dual-polarization multiband oval ring patch antenna with L-stubs and folded meander line for C-band/X-band/ Kuband/ K-band communications, ***6.Karteek Viswanadha*** and ***3.N.S.Raghava***, Electronics, DTU

23. Detecting Intrusions and Attacks in the Network Traffic using Anomaly based Techniques, ***3.Vinod Kumar***, ***8.Vinay Choudhary***, ***8.Vivek Sahrawat*** and ***8.Vinay Kumar***, Computer, DTU

24. Developing a framework for evaluating sustainability index for logistics service providers: graph theory matrix approach, ***6.Anchal Gupta*** and Rajesh Kumar Singh, DTU

25. Disease Detection in Coffee Plants Using Convolutional Neural Network, ***3.Manoj Kumar***, ***8.Pranav Gupta***, ***8.Puneet Madhav*** and ***8.Sachin***, Computer, DTU

26. Dual-Modal Emotion Recognition Using Discriminant Correlation Analysis, ***3.Dr. Rajiv Kapoor***, ***8.Suhail Saifi***, ***8.Tarun Kapoor***, ***8.Utkarsh Bisaria*** and ***8.Nikhil Singh***, Electronics, DTU

27. Effect of Distributed Damping on the Dynamic Behavior of Flexible Chassis of Heavy Road Vehicle Under Standardized Random Road Responses, Ashish Gupta and ***3.Vikas Rastogi***, Mechanical, DTU

28. Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, ***6.Husain Mehdi*** and ***3.R.S. Mishra***, Mech, DTU

29. Effect of Mesh Size on Soil-Structure Interaction in Finite Element Analysis, ***7.Nasasira Derrick*** and ***3.Amit Kumar Srivastava***, Civil, DTU

30. Efficient strip to slot waveguide mode converter using fast quasi-adiabatic approach, Nikhil Dhingra, Geetika Jain Saxena, ***3.Kamal Kishor*** and Enakshi Khular Sharma, Physics, DTU

31. Energy Efficient Transmission Trends Towards Future Green Cognitive Radio Networks (5G): Progress, Taxonomy and Open Challenges, ***6.Akanksha Srivastava***, Mani Shekhar Gupta and Gurjit Kaur, Electronics, DTU

32. Ensemble learning-based approach for crack detection using CNN, ***8.Vibhu Kailkhura***, ***8.S. Aravindh***, ***8.Shashank Shekhar Jha*** and ***3.N. Jayanthi***, Electronics, DTU

33. Environment Monitoring System for Agricultural Application using IoT and Predicting Crop Yield using Various Data Mining Techniques, ***8.Gautam Gupta***, ***8.Rashi Setia***, ***8.Archana Meena*** and ***3.Bhavnesh Jaint***, Electrical, DTU

34. Estimation of COVID-19 Prevalence under Closed Population using Capture-Recapture approach, ***6.Rakesh Kumar*** and ***3.R.S.Mishra***, Mechanical, DTU

35. Experimental studies of vapour compression refrigeration system with eco-Friendly primary refrigerant and brine mixed with nano particles as secondary refrigerant, Uma Shankar Prasad, ***3.R.S. Mishra*** and R.K. Das, Mechanical, DTU

36. Experimental Validation of Fully Informed Particle Swarm Optimization Tuned Multi-Loop L-PID Controllers for Stabilization of Gantry Crane System, ***3.Sudarshan K Valluru***, ***8.Daksh Dobhal***, ***8.Kumar Kartikeya***, ***8.Manpreet Kaur***, ***3.Madhusudan Singh*** and ***8.Arnav Goel***, DTU

37. Filter Based Scalable Blockchain for Domestic Internet of Things, ***8.Aditya Shyam Bazari***, ***8.Aditya Singh***, ***8.Abdul Ahad Khan*** and ***3.Rajni Jindal***, Computer, DTU

38. FLEXIBLE SOLAR POWER TREE - A CONCEPT, Mukesh Kumar, ***3.J P Kesari*** and Lokesh Rana, Mechanical, DTU

39. Forecasting the Probability of Cost Overrun Risk of Indian Construction Projects using Fuzzy Model, Savita Sharma, ***3.Pradeep K. Goyal*** and R.C. Chhipa, Civil, DTU

40. Fuzzy Logic based Automatic Generation Control, ***8.MD Shadan Salik***, ***8.Shaurya Priyal***, ***8.1.Neha Khokher*** and ***3.DR. Narendra Kumar***, Electrical, DTU

41. Generating Visible Spectrum Images from Thermal Infrared using Conditional Generative Adversarial Networks, ***8.Neeraj Bhat***, ***8.Navneet Saggu***, ***8.Pragati*** and ***3.Sanjay Kumar***, Computer, DTU

42. GloVeInit at SemEval-2020 Task 1: Using GloVe Vector Initialization for Unsupervised Lexical Semantic Change Detection, ***8.Vaibhav Jain***, Computer and Mathematics, DTU

43. Gratitude, social intelligence, and leadership among university teachers: mediation and moderation analysis, ***3.Naval Garg***, Anuradha Jain and B.K. Punia, USME, DTU

44. Implementation Strategy for Resource Optimization of FPGA-Based Adaptive Finite Control Set-MPC using XSG for a VSI System, Vijay Kumar Singh, ***7.1.Ravi Nath Tripathi***, and Tsuyoshi Hanamoto, Electrical, DTU

45. Modelling and optimisation of magnetic abrasive finishing process based on a non-orthogonal array with ANN-GA approach, ***6.S. Ahmad*** , ***3.R. M. Singari*** and ***3.R.S. Mishra***, Mechanical, DTU

46. Modelling Logistic Growth Model for COVID-19 Pandemic in India, ***3.Minni Jain***, ***8.Prajwal Kumar Bhati***, ***8.Pranav Kataria*** and ***8.Rohit Kumar***, CSE, DTU

47. Morphological and Swelling Potential Evaluation of Moringa oleifera Gum/Poly(vinyl alcohol) Hydrogels as a Superabsorbent, Suhail Ahmad, Kaiser Manzoor, ***3.Roli Purwar*** and Saiqa Ikram, Chemistry, DTU

48. Multi-dimensional aggregation: a viable solution for interval data, ***3.Shailender Kumar***, Computer, DTU

49. Node Classification in Complex Networks using Network Embedding Techniques, ***3.Sanjay Kumar***, ***8.Sanidhya Chaudhary***, ***8.Saksham Kumar*** and ***8.Raj Kumar Yadav***, Computer, DTU

50. Nonlinear dynamics of a time-delayed epidemic model with two explicit aware classes, saturated incidences, and treatment, ***6.Kanica Goel***, Abhishek Kumar and ***3.Nilam***, Mathematics, DTU

51. Numerical heat transfer analysis & predicting thermal performance of fins for a novel heat exchanger using machine learning, ***7.Gaurav Krishnayatra***, ***7.Sulekh Tokas*** and ***3.Rajesh Kumar***, Mechanical, DTU

52. On Black–Scholes option pricing model with stochastic volatility: an information theoretic approach, ***6.Luckshay Batra*** and ***3.H. C. Taneja***, Mathematics, DTU

53. ON CERTAIN GENERALIZATIONS OF S, ***3.S. SIVAPRASAD KUMAR*** AND ***6.KAMALJEET GANGANIA***, Mathematics, DTU

54. On the Outage Probability and BER of a DF-Based Multi-Hop System Over and □ □ -Extreme Fading Channels, Tau Raphael Rasethuntsa, ***8.Manpreet Kaur*** and Sandeep Kumar, Computer, DTU

55. Performance analysis of beamforming aided uncorrelated mm-Wave MIMO system for IEEE 802.15.3c standard, Anand Agrawal, Vimal Kumar Mishra, Ashish Gupta and ***6.1.Bajrang Bansal***, Electronics, DTU

56. Performance of Masonry Infill Framed Structures During Earthquake: A State-Of-The-Art Review, Priyanka, ***3.Alok Verma*** and Shobha Ram, Civil, DTU

57. PMS-Sharing : Framework for Automatically Authenticating users in a Group to Allow Sharing Storage, Nitesh Suthar and ***3.Rajesh Kr. Yadav***, Computer, DTU

58. Predicting Link Sign in Online Social Networks based on Social Psychology Theory and PageRank, ***3.Sanjay Kumar***, ***8.Sudhanshu Shekhar Singh*** and ***8.Vipul Gupta***, Computer, DTU

59. Real Time Experimental control and Design of FOP□□ and L-PID Controllers tuned by Invasiveweed optimization for Trajectory Control of TRAS, ***7.1.Hem Prabha*** and ***8.Rajul Kumar***, Electrical, DTU

60. Realization of Regula-Falsi Iteration based Double Precision Floating Point Division, ***8.Riya Jain*** and ***3.Neeta Pandey***, Electronics, DTU

61. Real-Time Implementation of the Optimal Predictor and Optimal Filter: Accuracy Versus Latency, SYED ASEEM UL ISLAM, ***8.1.ANKIT GOEL*** and DENNIS S. BERNSTEIN, Mechanical, DTU

62. Selection of sustainable transport system: a case study, ***8.Abhishek Singh***, Amulya Gurtu and Rajesh Kumar Singh, Civil and Environment, DTU

63. SIX SIGMA APPLICATION IN KETCHUP INDUSTRY FOR DAMAGED GLASS BOTTLES: A CASE STUDY, ***8.1.YATHARTH SHARMA***, ***8.SARMAD HASSAN*** and ***8.RAJAT KUMAR***, Mechanical, DTU

64. Sodium hydroxide effect on the mechanical properties of flyash-slag based geopolymer concrete, ***6.Manvendra Verma*** and ***3.Nirendra Dev***, Civil, DTU

65. Spatial Landslide Risk Assessment at Phuentsholing, Bhutan, Abhirup Dikshit, ***3.Raju Sarkar***, Biswajeet Pradhan, Saroj Acharya and Abdullah M. Alamri, Civil, DTU

66. SUBORDINATION AND RADIUS PROBLEMS FOR CERTAIN STARLIKE FUNCTIONS, ***3.S. SIVAPRASAD KUMAR*** AND ***6.KAMALJEET GANGANIA***, Mathematics, DTU

67. Suppression of Lower Order Harmonics for the full modulation range for a 2-level inverter fed IM drive with a switched capacitive filter technique forming a 42-sided Voltage Space Vector Structure, ***7.1.Rahul Dewani***, Rakesh R, K. Gopakumar, Umanand Loganathan Dariusz Zielinski and Leopoldo G. Franquelo, Electronics, DTU

68. The prospective application of a graphene/MoS2 heterostructure in Si-HIT solar cells for higher efficiency, Chandra Kamal Borah, ***3.Pawan K. Tyagi*** and Sanjeev Kumar, Physics, DTU

69. Theoretical analysis for transmission of Gaussian and sine time irradiance of electromagnetic beam in collisional dusty plasmas, ***6.Ruchi Sharma*** and ***3.Suresh C. Sharma***, Applied Physics, DTU

70. Theoretical Framework of Collaboration between MSME and Private Engineering Colleges to Co-exist during COVID-19, ***3.Seema Singh***, Humanities, DTU

71. Towards Social & Engaging Peer Learning: Predicting Backchanneling and Disengagement in Children, ***8.Mononito Goswami***, ***8.Minkush Manuja*** and ***8.Maitree Leekha***, Computer, DTU

72. Tribological characterisation of thermal sprayed CrC alloyed coating – A review, ***6.Navendu Gupta***, Shailesh Kumar Singh and Shailesh Mani Pandey, Mechanical, DTU

73. Ubiquitin biology in neurodegenerative disorders: From impairment to therapeutic strategies, ***6.Dhiraj Kumar***, ***6.Rashmi K. Ambasta*** and ***3.Pravir Kumar***, Biotechnological, DTU

74. Women Empowerment in India: A Critical Analysis, ***3.Seema Singh*** and ***6.Antra Singh***, Economics, DTU

1. *Chancellor*

2. *Pro Vice Chancellor*	***2.1.*** *Ex Pro Vice Chancellor*
3. *Faculty*	***3.1.*** *Ex Faculty*
4. *Teaching-cum-Research Fellow*	***4.1.****Asst. Librarian*
5. *Alumni*	
6. *Research Scholar*	***6.1.*** *Ex Research Scholar*
7. *PG Scholar*	***7.1.*** *Ex PG Scholar*
8. *Undergraduate Student*	***8.1.*** *Ex Undergraduate Student*

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

A Comparative Study of Image Dehazing Algorithms

Anil Singh Parihar, Yash Kumar Gupta, Yash Singodia, Vibhu Singh, Kavinder Singh
Machine Learning Research Lab, Dept. of Computer Science
Delhi Technological University
Delhi, India
parihar.anil@gmail.com, yashk.gupta97@gmail.com, yashsingodia@gmail.com, vibhusingh98@gmail.com,
kavinder85@gmail.com

***Abstract*—Haze poses a great challenge in modern-day applications. Removing haze is a challenging task because haze varies with the depth of the scenes in the image. Many automated systems like surveillance systems, object tracking, etc. use the dehazing methods internally to improve their overall performance in a hazy environment. Hence, recovering the original image from the hazy image is a crucial task. This paper provides a detailed survey of many algorithms that have been proposed for haze removal, which refines the image, either by using statistical observation of the scene or by network-based learning method. These algorithms have been assessed quantitatively and visually to compare based on their dehazing potential. The paper presents a comparative study of the various image dehazing methods having a different way of estimation of transmission and atmospheric light. This paper reviews different prior-based technique and learning-based technique image dehazing algorithms.**

***Index Terms*—Image dehazing, learning-based, CNN, Enhancement, Atmospheric scattering model**

I. Introduction

Image enhancement endeavours to improve the visual quality of an image. Image can degrade due to lighting, atmosphere, device issues and many other reasons. Researchers presented different approaches to deal with different issues and achieve image enhancement. Contrast enhancement [1]–[5], Haze removal [6]–[14], and Low-light enhancement [15]–[17] are a part of Image Enhancement.

The visibility of outdoor regions is deteriorated in adverse weather conditions with the presence of haze, fog, water droplets and various different particles in the atmosphere. The presence of haze is a dominant issue that decreases the contrast level and brings the noise to the image. Consequently, haze removal and improving the colour of the image becomes a difficult task. The volume of the haze or fog depends on the floating particles in the environment. These bits of particle absorb and scatter light, which reflects back from scene objects. The degree to which light scatters, rely on the separation gap between the object and the observer. Object tracking [18], Object detection [19], Action recognition [20], surveillance systems ,environment monitoring and other automated systems require clear and haze-free images to work properly. Moreover, It is difficult to perform any other vision-based task effectively in a hazy image. Thus, Haze removal becomes a necessity. Many researchers [21], [22] performed haze removal with the help of multiple images. These methods attempt to take more than one scene-image with an intent to learn the different features of the image. Separate images contain different features and thus, can help in capturing haze relevant feature effectively. However, it is sometimes not possible to have multiple images for feature extraction. It is a challenging task to distil out pertinent haze features from a single input image because of very less information provided by a single image. Various researchers [7], [8], [23]–[25] endeavoured to single image haze removal. The dehazed outputs produced by many single image dehazing algorithms appeared pleasing to the eye and the overall complexity also got reduced. Hence, addressing this problem is important in computer vision to recover images from the haze regions for useful information. The paper presents a contrasting examination of several image dehazing algorithms.

Section-II presents Atmospheric scattering model. Section-III contains Brief discussion of various single image dehazing methods. Section-IV propounds a detailed comparative analysis. Section-V is used for the conclusion.

II. Atmospheric Scattering Model

The Atmospheric scattering model [6] is extensively exploited to deal with hazy images. The model is stated as:

$$\mathbf{I}^c = \mathbf{R}^c\mathbf{T} + L^c(1 - \mathbf{T}) \tag{1}$$

here **I** depicts input-image, **R** denotes the scene radiance of the image, c denotes colour channels, and Transmission **T** depicts quantity of light that falls straight to lens of the

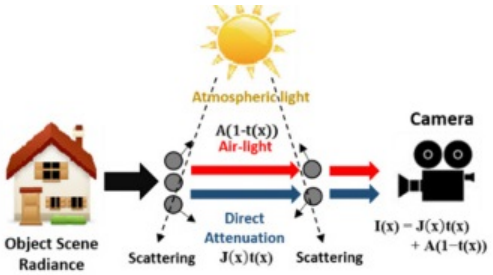


Fig. 1. Atmospheric Scattering Model [26]

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:00:15 UTC from IEEE Xplore. Restrictions apply.

camera without getting dispersed and L tells the atmospheric light. Also, the transmission is proportionate to the exponential distance between object and lens, therefore

$$\mathbf{T} = e^{-\beta \mathbf{D}} \tag{2}$$

here β is the atmosphere coefficient of scattering and **D** denotes the distance between the camera and the object. The model is widely employed in image dehazing to produce enhanced results **R** from input **I** by evaluating **T** and L.

III. LITERARY SURVEY

The main objective of image dehazing methods is to produce outputs that contains appropriate visible object in the image with almost no noise. This paper presents a brief deliberation of various image dehazing methods that are commonly used. Image dehazing can be used in variety of applications. Various methods are proposed by researchers to remove haze from an image and generate images with better image quality. The analysis contains prior-based and learning-based approaches. In the following section, various algorithms will be discussed.

A. DARK CHANNEL PRIOR

The dark channel prior (DCP) [7] forms on the statistical knowledge of outside scene without fog. It is built on the inspection that the local patchs Ω, which are not a part of sky region, have some pixels, with less intensity values in no less than one of the colour channel. These low intensities are accounted by the presence of shadows, colourful objects and dark surfaces. DCP is built on atmospheric model of image formation (1). The approach estimates Atmospheric light using dark-channel which is given as:

$$\mathbf{R}^{dark} = \min_{c \varepsilon (R,G,B)} (\min_{y \varepsilon \Omega} \mathbf{R}^c) \tag{3}$$

where $\mathbf{R}^{dark}$ denotes dark-channel, and Ω denotes a patch size. After estimation of dark-channel, it estimates Transmission map as:

$$\widehat{\mathbf{T}} = 1 - \omega \min_{c} (\min_{y \varepsilon \Omega} (\mathbf{I}^c / L^c)) \tag{4}$$

where $\widehat{\mathbf{T}}$ denotes Transmission map, and L^c denotes the channel-wise atmospheric light. DCP, assumes that the R^{dark} of a haze free image is nearly 0, and L^c is always non-negative.

$$\mathbf{R}^{dark} \rightarrow \min_{c \varepsilon R,G,B} (\min_{y \varepsilon \Omega} \mathbf{R}^c(y)) \tag{5}$$

Since, the Complete elimination of haze may appear displeasing to the eye. Thus, DCP preserves few traces of haze in the image to take care of the aerial perspective for distant objects. It is done by introducing a new parameter ω, which is set to 0.9.

$$\widehat{\mathbf{T}} = 1 - \omega \min_{c} (\min_{y \varepsilon \Omega} (\mathbf{I}^c(y) / L^c)) \tag{6}$$

The scene radiance can be recovered by using estimated Transmission, and Atmospheric light. Hence, It computes the final dehazed image as:

$$\mathbf{R} = \frac{\mathbf{I} - L}{max(\mathbf{T}, t_0)} \tag{7}$$

where t_0 denots the lower bound for Transmission, which keeps transmission at particular lower limit.

DCP work appropriately when images do not contain any white or brighter object. DCP generates artifacts while dealing with images having white object. In DCP, transmission is dependent on atmospheric light and prior. Thus, a improved prior can provide better enhancement.

B. COLOR ATTENUATION PRIOR

Colour attenuation prior (CAP) [8] is built on the scrutiny that, the difference in the intensity of brightness and that of an image saturation varies hazy regions of an image than haze-free regions greatly. For hazy parts, this difference is vast, whereas, the difference is low for the haze-free regions. Hence, It could be inferred that $\mathbf{D} \propto \rho \propto \mathbf{B}$ - **S**, where **D** symbolize the depth of scene, ρ is the haze concentration in the image, **S** is the intensity of saturation whereas **B** is the intensity brightness of the scene. CAP estimates depth map using the following learning model:

$$D(p) = \delta_0 + \delta_1 B(p) + \delta_2 S(p) + \varepsilon(p) \tag{8}$$

p is the position, in the image, $\delta_0, \delta_1, \delta_2$ are unknown parameter and θ_0 is equal to 1.0, θ_1 is equal to -1.0 and ϵ is a input image. The depth map can be estimated using equation (8). CAP estimates of atmospheric light by choosing the brightest 0.1% pixel in the depth and selecting the pixels intensity corresponding to them in the input image. Transmission map can be estimates by using Eq (3). Hence, The scene radiance can be estimated as:

$$\mathbf{R} = \frac{\mathbf{I} - L}{min(max(e^{-\beta \mathbf{D}}, 0.9), 0.1)} + L \tag{9}$$

It limits transmission map between a lower bound and upper bound to achieve natural enhancement. CAP performs better than DCP as it uses depth for transmission estimation. However, it sometimes provides limited enhancement. Since, it uses hard thresholding to limit the scene radiance.

C. RANKING CNN

Ranking Convolutional Neural Networks [11] attempts to cover the drawbacks of the results attained from [9] and [10].Any image has two features that play a major role in the formation of transmission map. They are called the structural and the statistical features. Some of the finest CNN may attain success in evaluating the structural features but most of them

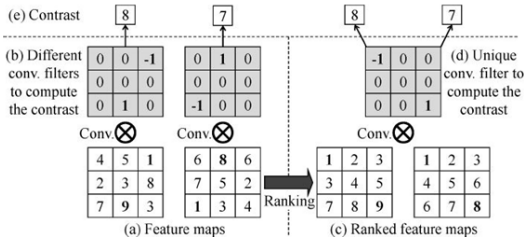


Fig. 2. Calculation of statistical feature (contrast) [11]

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:00:15 UTC from IEEE Xplore. Restrictions apply.

fall short while estimating statistical details. The Ranking layer was proposed keeping both these features in mind as it can estimate both the above said features. Ranking Layer takes feature maps as input and then return an output of ranked maps with the same shapes and sizes. The role of this layer is just to change the ordering of the input map. This aids the overall dehazing process by increasing its efficiency in calculation of features.

It can be observed from the 2 that for the calculation of the statistical feature other CNN may need many convolution filters but the ranking layer needs only one filter.This Ranking Layer is then inserted into the classical CNN to get an estimation over the pixel-wise transmission. The remaining part is a regression network followed by a Random Forest generator to increase performance. The obtained result is then passed for smoothing and rectified result is then coalesced with L, which is evaluated by calculating the 0.1% pixels having the highest intensities in the Dark Channel, to produce final results.

Ranking CNN attempts to learn haze characteristics using ranking layer. The other techniques are unable to distinguish statistical and structural details in the image. Ranking CNN evaluates statistical details with structural details from an image. While the other methods were taking many traversals of the filters to calculate the statistical details, the ordering of elements in the layer aids in the calculation of statistical details like contrast easily and does not require many traversals. This approach does not assumes that the transmission is constant locally instead it aims to find transmission at every pixel of the image, thereby increasing the overall image enhancement ability.

D. DEHAZE NET

DehazeNet [10] was proposed which being a [CNN] could be trained for memorizing haze characteristics and the effects they have in transmission. DehazeNet attempts to learn associations between hazy areas and their transmission. The attributes such as Dark Channel, Maximum Contrast, Colour Attenuation and Hue Disparity, if removed from an image makes the image haze-free. To improve the chances of convergence this method also brings with it a novel non-linear activation function known as Bilateral-Rectified Linear Unit (BReLU).

The structure of DehazeNet is a combination of many convolutional, pooling layers stacked back to back along with the presence of activation functions as shown in Fig. 3 The main feature of DehazeNet includes haze relevant characteristic extraction from the whole image using a non linear regression based on BReLU. Network learns the training parameters through Stochastic Gradient Descent (SGD) and Mean-Squared Error (MSE) is its loss function. The end result of this network is the efficient calculation of transmission map.

During the estimation of transmission near the sky region, some of the values of the outliers in CAP [8] fluctuates. However, DehazeNet is able to calculate the transmission with greater precision. Also, CAP performance degrades badly when the concentration of haze increase which is not observed in DehazeNet. Further, BReLU is an efficient activation function that helps in image restoration and conversion.

E. MULTISCALE CNN

To evaluate the transmission MSCNN [9] presented a two step approach. First, a rough outline depicting the transmission is obtained utilizing a Coarse-Scaled Network. The produced rough result containing the overall details of the image is then passed along to the second phase. Convolution, pooling, upsampling are some of the operations performed during the first phase. Second, the Fine-Scaled Network enhances the overall structure of the intermediate passed onto it. The coarse transmission is provided here as an add-on feature map. The transmission here gets refined due to the features learned by the fine-scaled model through forward and backward iterations. MSCNN model is as shown in Fig. 4.

Some methods depends on prior based knowledge and fails as compared to MSCNN because MSCNN tries to learn the haze relevant features with the help of an end to end CNN. MSCNN method includes the effects of priors and other details acquired by the forward and backward iterations of the network. Hence, it provides a better haze representation as compared to other dehazing techniques.

F. AOD NET

All-In-One Dehazing Network (AOD-Net) [12] evaluates haze-free results directly from the foggy image without separate evaluation of transmission and airlight. AOD Net changes the atmospheric model into a K-module based model. The

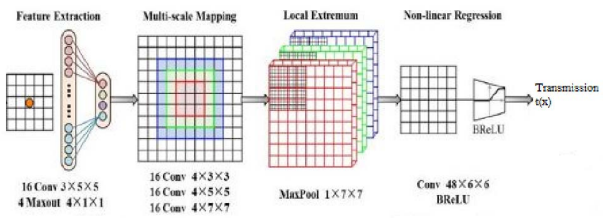


Fig. 3. Architecture of DehazeNet [10]

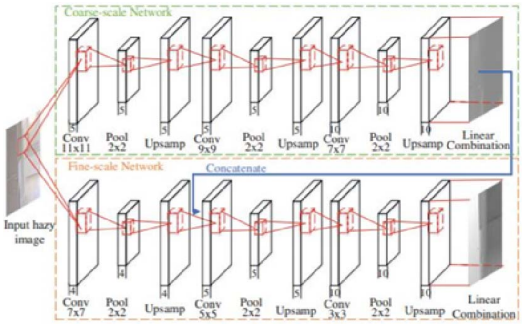


Fig. 4. MSCNN model [9]

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:00:15 UTC from IEEE Xplore. Restrictions apply.

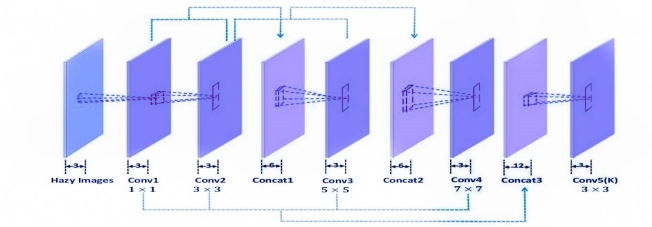


Fig. 5. Architecture of AODNet [12]

K-module based model incorporates the effect of the transmission, and airlight, Using Eq. (1).

$$\mathbf{R} = \frac{1}{\mathbf{T}}\mathbf{I} - L\frac{1}{\mathbf{T}} + L \tag{10}$$

The purpose here is to entangle L and $\frac{1}{T}$ into K.

$$\mathbf{R} = \mathbf{KI} - \mathbf{K} + b \tag{11}$$

where b has default value 1 and is a bias.Here,

$$\mathbf{K} = \frac{\frac{1}{\mathbf{T}}(\mathbf{I} - L) + (L - b)}{\mathbf{I} - 1} \tag{12}$$

After the model is converted as per the K module equation, the next step is the clean image conversion module that incorporates many pixel wise additions and multiplication layers. The results then obtained as the dehazed outputs. Because of the fact that the dehazed output is produced by a single network run, AOD Net has its application that it can be inserted with other learning based models for faster overall results. Fig. 5 shows AOD Net framework.

AOD Net is robust in cases when there are objects with colour alike atmospheric light. In those cases DehazeNet [10], MSCNN [9] show very less details. Also, the artifact problems are present in DCP [7], CAP [8], DEHAZENet [10], MSCNN [9], however AOD Net is artifact free method.

G. DCPDN (Densly Connected Pyramid Dehazing Network)

Zhang, and Patel [13] deveoped an end-to-end model for removal of haze using Generative Adversarial Network(GAN). The model learns structural relation from the images. The model consists of encoder-decoder structure which can learn transmission, airlight and can dehaze the image simultaneously. Transmission is better calculated with the help of multi-pool mapping. Also, atmospheric model has been included for better optimization for overall learning process. DCPDN has less ability for being temporally coherent. Fig. 6 represents the architecture of the model which contains 4 modules : 1. Pyramid Structure Transmission Estimation Map. 2. Atmospheric light estimation module. 3. Overall Dehazing. 4. GAN Joint Discriminator .

It optimizes GAN using the edge-preserving loss function. In practice they concatenated images free from fog with the transmission and then leveraged use of joint-discriminator loss function as:

$$\begin{aligned}\min_{G_t,G_d}\max_{D_j \text{ oint}} \mathbb{E}_{I\sim p_{\text{data }(I)}}\left[\log\left(1 - D_j \text{ oint }(G_t(I))\right)\right] + \\ \mathbb{E}_{I\sim p_{\text{data }(I)}}\left[\log\left(1 - D_j \text{ oint }(G_d(I))\right)\right] + \\ \mathbb{E}_{t,J\sim p_{\text{data }(t,J)}}\left[\log D_{\text{joint}}\left(t, J\right)\right]\end{aligned} \tag{13}$$

Several CNN-based methods like Ranking-CNN [11], MSCNN [9], restricts their ability by only taking into account of transmission map for their CNN framework whereas AOD-Net [12] solves this issue with their light-weight CNN.

H. FFA-NET

Qin *et al.* [14] presented a fusion end-to-end attention model for image dehazing. It uses Channel Attention and Pixel Attention modules which considers features based on different channels. The use of residual learning is stated to bring less vital features like thin haze. Feature Fusion Module is used to learn different features on different channels. It provides more weight to important features like thick haze regions with fusion of weights from different channels.

Fig. 7 shows Feature Attention module in FFA-Net. It consists of with basic residual learning which consists N group architectures which learns the less important features like low frequency and hazy region to avoid through multiple skip connections. The feature attention module improves the performance of the algorithm which consists of channel and pixel attention modules. Channel attention relies on the fact that the channels of image consists of different features and spatial-information is obtained by using global average pooling, with

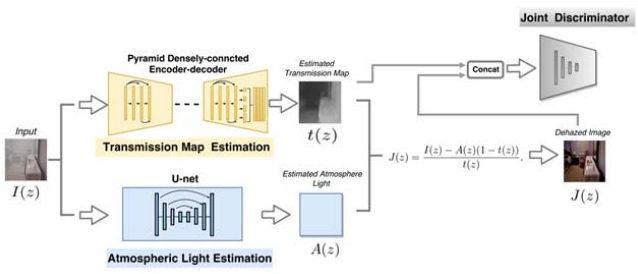


Fig. 6. Architecture of DCPDN Model [13]

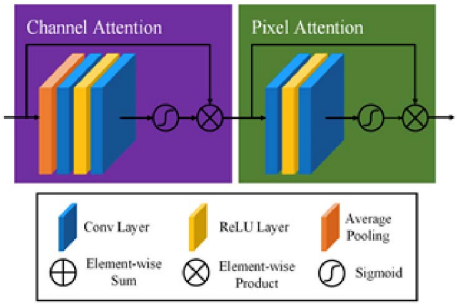


Fig. 7. Feature Attention Module [14]

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:00:15 UTC from IEEE Xplore. Restrictions apply.

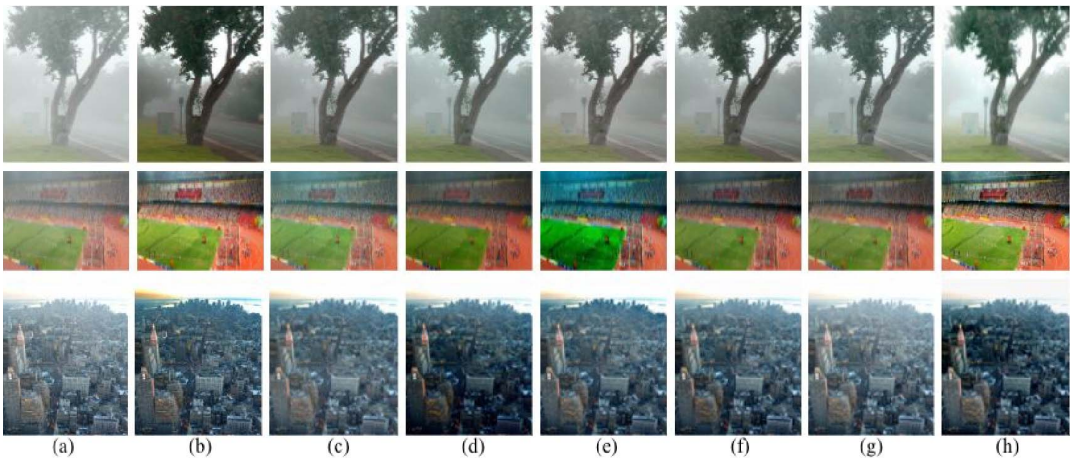


Fig. 8. Images (a) input; and result of (b) Dark Channel Prior [7]; (c) Colour Attenuation Prior [8]; (d) AOD NET [12]; (e) MSCNN [9] ; (f) Dehaze NET [10]; (g) FFA NET [14]; (h) DCPDN [13].

TABLE I
MEAN VALUE OF QUANTITATIVE MEASURES

Metrics	***He et al.'s [7]***	***Zhu et al.'s [8]***	***Cai et al.'s [10]***	***Ren et al.'s [9]***	***Li al.'s [12]***	***Zhang et al.'s [13]***	***Qin et al.'s [14]***
PSNR	17.71655	13.9468125	13.42545	14.71495	13.96785	13.06925	13.1842375
SSIM [27]	0.994475	0.98865	0.984375	0.9918375	0.9892375	0.9818625	0.9703
WPSNR [28]	20.8030625	17.9290625	16.6680625	18.91295	17.8942375	15.81585	15.12675
MSE	0.117425	0.2021125	0.2319625	0.1611125	0.175525	0.250725	0.3226125

convolution layers and element wise multiply with the weights of channels as:

$$g_{\mathrm{c}} = H_{\mathrm{p}}(F_{\mathrm{c}}) = \frac{1}{Height * Width} \sum_{i=1}^{Height} \sum_{j=1}^{Width} I_{\mathrm{c}}(i,j) \quad (14)$$

$$CA_{\mathrm{c}} = \sigma(Conv(\delta(Conv(g_{\mathrm{c}})))) \quad (15)$$

$$F^{*}{}_{\mathrm{c}} = CA_{\mathrm{c}} \otimes F_{\mathrm{c}} \quad (16)$$

,where I_c denotes $\mathrm{c_{th}}$ channel of Image with pixel position at (i,j), H_p indicates the global average pooling layer, σ is a sigmoid function, δ denotes the ReLu function, CA_c denotes the channel weights which is multiplied with input F_c.

Pixel Attention module is further used by the network to learn more information about thick hazy regions. It consists of convolution lanes and element-wise multiplication and outputs the Feature Attention module.

$$PA = \sigma(Conv(\delta(Conv(F^{*})))) \quad (17)$$

$$\tilde{F} = F^{*} \otimes PA \quad (18)$$

,where PA_c denotes output of Pixel Attention, which is further element-wise multiplied to obtain Feature Attention .

Previous CNN-based models like Ranking-CNN [11], DehazeNet [10], MSCNN [9], AOD-Net [12] handle the attention modules for channel and pixel-wise as uniform, but, as thin haze and thick haze should have different weighting factors and with the heterogeneous nature of haze distribution along different channels are also taken in consideration.

IV. QUALITATIVE AND QUANTITATIVE ANALYSIS

This section presents the analysis of above-discussed methods. For computing differences and evaluate efficiencies among different dehazing methods, both quantitative and qualitative analysis are necessary. Fig. 8 depicts the outputs that are produced by distinct dehazing techniques. Fig. 8(a) shows the input image with haze. Fig. 8(b-h) represents output of He [7], Zhu [8], Li [12], Ren [9], Cai [10], Qin [14], Zhang [13] respectively. It can be noted that Fig. 8(b) is removing haze efficiently but at times suffer from halo-artifacts near the boundaries of the subject such as tree or building. Fig. 8(c) does not suffer from excess saturation and brings out the real colours. The outputs of Fig. 8(d) have sharper details and maintains the colour accuracy of the image.In the presence of objects having colour close to white light, this method shows robustness and performs efficiently. Fig. 8(f) eradicates haze nicely and the colour distribution is also uniform thereby producing good results. But as the number of hazy images increases, the visual evaluation becomes harder and biased. Therefore it is required to present the quantitative evaluation of the above-mentioned methods. Learning-based techniques require a lot of data for learning complex features so after research, the datasets that are used in these methods for learning haze relevant features are: RESIDE [29], NYU Depth v2 [30], Middlebury Stereo [31], NTIRE [32], FRIDA [33], I-HAZE [34], and O-HAZE [35]. The codes and weights given by authors are used for fair comparison. Further, the image quality is tested using metrics such as Structure Similarity In-

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:00:15 UTC from IEEE Xplore. Restrictions apply.

dex(SSIM) [27], Peek Signal-to-Noise Ratio(PSNR), Weighted Peek-Signal-to-Noise Ratio(WPSNR) [28] , and Mean Squared Error. Since these are widely used metrices for analyzing image dehazing methods. The synthetic image data is used from the project sites of [36], [37] to gather the result with application of those metrics. Quantitative results are compared in Table 1. The value of these metrics are averaged out for the images of data set.

V. CONCLUSION

After analyzing Image dehazing approaches, it can be observed that each approach contains their own limitations. This paper depicts an extensive inspection of various state of art image dehazing approaches. It is observed that the statistical methods fail in regions where the object's colour was identical to the atmospheric light. Statistical methods performs better in quantitative analysis. However, in Qualitative analysis these methods shows various artifacts. The algorithms either produced a halo effect or erased the object. The learning based methods performed efficiently on most of the images. The contrast of learning based methods is similar to the statistical methods. The learning based approach performs appreciably in qualitative analysis. The performance of statistical methods can be improved using improved prior knowledge and parameters. Learning-based methods can overfit on a dataset thus, it must consider the overfitting conditions appropriately.

REFERENCES

[1] A. S. Parihar and O. P. Verma, "Contrast enhancement using entropy-based dynamic sub-histogram equalisation," *IET Image Processing*, vol. 10, pp. 799–808(9), November 2016.

[2] A. S. Parihar, "Histogram modification and dct based contrast enhancement," in *2018 2nd International Conference on Inventive Systems and Control (ICISC)*, 2018, pp. 614–618.

[3] A. S. Parihar, "Fuzzy adaptive gamma correction for contrast enhancement," in *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, 2017, pp. 625–629.

[4] A. S. Parihar, O. P. Verma, and C. Khanna, "Fuzzy-contextual contrast enhancement," *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1810–1819, 2017.

[5] A. S. Parihar, O. P. Verma, and D. Yadav, "Image contrast enhancement using differential evolution," in *Advances in Communication, Devices and Networking*. Singapore: Springer Singapore, 2018, pp. 517–526.

[6] E. J. McCartney, "Optics of the atmosphere: scattering by molecules and particles," *New York, John Wiley and Sons, Inc., 1976. 421 p.*, 1976.

[7] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010.

[8] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Transactions on image processing*, vol. 24, no. 11, pp. 3522–3533, 2015.

[9] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *European conference on CV*. Springer, 2016, pp. 154–169.

[10] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An end-to-end system for single image haze removal," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187–5198, 2016.

[11] Y. Song, J. Li, X. Wang, and X. Chen, "Single image dehazing using ranking convolutional neural network," *IEEE Transactions on Multimedia*, vol. 20, no. 6, pp. 1548–1560, 2017.

[12] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," in *IEEE Conference on Computer Vision*, 2017, pp. 4770–4778.

[13] H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," in *IEEE Conference on CVPR*, 2018, pp. 3194–3203.

[14] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "Ffa-net: Feature fusion attention network for single image dehazing," *arXiv preprint arXiv:1911.07559*, 2019.

[15] K. Singh and A. S. Parihar, "A comparative analysis of illumination estimation based image enhancement techniques," in *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, 2020, pp. 1–5.

[16] A. S. Parihar and K. Singh, "A study on retinex based method for image enhancement," in *2018 2nd International Conference on Inventive Systems and Control (ICISC)*, Jan 2018, pp. 619–624.

[17] A. S. Parihar and K. Singh, "Illumination Estimation for Nature Preserving low-light image enhancement," *10.36227/techrxiv.12236780.v1*, 5 2020. [Online]. Available: https://www.techrxiv.org/articles/Illumination_Estimation_for_Nature_Preserving_low-light_image_enhancement/12236780

[18] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. Torr, "Fully-convolutional siamese networks for object tracking," pp. 850–865, 2016.

[19] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," pp. 779–788, 2016.

[20] S. Ji, W. Xu, M. Yang, and K. Yu, "3d convolutional neural networks for human action recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 221–231, 2012.

[21] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, "Instant dehazing of images using polarization," in *IEEE Conference on CVPR*, vol. 1. IEEE, 2001, pp. I–I.

[22] Y. Y. Schechner, S. G. Narasimhan, and S. Nayar, "Polarization-based vision through haze," *Applied optics*, vol. 42, no. 3, pp. 511–525, 2003.

[23] J.-P. Tarel and N. Hautiere, "Fast visibility restoration from a single color or gray level image," in *IEEE Conference on Computer Vision*, 2009, pp. 2201–2208.

[24] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *IEEE conference on computer vision*, 2013, pp. 617–624.

[25] C. O. Ancuti and C. Ancuti, "Single image dehazing by multi-scale fusion," *IEEE Transactions on Image Processing*, vol. 22, no. 8, pp. 3271–3282, 2013.

[26] C.-C. Tsai, C.-Y. Lin, and J.-I. Guo, "Dark channel prior based video dehazing algorithm with sky preservation and its embedded system realization for adas applications," *Optics express*, vol. 27, no. 9, pp. 11 877–11 901, 2019.

[27] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[28] J. Mannos and D. Sakrison, "The effects of a visual fidelity criterion of the encoding of images," *IEEE transactions on Information Theory*, vol. 20, no. 4, pp. 525–536, 1974.

[29] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 492–505, 2018.

[30] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from rgbd images," in *European conference on CV*. Springer, 2012, pp. 746–760.

[31] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *IEEE Conference on CVPR*. IEEE, 2007, pp. 1–8.

[32] C. Ancuti, C. O. Ancuti, and R. Timofte, "Ntire 2018 challenge on image dehazing: Methods and results," in *IEEE Conference on CVPR Workshops*, 2018, pp. 891–901.

[33] J.-P. Tarel, N. Hautiere, L. Caraffa, A. Cord, H. Halmaoui, and D. Gruyer, "Vision enhancement in homogeneous and heterogeneous fog," *IEEE Intelligent Transportation Systems Magazine*, vol. 4, no. 2, pp. 6–20, 2012.

[34] C. O. Ancuti, C. Ancuti, R. Timofte, and C. D. Vleeschouwer, "I-haze: a dehazing benchmark with real hazy and haze-free indoor images."

[35] C. O. Ancuti, C. Ancuti, R. Timofte, and C. De Vleeschouwer, "O-haze: a dehazing benchmark with real hazy and haze-free outdoor images," in *IEEE CVPR*, 2018, pp. 754–762.

[36] R. Fattal, "Dehazing using color-lines," *ACM transactions on graphics (TOG)*, vol. 34, no. 1, pp. 1–14, 2014.

[37] D. Berman, S. Avidan *et al.*, "Non-local image dehazing," in *Proceedings of the IEEE conference on CVPR*, 2016, pp. 1674–1682.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:00:15 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

A Detailed Study on Generative Adversarial Networks

Shailender Kumar
Dept. of Computer Science and Engineering
Delhi Technological University
Delhi, India
shailenderkumar@dce.ac.in

Sumit Dhawan
Dept. of Computer Science and Engineering
Delhi Technological University
Delhi, India
sumitdhawan6@gmail.com

***Abstract*—From past decades, along with increase in computing power, different generative models have also been developed in the area of machine learning. Among all such generative models, one very popular generative model, called as Generative Adversarial Networks, has been introduced and researched upon, in past few years only. It is based on the concept of two adversaries constantly trying to outperform each other. Objective of this review is to extensively study the GAN related literature and provide a summarized form of the studied literature available on GAN, including the concept behind it, its objective function, proposed modifications in base model, and recent trends in this field. This paper will help in giving a thorough understating of GAN. This paper will give an overview of GAN and discuss the popular variants of this model, its common applications, different evaluation metrics proposed for it, and finally its drawbacks, conclusion of the paper and future course of action.**

Keywords—deep learning; generative adversarial networks; neural network; generative model; image processing

I. INTRODUCTION

There has been many advancements in the field of machine learning, especially deep learning as more and more computing or processing power is becoming available. Deep Learning helps us in extracting high level, useful and abstract features from the input data and using those features in classification and generation tasks. This approach is commonly known as representation learning and is based on how the human mind learns anything. Concept of generative models (based on deep learning) forms the basis of Generative Adversarial Networks (or GANs). Traditional generative models like Restricted Boltzmann machine [15] and Variational Auto Encoder [14] were based on concepts like Markov chains and maximum likelihood estimation. Based on the distribution of input data, they estimate the distribution of generated data, but as a result of not being good in generalization, their performance and outcome is affected. A new concept in the field of generative models, called as Generative Adversarial Networks (GANs) was introduced in the year of 2014 by Goodfellow et al. [1]. It consists of one generator and one discriminator, which are adversaries of each other and thus, constantly trying to outperform each other and thus, improving themselves in the process of doing so. GANs are based on learning the joint probability distribution.

Image Synthesis is a problem which has been researched a lot upon. It means generating an image based on the hidden and visible features of an already existing image. GANs are being used largely in the area of image synthesis (or image processing, in general) as they have been proven to work really well with the images. GANs consist of two models which are trained simultaneously against each other. Generator is tasked with generating new data points based on how the data points are distributed in input samples and thus, deceiving the Discriminator with the generated/fake data points as real data points. Discriminator is tasked with catching the bluff of Generator by classifying the received data points as generated or a real (taken directly from the input sample space). It is similar to a zero-sum game between two opponents. Models are trained using back-propagation [17] and dropouts (to prevent over-fitting).

GANs can be classified based on their architecture and loss function used to train the generator [16]. It has been shown that GANs can produce pretty good and high quality images which seem convincing enough to be considered as real images. GANs have grown exponentially since they were first proposed in 2014. As of now, there are many variants of generative adversarial networks, each created specifically for a specific problem in multiple research areas like image synthesis, style transfer, image enhancement, image to text conversion, text to image conversion, detecting an object, etc., making GANs a hot topic for research purpose as well as in real world applications. Despite all the growth and research going on, GANs suffer from some problems and have shortcomings [9] like vanishing gradient, mode collapse, non-convergence, absence of universal performance evaluation metrics. Also, various solutions have been proposed for above mentioned drawbacks.

This survey analyses and presents theory, applications, shortcomings, state of the art variants of GANs and recent trends in this field. This survey is organized in parts as follows. Section II defines the background, structure and loss function of GANs. A comparison of variants of GANs is presented in Section III. Section IV discusses different evaluation metrics. The applications of GANs are given in Section V. Section VI

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:56:00 UTC from IEEE Xplore. Restrictions apply.

discusses the shortcomings of GANs. Finally, Section VII gives the conclusions and future scope.

II. THEORY AND STRUCTURE OF GAN

A. Basic Theory

Behind the motivation for GANs, is Nash equilibrium of Game Theory [1], which is a solution for a non-cooperative game between two adversaries, in which each player already knows all the strategies of other player, therefore no player gains anything by modifying their own strategy.

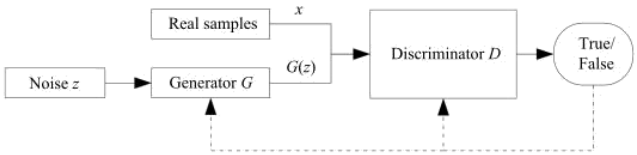


Fig. 1. Structure of Generative Adversarial Networks

The main aim of GAN is to arrive to this Nash equilibrium. Fig. 1 depicts the basic structural model of a standard GAN. Any function which can be differentiated can be used as the function for equations of generator as well as discriminator. Here, G and D are two differentiable functions that represent the generator and the discriminator respectively. Inputs given to D is x (real data), and z (random data or simply, noise). Output of G is fake data produced as per the probability distribution of actual data (or pdata), which is termed as G(z). If actual data is given as input to Discriminator, it should classify the input data as real data, labeling it 1. If fake or generated data is given as input to Discriminator, it should classify the input data as fake data labeling it 0. Discriminator strives to classify the input data correctly as per the source of data, On the other hand, Generator strives to deceive the Discriminator by making generated data G(z) similar and in line with the real data x. This game like adversarial process helps in improving the performance of both Discriminator and Generator slowly and gradually throughout the process. Therefore, slowly, Generator is able to generate better images that look more real, because it has to fool the improved and more efficient Discriminator (when compared to the previous iteration of training).

B. Loss Function

In this part of the section, loss function and learning methodology of GANs will be discussed. Following is the mathematical representation of loss function (or objective function) of GANs, as discussed in [1].

$$min_G max_D V(D, G) = E_{x\sim p(x)}[log\, D(x)] + E_{x\sim p(z)}[log(1-D(G(z)))] \quad (1)$$

where,
x - Obtained by sampling the distribution of real data or p(x).
z - Obtained by sampling the distribution of prior data or p(z) (which can be Gaussian or Uniform distribution).
E[x] - Expected value of any random variable x.
D(x) - The probability that x is sampled from actual data and not from the generated data.

Main aim of Discriminator is to maximize the objective function by minimizing D(G(z)) along with maximizing D(x), which signifies Discriminator is classifying real data as real and fake data as fake. Main aim of Generator is to minimize the objective function by maximizing D(G(z)) which signifies that Discriminator is classifying generated data (which is fake) as real. The training data consists of real data as well as generated data. Discriminator is initially trained on real as well as fake data. After that, its weights are made non trainable. Then the generator is trained. This process is repeated until required level of performance is achieved. In initial phase of learning, due to poor performance of Generator, Discriminator can catch the lie and reject generated samples easily as they are easily distinguishable from the real samples. Therefore, saturation of log(1 − D(G(z))) happens and derivative tends to 0, preventing gradient descent from occurring. So, to correct this, Generator is now trained to maximize the value of log(D(G(z))) instead of minimizing the value of (1 − D(G(z))). This modified objective function results in the derivative to be high and provides much better gradients in the initial phases of learning

III. GAN MODELS

Since the idea of GAN was proposed, many variations have been done in the original model, resulting in different models of GANs. These variations include the improvement of structure for efficiency, changes done for any specific application or problem statement like, style transfer from one image into another [18], image enhancement [13], completing an incomplete image [20-21], and generating an image from text [19].

A. Fully Connected or Vanilla GAN

It was first introduced as a baseline model for GANs in 2014 by Goodfellow et al. [1]. As the overlap of probability distributions of actual data and fake data is very little, the measure of similarity between two distributions, also known as the Jensen-Shannon divergence, may become constant, which results in the problem of vanishing gradient. It doesn't perform that well in the case of more complex images.

B. Deep Convolutional GAN (DCGAN)

Another variant of GANs was proposed by Radford et al. (2015) [4], known as Deep Convolutional Generative Adversarial Networks (DCGANs), which is based on Convolutional Neural Networks with following changes:

- Removing fully connected hidden layers.
- In the generator, a fractional-strided convolutions are used instead of pooling layers. In the discriminator, strided convolutions are used instead of pooling layers.
- Using the concept of batch normalization in generator as well as discriminator.
- Applying ReLU activation function in all but the last layer of generative model. LeakyReLU activation function is applied in each and every layer of the discriminator.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:56:00 UTC from IEEE Xplore. Restrictions apply.

C. Wasserstein GAN (WGAN)

Wasserstein GAN (WGAN) was proposed by Arjovsky et al. [5] to resolve the issue of vanishing gradient. Instead of the Jensen-Shannon divergence, it uses the Earth-Mover distance, for calculating the similarity between probability distribution of real or actual data and fake or generated data. Discriminator is represented using f, which is a critic function, and is built on Lipschitz constraint. WGAN, though useful in a comparatively stable training of GANs, still produces data samples which are low in quality and even sometimes fail in the process of converging.

D. Wasserstein GAN with Gradient Penalty (WGAN-GP)

This model can be considered as an improvement to the basic WGAN, proposed by Gulrajani et al. [12]. They found that the weight clipping in GAN, which is done in order to apply a Lipschitz constraint on the critic, leads to failure in training. This causes abnormal behavior. So an alternative was proposed to apply the Lipschitz constraint on critic, that is, it works by penalizing the value of norm of the gradient of the f (or critic function) with respect to the input, instead of clipping of weights. This method tends to converge quickly and generates samples of relatively high quality when compared to WGAN version with weight clipping.

E. Least Square GAN (LS-GAN)

In conventional GANs, during the process of back-propagation, sigmoid cross entropy loss function is used. However, sometimes, during the learning process, the issue of vanishing gradient occurs due to this particular loss function. As a solution, Mao et al. [22] came up with LS-GAN. It uses the least square loss function. LS-GANs generate higher quality images and also are more stable during the whole learning process compared to normal GANs.

F. Semi-GAN (SGAN)

In basic GAN, during training, only one class label is required which is used to specify data source (real or fake). Odena [10] proposed Semi GAN (SGAN) by further adding class labels (say, N) of actual data while discriminator D is trained. After training, D will classify the input to one of the N+1 classes, where the extra class is used to specify the data source as per G. It is shown that this method can create a better classifier which generates samples of high quality when compared to standard GAN, as it is trained to learn the features and associates them with correct class labels.

G. Conditional GAN (C-GAN)

In conventional GAN, only latent space is provided to generator. The conditional GAN, as proposed by Mirza [11] changes that by providing an additional parameter (label y) along with latent space, to the generator and train it to produce corresponding images. Discriminator is provided with real images and label as its input, to distinguish real images better. It is shown that this model can produce digits similar to those of MNIST dataset, conditioned on class labels (0, 1, 2, 3…9). Hence, the name is Conditional GAN.

H. Bidirectional GAN (BiGAN)

In normal GANs, the generator learns by mapping latent feature vector to real data probability distribution. But it lacks an efficient mechanism which maps the real data to data in latent space. Bidirectional GANs (BiGANs) was proposed by Donahue et al. [23]. It maps the actual data probability distribution on to latent space, hence helping to learn to how to extract relevant features.

IV. EVALUATION METRICS

Apart from the manual inspection and evaluation, of samples produced by Generator module of GANs, there are some quantitative measures like Average Log Likelihood [24], Inception Score (IS) [9], Wasserstein metric [5], Frechet Inception Distance (FID) [25], etc. IS should be high and FID should be low for high quality generated images. Based on these evaluation metrics, GAN model's performance can be better judged and thus can be improved by introducing the required modifications in model.

Log-likelihood (also called as Kullback-Leibler divergence or KL divergence) is one of the widely used standard metrics for evaluating and measuring the performance of generative models [24]. It helps in measuring the likelihood of generated distribution being consistent with actual data distribution. Maximum likelihood or 0 KL divergence will produce perfect samples which will seem real.

Inception Score is probably one of the most used evaluation metric for GAN evaluation, and it was proposed by Salimans [9]. A pre-trained neural network (for example, Inception v3 model) is used to capture the relevant and important properties of generated samples, i.e., image quality and image diversity. High image quality is represented by narrow distribution of label and high image diversity is represented by uniform distribution of labels of all samples combined. As both distributions are very much dissimilar (one is narrow, another is uniform), KL divergence should be high, further giving good Inception score.

$$KL\ divergence = KL(C||M) = p(y|x) * (log(p(y|x)) - log(p(y))) \quad (2)$$

where,
C is conditional probability distribution, p(y|x) and,
M is marginal probability distribution, p(y)

Inception score doesn't measure the variations in fake images if contrasted with actual images, therefore Heusel et al. [25] introduced Frechet Inception Distance. It works on the features and transforms generated data and maps it to feature space. This feature space is obtained by the last pooling layer, just before output classification. It is summarized by calculating the covariance as well as mean for fake and actual data. By calculating this Frechet Inception Distance (also known as Wasserstein-2 distance) between these Gaussian distributions, quality and variation of produced data is measured.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:56:00 UTC from IEEE Xplore. Restrictions apply.

As of yet, there is no unanimous decision regarding which one is a better evaluation measure. Different scores focus on different aspects of the image generation process. But some metrics seem more plausible than others, like FID has the advantage of being more robust to noise. And as claimed in [25], FID can tell us how similar are real and fake images, and this measure is considered more efficient than IS.

V. APPLICATIONS OF GANS

GANs are being used widely in fields like Image processing and Natural Language Processing. Following are few of the main applications of GANs apart from classical application of generating data similar to the input data.

1) Inter-domain conversion of content of one image to another using CGAN was proposed by Isola et al. [18]. It is named pix2pix which is effective at generating photos from label maps and colorizing images.
2) Zhu et al. [18] propose CycleGAN, which can also be used for translating image from one domain to another, in the absence of a training example of paired images. It can perform style transfer, photo enhancement, object transfiguration, attribute transfer, etc.
3) Solution to the problem of Image super-resolution was provided in [13], where Ledig et al. present SRGAN. VGG network [26] is used as the discriminator. It is shown in [13] that it is possible to generate photorealistic image with high resolution. But the texture information which SRGAN generates, doesn't seem that much real, and is noisy. Therefore, Wang et al. [29] propose an Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) which aims to improve the network's architecture.
4) Zhang et al. [28] propose that GAN can generate such sentences which seem too real by utilizing the concept of long short term memory (LSTM). Li et al. [30] use GANs to perform speech to text conversion by capturing the relevance of the dialogue. SeqGAN [31] uses reinforcement learning to not only generates speech but even poems and music.
5) Antipov et al. [27] introduced a model on the concept of GAN that can be used for automatic face aging. It produces synthetic images. Facial attributes are modified in similar kinds of work but in this, emphasis is given on preserving basic facial attributes and original person's identity in aged version of the face.



Fig. 2. Few examples of applications of GANs [8]

VI. DRAWBACKS

The main shortcoming in the training process of GANs is that there is always a risk of mode collapse. It occurs when data produced by GANs is mostly focused on very less modes or sometimes even on one single mode, which results in comparatively less diverse samples. A solution to the above problem is batch regularization or mode regularization. In this, decent size of batches of data samples or various data points are included so that the diversity is improved. Another solution is proposed in [33], in which the samples generated by different models are combined together. Moreover, optimizing the objective function (like done in WGAN [5]) can resolve the problem as well.

The lack of stability in the process of training is a big issue in itself. The GAN model's parameters oscillate, destabilize and never converges to the Nash equilibrium, that is each opponent always countermeasures other opponent actions, making the models harder to converge. In contrast to the other generative models, GANs evaluation problem is much more challenging as there is no consensus on what evaluation metrics should be used for GANs [34]. Another shortcoming of GANs is the problem of diminished gradient. It occurs if the discriminator gets very powerful while training, causing the gradient of generator's function to reduce and slowly vanish, and in turn, the generator learns nothing. This imbalance between models of generator and discriminator, results in overfitting. Also, GANs are very much sensitive to the selection of hyper parameters.

VII. CONCLUSION AND FUTURE SCOPE

This paper analyses and summarizes the background of GANs, its basic theory, structure, variants, evaluation metrics, applications, shortcomings and future scope. It gives an overall and broad understanding of the literature present on GAN. A wide variety of applications of GANs is shown in this paper. New as well as better solutions to new and existing challenges in GANs are yet to be explored with an aim to increase the efficiency of GANs. The field of GAN is a promising research area, but even with all the development in its research area, it has its own problems like unstable training, non-convergence, lack of consensus on an evaluation metric, requirement of more computing power and the model's complexity. In conclusion, GAN is an interesting and useful research field with many applications but also, a lot of work needs to be done to overcome existing challenges, as it is still comparatively a new field.

New work is being continuously done to overcome the limitations of GANs. For example, WGAN can resolve the mode collapse problem as well as the training instability problem, but only partially. Therefore, preventing the problem of mode collapse in GANs remains an open research problem. Also, other research areas include the presence of Nash equilibrium and the theory of the convergence of a GAN model. GANs are being widely utilized in the area of Computer Vision but comparatively not that much in other fields like Natural Language Processing. This limitation is

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:56:00 UTC from IEEE Xplore. Restrictions apply.

because of the different properties related to image and non-image data. GANs can be used for interesting applications in various fields, therefore research is going on for that and also on how to increase the efficiency and improve the performance of GAN.

REFERENCES

[1] I. Goodfellow et al., ‘‘Generative adversarial nets,’’ in Proc. Adv. Neural Inf. Process. Syst., 2014, pp. 2672–2680. [Online]. Available: http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf.

[2] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. New York, USA: MIT Press, 2016.

[3] I. Goodfellow, “NIPS 2016 tutorial: generative adversarial networks,” arXiv: 1701.00160, 2016.

[4] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” arXiv: 1511.06434, 2015.

[5] Arjovsky, Martin, Soumith Chintala, and Léon Bottou. "Wasserstein gan." arXiv preprint arXiv:1701.07875 ,2017.

[6] L. J. Ratliff, S. A. Burden, and S. S. Sastry, “Characterization and computation of local Nash equilibria in continuous games,” in Proc. 51st Annu. Allerton Conf. Communication, Control, and Computing (Allerton), Monticello, IL, USA, 2013, pp. 917−924.

[7] Hitawala, Saifuddin. "Comparative Study on Generative Adversarial Networks." arXiv preprint arXiv:1801.04271, 2018.

[8] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In CVPR, 2017

[9] Salimans, Tim, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. "Improved techniques for training gans." In Advances in Neural Information Processing Systems, pp. 2234-2242. 2016.

[10] A. Odena, “Semi-supervised learning with generative adversarial networks,” arXiv: 1606.01583, 2016.

[11] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” arXiv: 1411.1784, 2014.

[12] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville. “Improved Training of Wasserstein GANs”. arXiv preprint arXiv:1704.00028, 2017

[13] Ledig, Christian, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew P. Aitken et al. "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network." In CVPR, vol. 2, no. 3, p. 4. 2017.

[14] Kingma, Diederik P., and Max Welling. “Auto-Encoding Variational Bayes.” ArXiv:1312.6114 [Cs, Stat], May 2014. arXiv.org, http://arxiv.org/abs/1312.6114.

[15] Fischer, Asja, and Christian Igel. “An Introduction to Restricted Boltzmann Machines.” Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, edited by Luis Alvarez et al., Springer, 2012, pp. 14–36. Springer Link.

[16] Wang, Zhengwei, et al. “Generative Adversarial Networks: A Survey and Taxonomy.” ArXiv:1906.01529 [Cs], June 2019. arXiv.org, http://arxiv.org/abs/1906.01529.

[17] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, ‘‘Learning representations by back-propagating errors,’’ Nature, vol. 323, pp. 533-536, Oct. 1986.

[18] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” arXiv preprint arXiv:1703.10593v6, 2017.

[19] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, “Generative adversarial text to image synthesis,” arXiv preprint arXiv:1605.05396, 2016.

[20] Z. Chen, S. Nie, T. Wu, and C. G. Healey, “High resolution face completion with multiple controllable attributes via fully end-to-end progressive generative adversarial networks,” arXiv preprint arXiv:1801.07632, 2018.

[21] Y. Li, S. Liu, J. Yang, and M.-H. Yang, “Generative face completion,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3911–3919.

[22] Mao, Xudong, et al. “Least Squares Generative Adversarial Networks.” ArXiv:1611.04076 [Cs], Apr. 2017. arXiv.org, http://arxiv.org/abs/1611.04076.

[23] Donahue, Jeff, et al. “Adversarial Feature Learning.” ArXiv:1605.09782 [Cs, Stat], Apr. 2017. arXiv.org, http://arxiv.org/abs/1605.09782.

[24] Hamid Eghbal-zadeh, Gerhard Widmer, “Likelihood estimation for generative adversarial networks”, 2017.

[25] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, Gans trained by a two time-scale update rule converge to a local nash equilibrium, in: Advances in Neural Information Processing Systems, 2017, pp. 6629–6640.

[26] Simonyan, Karen, and Andrew Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition.” ArXiv:1409.1556 [Cs], Apr. 2015. arXiv.org, http://arxiv.org/abs/1409.1556.

[27] Antipov et al., “Face aging with conditional generative adversarial networks”, 2017.

[28] Y. Z. Zhang, Z. Gan, and L. Carin, “Generating text via adversarial training,” Proc. Workshop on Adversarial Training, Barcelona, Spain, 2016

[29] Wang, Xintao, et al. “ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks.” ArXiv:1809.00219 [Cs], Sept. 2018. arXiv.org, http://arxiv.org/abs/1809.00219.

[30] J. W. Li, W. Monroe, T. L. Shi, S. Jean, A. Ritter, and D. Jurafsky, “Adversarial learning for neural dialogue generation,” arXiv: 1701.06547, 2017.

[31] L. T. Yu, W. N. Zhang, J. Wang, and Y. Yu, “SeqGAN: sequence generative adversarial nets with policy gradient,” arXiv: 1609.05473, 2016.

[32] Killoran, Nathan, et al. “Generating and Designing DNA with Deep Generative Models.” ArXiv:1712.06148 [Cs, q-Bio, Stat], Dec. 2017. arXiv.org, http://arxiv.org/abs/1712.06148.

[33] A. Ghosh, V. Kulharia, V. P. Namboodiri, P. H. S. Torr, and P. K. Dokania, ‘‘Multi-agent diverse generative adversarial networks,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 8513–8521

[34] Borji, Ali. “Pros and Cons of GAN Evaluation Measures.” ArXiv:1802.03446 [Cs], Oct. 2018. arXiv.org, http://arxiv.org/abs/1802.03446.

[35] Vijayakumar, T. (2019). Comparative study of capsule neural network in various applications. Journal of Artificial Intelligence, 1(01), 19-27

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:56:00 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

Adaptive Droop based Control for Active Power Sharing in Autonomous Microgrid for Improved Transient Performance

Surya Prakash, Student Member IEEE, Vaibhav Nougain, and Sukumar Mishra, *Senior Member, IEEE*

***Abstract*—This paper presents a decentralised adaptive droop based control for active power sharing between parallel inverters in autonomous microgrids. The work is focused on improving the transient performance of the system proposing a robust controller. The analysis does not require load current measurement where the Lyapunov analysis based control takes load current as a disturbance to the system. The mathematical analysis takes a more practical form of *LC* filter configuration where the dynamics due to resistances across the *LC* is also taken into consideration. Controller gain conditions are taken into consideration employing a detailed Lyapunov analysis. A comparative study of the conventional droop control with the proposed adaptive droop control is included in the work. MATLAB has been employed as the simulink platform for the comparative study. In addition, for experimental verification, FPGA based hardware setup has been developed and the related results are presented to verify the real-time operation of proposed adaptive control.**

***Index Terms*—Point of Common Coupling (PCC), Low Voltage (LV), Medium Voltage (MV), Backstepping Control, Voltage Source Converter (VSC), Phase Locked Loop (PLL), Constant Power Load (CPL), Adaptive Droop**

NOMENCLATURE

p	Instantaneous active power
q	Instantaneous reactive power
P	Filtered active power
Q	Filtered reactive power
V_{odq}	PCC point voltage in *dq* frame of reference
V^*_{odq}	Virtual PCC point voltage in *dq* frame of reference
V_{idq}	Inverter output voltage in *dq* frame of reference
I_{idq}	Inverter current in *dq* frame of reference
I^*_{idq}	Virtual Inverter current in *dq* frame of reference
I_{odq}	Load current in *dq* frame of reference
L_f	Filter Inductance
R_f	Resistance associated with the filter inductance
C_f	Filter Capacitance
R_d	Equivalent leakage Resistance associated with the filter capacitance
ω^*	Nominal system frequency
ω_o	Reference system frequency
m	Static droop for Real Power sharing
α	Adaptive droop for Real Power sharing
e_1, e_2	Current Tracking Errors
e_3, e_4	Voltage Tracking Errors

This work was supported in part by the Science & Energy Research Board (SERB), Department of Science & Technology (DST), Govt. of India through the Distinguished Investigator Award (DIA) scheme (File number - DIA/2018/000032).

The authors are with the Department of Electrical Engineering, Indian Institute of Technology Delhi, India. (e-mail: eeesuryaprakash@gmail.com, nougainvaibhav@gmail.com and sukumar@ee.iitd.ernet.in)

I. INTRODUCTION

THE penetration of power electronic interfaces in the power system has changed the course of how electricity is perceived. Power electronic interfaces have made it possible for the renewable sources (solar, wind, fuel cell etc.) to be either interface with the conventional utility grid or to be able to get locally consumed upon generation forming an autonomous microgrid [1]. The sources of such microgrid, i.e., distributed generators (DGs) are interfaced with Voltage source converters (VSCs). The VSCs, running in parallel, is responsible for controlling the frequency and voltage profile of the system due to the absence of grid [2]. The multi-DG configuration, in parallel shares the total power of the microgrid. The discussed control objective could be met by employing centralised [3], distributed [4] or decentralised methods [5].

The primary challenge for the autonomous microgrid is maintaining voltage and frequency of operation for proper operation of the loads. Once the primary challenge is dealt, the secondary challenge surfaces which are meeting the load demand in a microgrid. Once the secondary problem is addressed, the tertiary and final challenge is, maintaining the power-sharing between different sources proportionately to ensure the sources not hitting their current limit for their operational longevity. The droop control based power-sharing is widely accepted due to its decentralised nature [6]-[7]. However, droop based approach has the inherent problem of a trade-off between the voltage regulation and power-sharing accuracy [8]. To elaborate, a larger value of droop gives better power-sharing. However, the oscillations increase for such cases [9]. Transient operations in power-sharing in such microgrid setups cause overstressing, which affects the life of DG units. Mitigation of such transients is necessary to ensure the longevity of VSCs [10]-[11]. The transient and steady-state performance of the system is dependent on the control architecture employed. The conventional Proportional-Integral (PI) control has **I** (Integral) which is responsible for the robust steady-state performance. However, it is also responsible for the sluggish transient performance of the controller. The hierarchical droop based control [12]-[14] has been discussed employing radial basis

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Exeter. Downloaded on May 05,2020 at 15:09:36 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

function neural network (RBFNN). However, such schemes are sluggish due to multiple control layers. The idea of single control layer modern non-linear controllers have surfaced in the power domain for improved transients performance of such microgrid setups. Sliding mode control (SMC) [15]-[16] has been proposed to improve the transient response of the system. However, SMC inherently suffers from the problem of chattering along the sliding surface. The authors in [17] employ Adaptive Neural Network (ANN) control to conclude improved transient performance. However, the gain re-tuning for different configurations is a problem of control in [17]. The authors in [18] introduce a robust Takagi-Sugeno (T-S) fuzzy control for a parallel islanded AC microgrid. However, the complexity of such control makes its implementation difficult [19].

The idea of the proposed work is improvement of transient performance for active power sharing employing an Adaptive control. Virtual inputs [20] are introduced to frame a non-linear systematic recursive control design. The transient performance of such virtual input based control has been validated in [21]. The proposed adaptive control is a simple, robust approach without the problem of chattering as in SMC. A detailed Lyapunov analysis is used to design the gain conditions for the proposed control. The performance of the proposed adaptive control is tested under dynamic AC (RLC) transients, CPL transients and fault transients.

The rest of the paper is organised as follows. Section II presents the system architecture of the autonomous microgrid. Section III gives the modelling of parallel VSCs, including system dynamics, control problem formulation, control synthesis, generation of firing signal and tuning of control gains. Section IV gives the simulation results. Section V presents the hardware results for the proposed adaptive control.

II. System Architecture

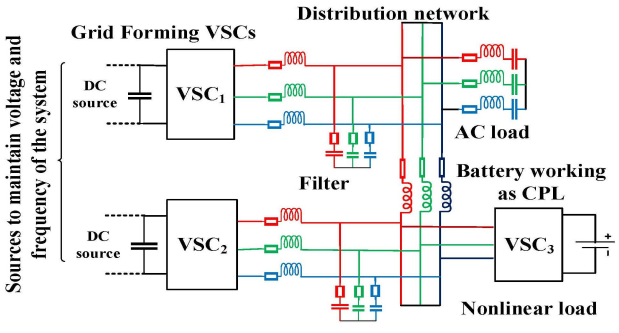


Fig. 1: Autonomous Microgrid setup

The system under consideration is an autonomous microgrid with its abstract presentation depicted in Fig. 1. Since the network is a distribution type in nature, the operating voltage of the system is 400V (phase to phase). The coupling between active and reactive power is not taken into consideration since the filter inductance value is considerable in such cases [22]. A Constant Power Load (CPL) has been taken into account due to its nature of negative incremental resistance which tends to destabilise the distribution network [23]. The different loads are distributed in nature, and hence, if the load current is a requirement for controlling the autonomous microgrid, a large number of current sensors are required. The proposed analysis does not require load current measurement, saving the cost of multiple load current sensors. The DG units and the loads are interfaced employing filter and distribution network, as shown in Fig. 1. The filter configuration employed is *LC* type where the resistances R_f (across L_f) is also included in the dynamics of the proposed analysis. The control architecture is built around the degree of freedom of the filter parameters (L_f, R_f and C_f). The power electronic interfaces, i.e., VSCs, are modulated using a control strategy obtained from Lyapunov based Adaptive technique which has been discussed in the following section.

III. Modelling of Parallel VSCs in Autonomous Microgrid

The control objective of the VSCs operating in parallel is to maintain the voltage and frequency of the autonomous microgrid, sharing the total load power in the microgrid. The instantaneous active power, p shared by each VSC is the microgrid can be calculated as

$$p = V_{od} I_{id} + V_{oq} I_{iq} \tag{1}$$

where the parameters employed can be seen in Fig. 3. The fundamental component of power, P is obtained using a low pass filter and is expressed as

$$P = \frac{\omega_c}{s + \omega_c} p \tag{2}$$

where ω_c is the cut-off frequency of the low pass filter.

The approach applied widely for such control objective is Proportional-Integral (PI) based static droop control [24]-[25]. The proposed adaptive method employs an adaptive droop control governed by,

$$\omega_0 = \omega^* - (m + \alpha) P \tag{3}$$

where ω_0 is the reference frequency of the autonomous microgrid, ω^* is the nominal frequency which is taken to be 314 *rad/s*, m is the static droop and α is the adaptive droop. Intuitively, the adaptive droop term is responsible for the improved transient response of the power-sharing in the autonomous setup. The idea behind introducing α is, increasing static droop, m improves the power-sharing; however, a low-frequency is introduced in the system if the value of static droop is very high [24]. Hence, a low value of m with adaptive α improves the power-sharing and helps to eliminate the problem of low-frequency oscillations in the system.

A. Control Design Using Lyapunov analysis based Adaptive Technique

The control design is derived using Lyapunov energy function of the autonomous system. The analysis employs adaptive technique to introduce parameter estimate ($\hat{\alpha}$) for adaptive term (α) and virtual control inputs (i^*_{1dq}, V^*_{odq}) in

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Exeter. Downloaded on May 05,2020 at 15:09:36 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

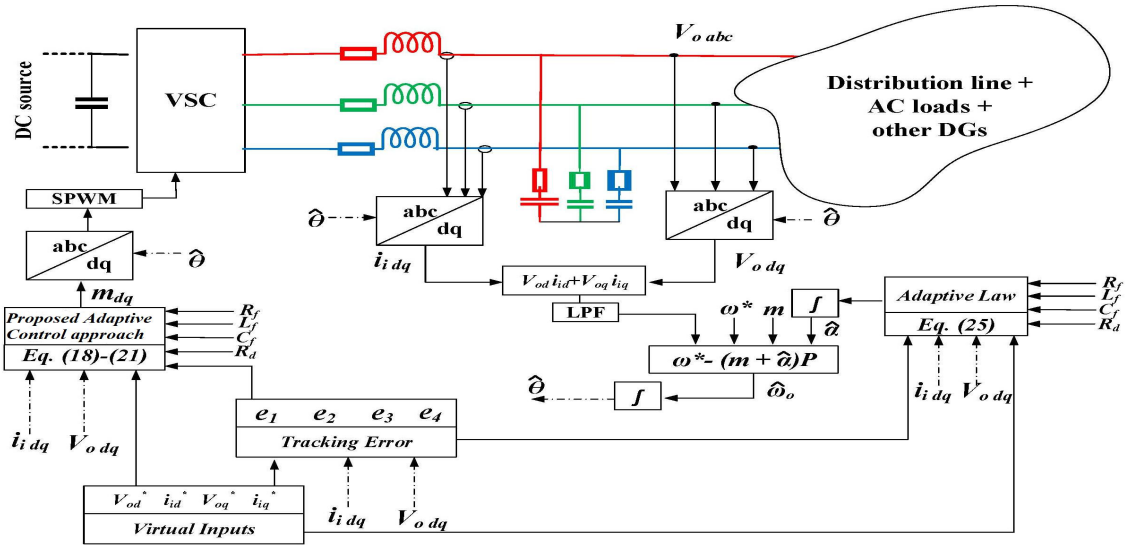


Fig. 2: Proposed Adaptive Control

the system dynamics [20]-[21], [26] (the terms are elaborated subsequently). The control technique takes load current (i^*_{odq}) as disturbance focusing on the notion of boundedness of the tracking error (e_1, e_2, e_3, e_4) dynamics. The results ensure exponential convergence of the tracking errors to a bound. Control gains introduced (γ, k_1, k_2, k_3, k_4) in the analysis are employed to vary the bound for the errors.

B. Mathematical analysis

The dynamics of the 3ϕ autonomous inverter setup is given using Kirchoff's law where the system is transformed into the direct-quadrature (*d-q*) form from the 3ϕ form. R_f are considered in the dynamics to present a more practical filter configuration. The system after the *d-q* transformation can be represented with the following differential equations where the notations are as indicated in Fig. 3.

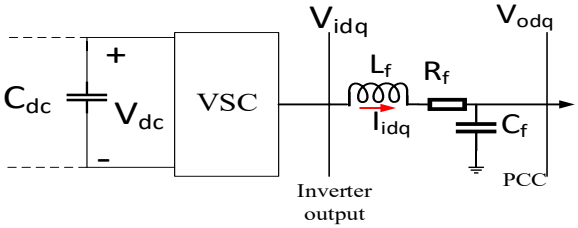


Fig. 3: The direct-quadrature equivalent of the 3ϕ (abc) single VSC configuration

1) System Dynamics:

$$\dot{i}_{1d} = -R_f i_{1d}/L_f - V_{od}/L_f + V_{id}/L_f + \omega_0 i_{1q} \quad (4)$$

$$\dot{i}_{1q} = -R_f i_{1q}/L_f - V_{oq}/L_f + V_{iq}/L_f - \omega_0 i_{1d} \quad (5)$$

$$\dot{V}_{od} = i_{1d}/C_f - i_{od}/C_f + \omega_0 V_{oq} \quad (6)$$

$$\dot{V}_{oq} = i_{1q}/C_f - i_{oq}/C_f - \omega_0 V_{od} \quad (7)$$

The discussed technique does not require load current measurement and hence i_{od} and i_{oq} are considered to be disturbances with an upper bound.

$$|i_{od}| < \bar{d}_1 \qquad |i_{oq}| < \bar{d}_2 \quad (8)$$

2) Control Problem Formulation: The closed loop system is considered for the control problem formulation. The tracking error dynamics for the system is defined as,

$$\begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} = \begin{bmatrix} i_{id} - i_{id}* \\ i_{iq} - i_{iq}* \\ V_{od} - V_{od}* \\ V_{oq} - V_{oq}* \end{bmatrix}$$

where $i_{1d}*$ and $i_{iq}*$ are the virtual input current references to the *d-axis* and *q-axis*; $V_{od}*$ and $V_{oq}*$ are the static virtual *d-axis* and *q-axis* voltage references. Inclusion of the tracking error dynamics modify eq. (4)-(7) to represent the closed loop error dynamics for the system (eq. (9)-(12)).

$$\dot{e}_1 = -R_f e_1/L_f - R_f i^*_{1d}/L_f - V_{od}/L_f + V_{id}/L_f + \omega_0 i_{1q} - \dot{i}^*_{1d} \quad (9)$$

$$\dot{e}_2 = -R_f e_2/L_f - R_f i^*_{1q}/L_f - V_{oq}/L_f + V_{iq}/L_f + \omega_0 i_{1d} - \dot{i}^*_{1q} \quad (10)$$

$$\dot{e}_3 = e_1/C_f + i^*_{1d}/C_f + \omega_0 e_4 + \omega_0 V^*_{oq} - i_{od}/C_f \quad (11)$$

$$\dot{e}_4 = e_2/C_f + i^*_{1q}/C_f - \omega_0 e_3 - \omega_0 V^*_{od} - i_{oq}/C_f \quad (12)$$

3) Adaptive Control Synthesis: The Lyapunov Energy Function for the state dynamics of the autonomous inverter is defined as;

$$V_1(e_1, e_2, e_3, e_4) = \frac{1}{2}L_f e_1^2 + \frac{1}{2}L_f e_2^2 + \frac{1}{2}C_f e_3^2 + \frac{1}{2}C_f e_4^2 \quad (13)$$

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Exeter. Downloaded on May 05,2020 at 15:09:36 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

where the energy function is positive definite and radially unbounded. The derivative of the energy function is further realized as;

$$\dot{V}_1(e_1,e_2,e_3,e_4) = L_f e_1\dot{e}_1 + L_f e_2\dot{e}_2 + C_f e_3\dot{e}_3 + C_f e_4\dot{e}_4 \quad (14)$$

Parameter estimation error for the adaptive term is defined as,

$$\tilde{\alpha} = \alpha - \hat{\alpha} \quad (15)$$

where $\hat{\alpha}$ is the known estimate of adaptive droop, α.

$$\eta = m + \hat{\alpha} \quad (16)$$

where η is a parameter used for simplification. The estimated reference system frequency is defined as,

$$\hat{\omega_o} = \omega^* - \eta P \quad (17)$$

The direct-quadrature derivative of virtual current inputs (i^*_{1d} and i^*_{1q}) and the direct-quadrature modulated inverter output voltages (V_{id} and V_{iq}) of the autonomous inverter are defined as,

$$V_{id} = R_f i^*_{1d} + V^*_{od} + L_f \dot{i}^*_{1d} - \hat{\omega_o} L_f i^*_{1q} - k_1 e_1 \quad (18)$$

$$V_{iq} = R_f i^*_{1q} + V^*_{oq} - L_f \dot{i}^*_{1q} - \hat{\omega_o} L_f i^*_{1d} - k_2 e_2 \quad (19)$$

$$i^*_{1d} = -C_f \hat{\omega_o} V^*_{oq} - k_3 e_3 \quad (20)$$

$$i^*_{1q} = -C_f \hat{\omega_o} V^*_{od} - k_4 e_4 \quad (21)$$

Hence, incorporating the derivative of virtual current inputs and the modulating inverter output voltages (eq. (18)-(21)), the derivative of the energy function is given as;

$$\begin{aligned}\dot{V}_1(e_1,e_2,e_3,e_4) = &-(R_f+k_1)e_1^2 - (R_f+k_2)e_2^2 - k_3 e_3^2 \\ &-k_4 e_4^2 - e_3 i_{od} - e_4 i_{oq} + L_f e_1 i^*_{1q}\tilde{\alpha}P - L_f e_2 i^*_{1d}\tilde{\alpha}P \\ &+C_f e_3 V^*_{oq}\tilde{\alpha}P - C_f e_4 V^*_{oq}\tilde{\alpha}P\end{aligned} \quad (22)$$

Now, the idea is to introduce another energy function to take the terms with $\tilde{\alpha}$ into consideration.

$$V_2(e_1,e_2,e_3,e_4,\tilde{\alpha}) = V_1 + 0.5\tilde{\alpha}^2 \quad (23)$$

Taking the derivative of new energy function,

$$\dot{V}_2(e_1,e_2,e_3,e_4,\tilde{\alpha}) = \dot{V}_1 - \tilde{\alpha}\dot{\hat{\alpha}} \quad (24)$$

The derivative of the estimate of adaptive term is defined as,

$$\dot{\hat{\alpha}} = -P(L_f e_1 i^*_{1q} + L_f e_2 i^*_{1d} - C_f e_3 V^*_{oq} + C_f e_4 V^*_{od}) \quad (25)$$

The idea of designing $\dot{\hat{\alpha}}$ is to shape the energy function, V_2 eliminating the terms which are impeding the decay of energy function derivative. Accounting eq. (25), the derivative of the energy function is modified to the form,

$$\dot{V}_2(e_1,e_2,e_3,e_4,\tilde{\alpha}) = -(R_f+k_1)e_1^2 - (R_f+k_2)e_2^2 - k_3 e_3^2 - k_4 e_4^2 - e_3 i_{od} - e_4 i_{oq} \quad (26)$$

The idea is to upper bound the derivative of the energy function considering the direct-quadrature profile of load current and their derivative as disturbances in the system.

$$\dot{V}_2(e_1,e_2,e_3,e_4,\tilde{\alpha}) \le -(R_f+k_1)e_1^2 - (R_f+k_2)e_2^2 - k_3 e_3^2 - k_4 e_4^2 + |e_3|\bar{d}_1 + |e_4|\bar{d}_2 \quad (27)$$

The evaluation of control gains (k_1–k_4) is based on the philosophy of exponential convergence of tracking errors (e_1–e_4) towards a bounded region around equilibrium point. The exponential convergence of tracking errors is guaranteed if the derivative of the energy function can be expressed in the form, $\dot{V}_2 \le -\gamma V_2 + c$ ensuring uniform ultimate boundedness of the energy function and the tracking errors.

$$\dot{V}_2 \le -(R_f+k_1)e_1^2 - (R_f+k_2)e_2^2 - \gamma C_f e_3^2 - \gamma C_f e_4^2 - |e_3|\left[(k_3-\gamma C_f)|e_3| - \bar{d}_1\right] - |e_4|\left[(k_4-\gamma C_f)|e_4| - \bar{d}_2\right] \quad (28)$$

The control gains (k_1–k_4) are defined as,

$$k_1 = k_2 = \gamma L_f - R_f \quad (29)$$

$$k_3 \ge \gamma C_f + \bar{d}_1/|e_3| \quad (30)$$

$$k_4 \ge \gamma C_f + \bar{d}_2/|e_4| \quad (31)$$

4) Generation of the Firing Signal for the inverter: The autonomous inverter is fired using the proposed control philosophy. The SPWM technique employed for firing includes generation of m_{dq}, inverse Park's Transformation (m_{dq} to m_{abc}), comparing with carrier sinusoid to generate the firing signal. The generation of m_{dq} is defined employing eq. (18)-(19) as,

$$\begin{bmatrix} m_d & m_q \end{bmatrix} = \frac{2}{V_{dc}} \begin{bmatrix} V_{id} & V_{iq} \end{bmatrix}$$

5) Eigen Value Synthesis: The control objective of exponential convergence of tracking error towards a bound requires tuned value of introduced control gains (k_1–k_4). Eq(29)-(31) give the range of control gains. The dynamics is then employed to do the eigenvalue synthesis studying variation of control gains on the system performance. The synthesis gives the idea of small-signal stability of the system with the variation of control gain parameters. The state-space representation is expressed as,

$$\dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{B}u(t) + \boldsymbol{D}d(t) \quad (33)$$

where $\boldsymbol{A}$ is the state matrix, $\boldsymbol{B}$ is the input matrix and $\boldsymbol{D}$ is the disturbance matrix [see eq. (32)]. $x = [e_1\ e_2\ e_3\ e_4]^T$ gives the state variables, $u = [V^*_{od}\ V^*_{oq}]^T$ gives the inputs, $d = [d_1\ d_2\ d_3\ d_4]^T$ gives the disturbances in VSC state-space representation. The eigenvalues are plotted for varying control gains. Fig. 4 shows the eigenvalue loci for varying k_1,k_2,k_3,k_4. Table I & II shows the eigenvalues for different k_1,k_2,k_3,k_4. It can be seen from table I that $k_{1min} > 0$. Similar studies are presented subsequently to verify the values of k_{2min}, k_{3min}, k_{4min} for system stability.

TABLE I: Eigen value loci with the variation of gain parameters, k_1 and k_2

λ	k_1=1000	k_1=500	k_1=0
λ_1	-471.23	-471.08	52.57
λ_2, λ_3	-436.61± 399.13i	-436.61±399.17i	-426.49±392.64i
λ_4	-1443.82	-848.15	-546.69
λ	k_2=1000	k_2=500	k_2=0
λ_1	-471.23	-471.08	52.57
λ_2, λ_3	-436.61± 399.13i	-436.61±399.17i	-426.49±392.64i
λ_4	-1443.82	-848.15	-546.69

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Exeter. Downloaded on May 05,2020 at 15:09:36 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

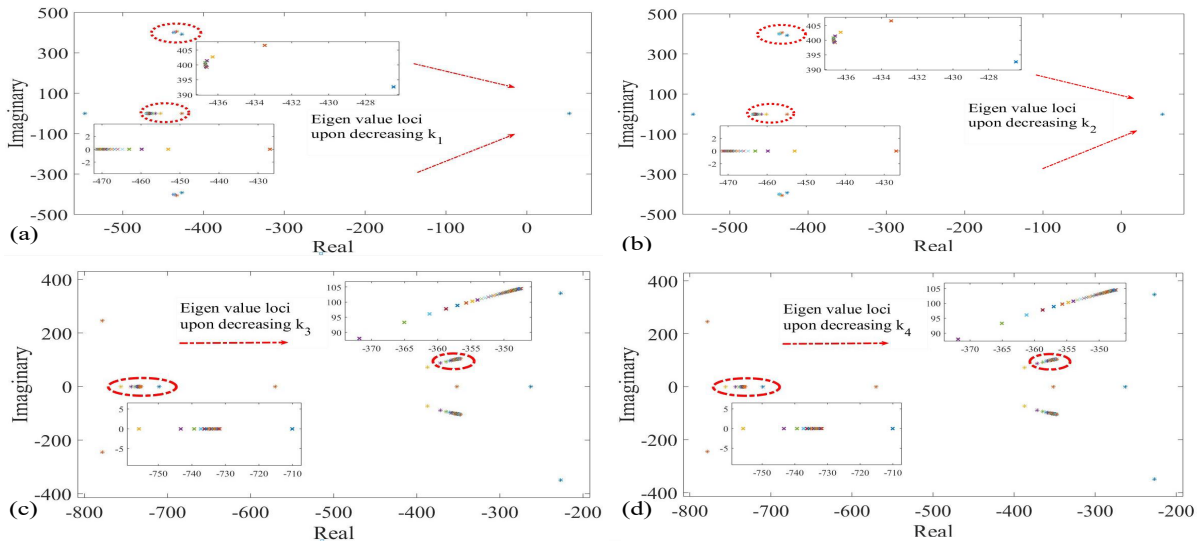


Fig. 4: Eigen Value loci for variation of (a) k_1, (b) k_2, (c) k_3, (d) k_4

$$\frac{d}{dt}\begin{bmatrix}\Delta e_1\\ \Delta e_2\\ \Delta e_3\\ \Delta e_4\end{bmatrix} = \begin{bmatrix} -\frac{R_f+k_1}{L_f} - \eta V_{od} e_2 & \hat{\omega_o} - \eta V_{oq} e_2 & -\frac{1}{L_f} - \eta i_{1d} e_2 & -\eta i_{1q} e_2 \\ \hat{\omega_o} - \eta V_{od} e_1 & -\frac{R_f+k_2}{L_f} - \eta V_{oq} e_1 & -\eta i_{1d} e_1 & -\frac{1}{L_f} - \eta i_{1q} e_1 \\ \frac{1}{C_f} - \eta V_{od} e_3 & -\eta V_{oq} e_4 & -\frac{k_3}{C_f} - \eta i_{1d} e_4 & \hat{\omega_o} - \eta i_{1q} e_4 \\ \eta i_{1d} e_3 & \frac{1}{C_f} + \eta V_{oq} e_3 & -\hat{\omega_o} + \eta i_{1d} e_3 & -\frac{k_4}{C_f} + \eta i_{1q} e_3 \end{bmatrix} \begin{bmatrix}\Delta e_1\\ \Delta e_2\\ \Delta e_3\\ \Delta e_4\end{bmatrix} + \eta \begin{bmatrix} -i_{1d} e_2 & -i_{1q} e_2 \\ -i_{1d} e_1 & -i_{1q} e_1 \\ -i_{1d} e_4 & -i_{1q} e_4 \\ i_{1d} e_3 & i_{1q} e_3 \end{bmatrix} \begin{bmatrix}\Delta V_{od}^* \\ \Delta V_{oq}^*\end{bmatrix} + \begin{bmatrix} 0 & 0 \\ -\frac{1}{C_f} & 0 \end{bmatrix} \begin{bmatrix}\Delta d_1 \\ \Delta d_2\end{bmatrix} \quad (32)$$

TABLE II: Eigen value loci with the variation of gain parameters, k_3 and k_4

λ	k_3=1000	k_3=500	k_3=0
λ_1	-731.71	-531.83	-262.70
λ_2, λ_3	-347.55± 390.81i	-347.87± 104.24i	-226.64± 349.02i
λ_4	-2105.25	-1052.61	-710.05
λ	k_4=1000	k_4=500	k_4=0
λ_1	-731.71	-531.83	-262.70
λ_2, λ_3	-347.55± 390.81i	-347.87± 104.24i	-226.64± 349.02i
λ_4	-2105.25	-1052.61	-710.05

IV. Simulation Results and Discussion

MATLAB/Simulink platform is employed for the simulation-based validation. The system considered is shown in Fig. 1. The simulation events considered for the comparison between the proposed control and the conventional PI control are elaborated in Table III. The events are kept similar to ensure proper comparison of control actions.

A. Simulation Events

1) AC load transients: Initially, both VSCs operating in parallel are contributing *320W* each (*m=22.86e-4*, α*=0*) when the AC load is consuming *640W* as can be seen in Fig. 6(a). A load transient of *640W* occurs at *t=1.3s* making the total load *1280W* as explained in Table III. The value of the adaptive droop increases to *10.27e-4* from *0* for both VSCs

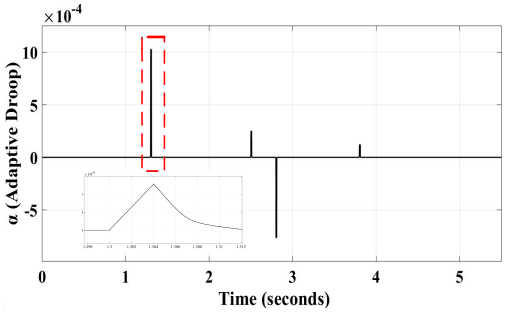


Fig. 5: Variation of adaptive droop, α with time with Event transients indicated in Table III

TABLE III: Event transients captured for comparison in Fig. 5

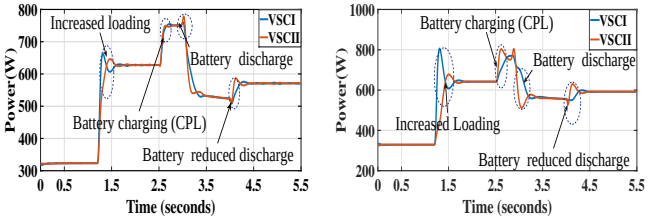
Parameter	Time of operation	Value
$P_1 + P_2$ (W)	0.0 s to 1.3 s	640 W (320 W each)
$P_1 + P_2$ (W)	1.3 s to 2.5 s	1280 W (640 W net change)
$P_1 + P_2$ (W)	2.5 s to 2.8 s	1480 W (200 W net change)
$P_1 + P_2$ (W)	2.8 s to 3.8 s	1020 W (460 W net change)
$P_1 + P_2$ (W)	3.8 s to 5.5 s	1120 W (100 W net change)

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Exeter. Downloaded on May 05,2020 at 15:09:36 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

and again reduces to *0*. The idea is to obtain better power sharing and also to eliminate the low-frequency oscillations with the adaptive term. The variation of α can be seen in Fig. 5. The proposed adaptive control also mitigates the transients better than PI control as can be seen, comparing Fig. 6(a) and Fig. 6(b). The peak overshoot for the proposed control is lesser than that for PI control. Fig. 6(c) shows the frequency profile for the proposed control and Fig. 6(d) for PI control. Similar to the power response, the proposed control mitigate the transients in frequency in a better manner.

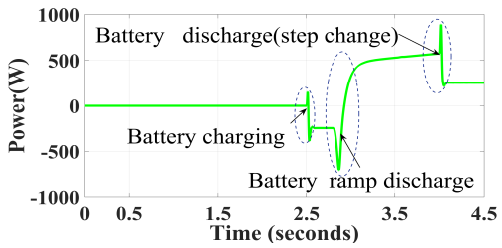
2) CPL transients: A *42V* lead-acid battery connected on the DC side of VSC_3 acts as CPL for the autonomous microgrid. At *t=2.5s*, CPL switching of *200W* (battery charging) occurs in the system. Another CPL switching of *460W* (battery discharging) occurs at *t=2.8s*. Fig. 7 validates the performance of proposed controller mitigating the CPL transients. The performance is validated to be better than the conventional PI control.



(a) Active power(proposed control) (b) Active power(PI)

(c) Frequency(proposed control) (d) Frequency(PI)

Fig. 6: MATLAB/simulink responses for transient events explained in Table III

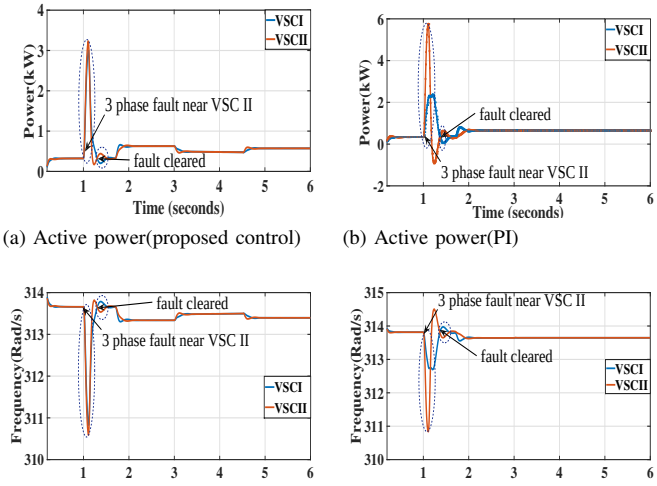


(a) CPL response

Fig. 7: MATLAB/Simulink responses of CPL operation explained in Table III

3) Fault transients: The case of transient mitigation during an AC faults contingency is discussed here. A 3ϕ fault occurs near VSC_2 for *100ms* at *t=1s* and is cleared at *t=1.1s* [see Fig. 8]. The contribution to the fault is dependent on the control employed. It can be seen that power-sharing has a better profile for the proposed control than for PI control. Fig. 8(a) shows both VSCs contributing *P=3.2 kW* whereas Fig. 8(b) shows VSC_2 contributing *P=5.8 kW* alone while VSC_1 contributing around *P=2 kW*. The frequency output of both VSCs are shown in Fig. 8(c) & (d). The proposed controller has better performance in terms of maintaining the same frequency to the distribution system. As it can be seen that during the transient, both VSC maintain the almost identical frequency when the proposed controller is applied. However, with the conventional PI controller VSC, I & VSC II have different frequency response during the transient as the drop in VSC II frequency output is higher than the VSC I frequency output. The settling time for the proposed control is also lesser than that for PI control. It means that for power restoration after a fault isolation, the proposed approach restores the system at *t=1.45s* whereas the conventional PI restores the system at *t=1.92s*.

However, the steady-state error is zero for the case of PI control, whereas the steady-state error is in a bound for the proposed approach. The proposed analysis gives uniform ultimate boundedness of the tracking errors. Hence, the error converges exponentially to a bound around the equilibrium point (zero).



(a) Active power(proposed control) (b) Active power(PI)

(c) Frequency(proposed control) (d) Frequency(PI)

Fig. 8: MATLAB responses for 3 phase ground fault.

V. Experimental Results

A. Experimental setup

The hardware validation of the proposed adaptive backstepping is carried out using an FPGA based NI- PXIe system. The scaled-down assembly for hardware validation is shown in Fig. 9 where the system shown in Fig. 1 has been implemented in a lab scale setup at *110V*. The two VSCs responsible for maintaining the voltage and frequency of the system and also for sharing the load proportionately are shown in the figure as VSC1 and VSC2. For constant power load (CPL), series combination of *12V* lead-acid batteries are connected on the DC side of a VSC through a bi-directional converter. The voltages and currents measurements are taken employing voltage sensors (LV25-P) and current sensors (LA55-P), shown in Fig. 9. Also, the RLC type loads connected to the system is shown as load banks in the figure. FPGA board (NI sbRIO 9606 and

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Exeter. Downloaded on May 05,2020 at 15:09:36 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

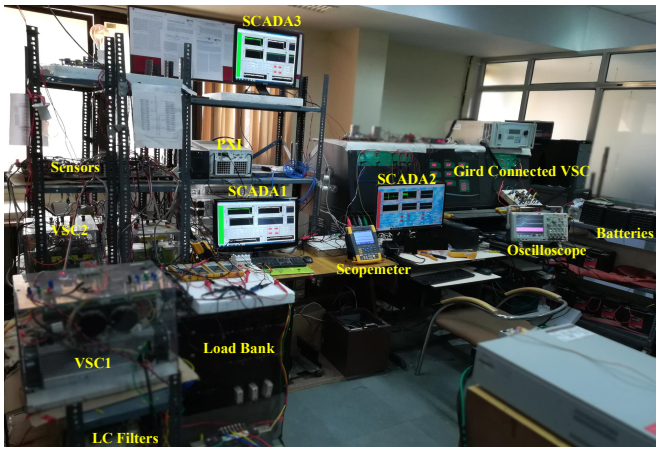
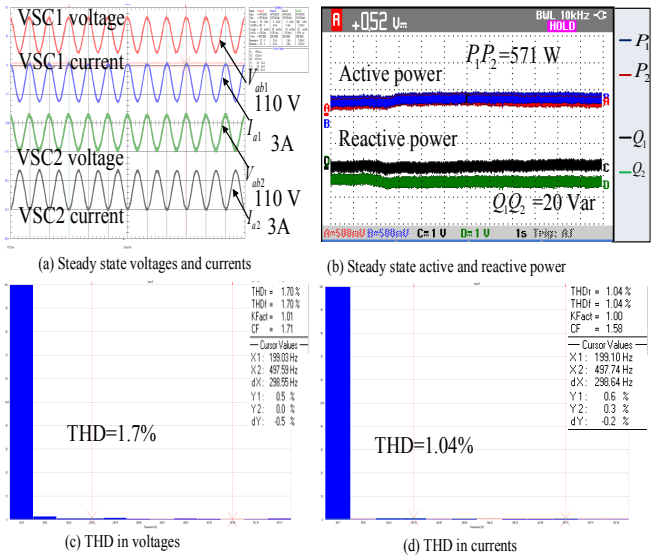


Fig. 9: Hardware setup for experimental validation

9683) is used for data acquisition and processing of signals. NI LabView is employed to design the control algorithms shown as SCADA1, SCADA2 and SCADA3 in Fig. 9.



(a) Steady state voltages and currents

(b) Steady state active and reactive power

(c) THD in voltages

(d) THD in currents

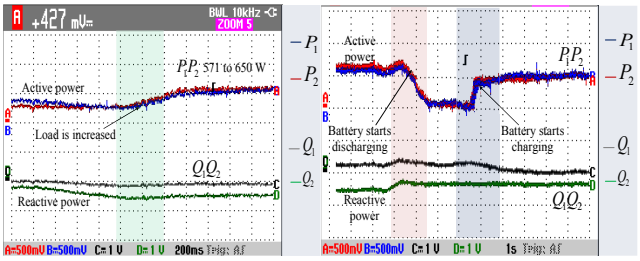
Fig. 10: Experimental Results (steady state)

B. Steady state response (VSC_1 & VSC_2)

Fig. 10(a) shows the voltage and current profile of VSCs running in parallel. Fig. 10(a) includes the VSC_1 and VSC_2 voltages and currents while the system is operating in steady-state. The output voltage of VSCs is at 110 V while both are supplying 3 A current to the loads connected to the system. The active and reactive powers of both VSC combined is shown in Fig. 10(b). The net active power transferred by VSC1 and VSC2 is of 571 W and reactive power transferred is of 20 Var. The VSCs share *571W* as seen in Fig. 10(b). Total Harmonic Distortion (THD) in AC phase voltage for the proposed backstepping control is 1.7% (see Fig. 10(c)) while THD in the current plot is 1.04%. (see Fig. 10(d)).

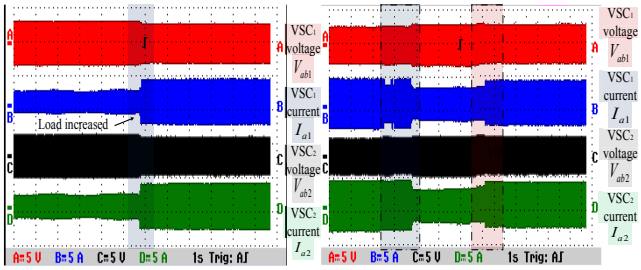
C. Dynamic state response (VSC_1 & VSC_2)

The performance of power-sharing during a load transient (*571W* to *650W*) is studied next. Fig. 11(a) shows the pro-



(a) Active and reactive power sharing in dynamic loading

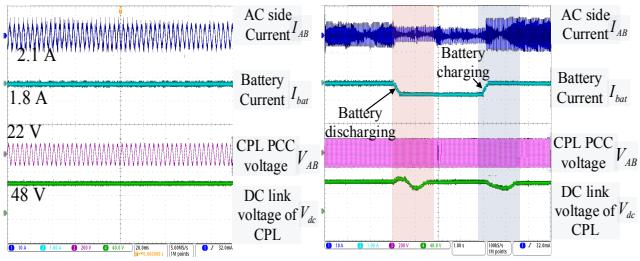
(b) Active and reactive power sharing in CPL loading



(c) Voltages and currents in increased loading

(d) Voltages and currents in CPL loading

Fig. 11: Experimental Results (load changes)



(a) Steady state operation of battery discharging

(b) Battery changed from charging to discharging

Fig. 12: Experimental Results (load changes)

portional power-sharing profile of both VSCs for the load transient. Fig. 11(c) shows the voltage and current profile of VSCs for such an operation. It can be seen that the current by each VSC increases from *3A* to *4.1A*. The loads are increased to make the loading dynamic, and it can be seen that the active power is increased from 571 W to 650 W and similarly the reactive power demand is also increased from 20 Var to 32 Var. However, this increased in active power demand is proportionally shared by both VSCs, which support the claim of the proposed control philosophy.

D. Dynamic state response with CPL (VSC_1, VSC_2 & VSC_3)

CPL operations in the form of battery charging and discharging are studied next. Three lead-acid batteries (*12V, 42Ah* each) in series are interfaced with a bi-directional converter operating in current control mode. The output of the bi-directional converter is connected to the DC link of VSC_3. The DC link of VSC is maintained at *48V*. Initially, battery discharges which reduce the power requirement (*571W* to *450W*) from the VSC set. The battery is made to charge next,

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Exeter. Downloaded on May 05,2020 at 15:09:36 UTC from IEEE Xplore. Restrictions apply.

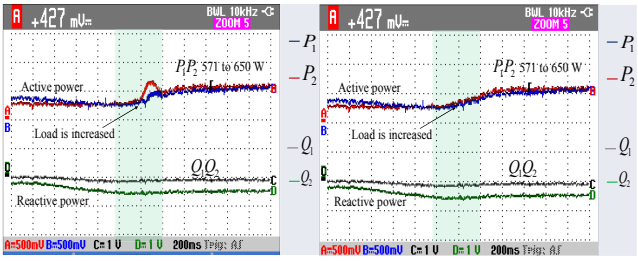
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

increasing power requirement from *450W* to *560W*. Fig. 11(b) shows the contribution of each VSC in the CPL operation. The current contribution drops from *3A* to *1.9A* when the battery starts to discharge. The current contribution starts to increase again from *1.9A* to *2.9A* when the battery starts charging. The sequence of events and the following current contributions are elaborated in Table IV. CPL dynamics has been presented in Fig. 12(a), 12(b). In steady-state, the battery is discharging at *1.8A*. A step-down transformer (*5:1*) is connected between distribution lines and VSC_3. The AC side voltage of VSC_3 is at *22V*.

TABLE IV: Events during real time CPL operation

Parameter	Time of operation	Value
$P_1 + P_2$ (W)	0.0 s to 1.5 s	571 W (285.5W each)(Fig. 9(b))
$P_1 + P_2$ (W)	1.5 s to 4.5 s	450 W (79W net change)(Fig. 9(b))
$P_1 + P_2$ (W)	4.5 s to 6 s	560 W (110W net change)(Fig. 9(b))
$i_{a1}\&i_{a2}$ (W)	0 s to 2.5 s	3A each (Fig. 9(b))
$i_{a1}\&i_{a2}$ (W)	2.5 s to 4.5 s	1.9A each (Fig. 9(b))
$i_{a1}\&i_{a2}$ (W)	4.5 s to 6 s	2.9A each (Fig. 9(b))

E. Comparision between PI based control and proposed control reponse (VSC_1 & VSC_2)



(a) Active power sharing (PI control) (b) Active power sharing (proposed control)

Fig. 13: Experimental Results (comparision between PI and proposed control)

Similar to the simulation results, a comparison between the PI-based control and the proposed control is carried out in real-time. The performance of power-sharing during a load transient (*571W* to *650W*) is studied next. Fig. 13(a) shows the proportional power-sharing profile of both VSCs for the load transient when PI control is applied for both VSCs. The loads are increased to make the loading dynamic, and it can be seen that the active power is increased from 571 W to 650 W. The same steps are carried out for the proposed adaptive droop based control to maintain the uniformity. The active power-sharing dynamics for the proposed control is shown in Fig. 13(b) The proposed adaptive backstepping control mitigates the transients better than PI control as can be seen in Fig. 13, as the peak overshoot for the proposed control, in case of VSC_2 is lesser than that for PI control.

VI. Conclusion

The work proposes an adaptive droop based control for active power-sharing between VSCs in an autonomous microgrid. The adaptive term introduced in the droop control acts as a feed-forward term and is responsible for the transient performance of the system. A Lyapunov based technique has been proposed, which ensures uniform ultimate boundedness (UUB) of the tracking errors. The focus of the work is on transient mitigation for different transients in the system. A thorough comparison of the proposed control with the conventional PI control is presented employing MATLAB/Simulink as the platform. AC load variations, Constant Power Load (CPL) switching, AC faults restoration transients are considered for the comparison. FPGA based hardware setup is employed to study the real-time performance of the proposed adaptive control.

References

[1] IEEE Guide for Design, Operation, and Integration of Distributed Resource Island Systems with Electric Power Systems," in *IEEE Std 1547.4-2011* , vol., no., pp.1-54, 20 July 2011.

[2] Q. Zhong, "Robust Droop Controller for Accurate Proportional Load Sharing Among Inverters Operated in Parallel," in *IEEE Transactions on Industrial Electronics,* vol. 60, no. 4, pp. 1281-1290, April 2013.

[3] J. He and Y. W. Li, "An Enhanced Microgrid Load Demand Sharing Strategy," in *IEEE Transactions on Power Electronics,* vol. 27, no. 9, pp. 3984-3995, Sept. 2012.

[4] R. Han, L. Meng, G. Ferrari-Trecate, E. A. A. Coelho, J. C. Vasquez and J. M. Guerrero, "Containment and Consensus-Based Distributed Coordination Control to Achieve Bounded Voltage and Precise Reactive Power Sharing in Islanded AC Microgrids," in *IEEE Transactions on Industry Applications,* vol. 53, no. 6, pp. 5187-5199, Nov.-Dec. 2017.

[5] T. Wu, Z. Liu, J. Liu, S. Wang and Z. You, "A Unified Virtual Power Decoupling Method for Droop-Controlled Parallel Inverters in Microgrids," in *IEEE Transactions on Power Electronics,* vol. 31, no. 8, pp. 5587-5603, Aug. 2016.

[6] T. Nguyen, H. Yoo and H. Kim, "A Droop Frequency Control for Maintaining Different Frequency Qualities in a Stand-Alone Multimicrogrid System," in *IEEE Transactions on Sustainable Energy,* vol. 9, no. 2, pp. 599-609, April 2018.

[7] G. Lou, W. Gu, Y. Xu, M. Cheng and W. Liu, "Distributed MPC-Based Secondary Voltage Control Scheme for Autonomous Droop-Controlled Microgrids," in *IEEE Transactions on Sustainable Energy,* vol. 8, no. 2, pp. 792-804, April 2017.

[8] X. Huang, K. Wang, J. Qiu, L. Hang, G. Li and X. Wang, "Decentralized Control of Multi-Parallel Grid-Forming DGs in Islanded Microgrids for Enhanced Transient Performance," in *IEEE Access,* vol. 7, pp. 17958-17968, 2019.

[9] D. E. Olivares et al., "Trends in Microgrid Control," in *IEEE Transactions on Smart Grid,* vol. 5, no. 4, pp. 1905-1919, July 2014.

[10] H. R. Baghaee, M. Mirsalim, and G. B. Gharehpetian, Real-time verification of new controller to improve small/large-signal stability and fault ridethrough capability of multi-DER microgrids, *IET Gener. Transmiss. Distrib.,* vol. 10, no. 12, pp. 30683084, Sep. 2016.

[11] H. R. Baghaee, M. Mirsalim, G. B. Gharehpetian, and H. A. Talebi, A generalized descriptor-system robust H control of autonomous microgrids to improve small and large signal stability considering communication delays and load nonlinearities, *Int. J. Elect. Power Energy Syst.,* vol. 92, no. 1, pp. 6382, Nov. 2017.

[12] H. R. Baghaee, M. Mirsalim, G. B. Gharehpetan and H. A. Talebi, "Nonlinear Load Sharing and Voltage Compensation of Microgrids Based on Harmonic Power-Flow Calculations Using Radial Basis Function Neural Networks," in *IEEE Systems Journal,* vol. 12, no. 3, pp. 2749-2759, Sept. 2018.

[13] H. R. Baghaee, M. Mirsalim and G. B. Gharehpetian, "Power Calculation Using RBF Neural Networks to Improve Power Sharing of Hierarchical Control Scheme in Multi-DER Microgrids," in *IEEE Journal of Emerging and Selected Topics in Power Electronics,* vol. 4, no. 4, pp. 1217-1225, Dec. 2016.

[14] H. R. Baghaee, M. Mirsalim, G. B. Gharehpetian and H. A. Talebi, "Unbalanced harmonic power sharing and voltage compensation of microgrids using radial basis function neural network-based harmonic power-flow calculations for distributed and decentralised control structures," in *IET Generation, Transmission & Distribution,* vol. 12, no. 7, pp. 1518-1530, 10 4 2018.

[15] M. B. Delghavi and A. Yazdani, "Sliding-Mode Control of AC Voltages and Currents of Dispatchable Distributed Energy Resources in Master-Slave-Organized Inverter-Based Microgrids," in *IEEE Transactions on Smart Grid,* vol. 10, no. 1, pp. 980-991, Jan. 2019.

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Exeter. Downloaded on May 05,2020 at 15:09:36 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

[16] M. Cucuzzella, G. P. Incremona and A. Ferrara, "Decentralized Sliding Mode Control of Islanded AC Microgrids With Arbitrary Topology," in *IEEE Transactions on Industrial Electronics,* vol. 64, no. 8, pp. 6706-6713, Aug. 2017.

[17] D. O. Amoateng, M. Al Hosani, M. S. Elmoursi, K. Turitsyn and J. L. Kirtley, "Adaptive Voltage and Frequency Control of Islanded Multi-Microgrids," in *IEEE Transactions on Power Systems,* vol. 33, no. 4, pp. 4454-4465, July 2018.

[18] T. Hosseinalizadeh, H. Kebriaei and F. R. Salmasi, "Decentralised robust T-S fuzzy controller for a parallel islanded AC microgrid," in *IET Generation, Transmission & Distribution,* vol. 13, no. 9, pp. 1589-1598, 7 5 2019.

[19] Graeme Heald "Issues with Reliability of Fuzzy Logic" Published in *International Journal of Trend in Scientific Research and Development (ijtsrd)*, ISSN: 2456-6470, Volume-2 — Issue-6, October 2018, pp.829-834,

[20] H. K. Khalil, Nonlinear Systems, 2nd ed., *Englewood Cliffs,* NJ, USA: Prentice-Hall, 1996.

[21] Yaolong Tan, Jie Chang, Hualin Tan and Jun Hu, "Integral backstepping control and experimental implementation for motion system," in Proceedings of the 2000 *IEEE International Conference on Control Applications,* Anchorage, AK, USA, 2000 pp. 367,368,369,370,371,372.

[22] D. Rajicic and A. Bose, "A modification to the fast decoupled power flow for networks with high R/X ratios," in *IEEE Transactions on Power Systems,* vol. 3, no. 2, pp. 743-746, May 1988.

[23] A. Emadi, A. Khaligh, C. H. Rivetta and G. A. Williamson, "Constant power loads and negative impedance instability in automotive systems: definition, modeling, stability, and control of power electronic converters and motor drives," in *IEEE Transactions on Vehicular Technology,* vol. 55, no. 4, pp. 1112-1125, July 2006.

[24] A. Firdaus and S. Mishra, "Mitigation of Power and Frequency Instability to Improve Load Sharing Among Distributed Inverters in Microgrid Systems," in *IEEE Systems Journal,* vol. 14, no. 1, pp. 1024-1033, March 2020.

[25] M. Yazdanian and A. Mehrizi-Sani, "Internal Model-Based Current Control of the RL Filter-Based Voltage-Sourced Converter," in *IEEE Transactions on Energy Conversion,* vol. 29, no. 4, pp. 873-881, Dec. 2014.

[26] P. D. Achlerkar, V. Nougain and B. K. Panigrahi, "Backstepping Technique based Customer-end Voltage Control Strategy of DC Distribution Network," in *IEEE Journal of Emerging and Selected Topics in Power Electronics.*



Surya Prakash received the B.Tech. degree in Electrical Engineering from School of Engineering, CUSAT, Kochi, India in 2009 and the M.Tech. degree in Power System from the NIT Trichy, Tamilnadu, India, in 2015. He has worked as Design Engineer at Hydro Machinery Development Station, BHEL, Bhopal India for 3 years. He is currently working toward the Ph.D. degree at Electrical Engineering Department, Indian Institute of Technology, Delhi, India.

His current research interests include control of grid connected inverter system under varying grid conditions.



Vaibhav Nougain Vaibhav Nougain received B.Tech degree in electrical engineering in 2017 from Delhi Technological University (Formerly Delhi College of Engineering), New Delhi, India. He is currently working toward the doctoral degree at the Department of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi, India. He was a Visiting Student with The Center for Electric Power and Energy (CEE), Technical University of Denmark in 2019. His research interests include protection and control of DC and AC microgrids, islanding detection, and protection of HVDC systems.



Sukumar Mishra (M'97-SM'04) received the B.Sc (Engg) degree in electrical engineering from University College of Engineering Burla, Odisha, India, in 1990, and the M.Sc (Engg) and Ph.D. degrees in electrical engineering from National Institute of Technology Rourkela, Odisha, India, in 1992 and 2000, respectively.

He a Professor with the Indian Institute of Technology Delhi, New Delhi, India, and has been a part of the institute for the past 15 years. He has authored or coauthored over 200 research articles (including papers in international journals, conferences, and book chapters). His research interests include power systems, power quality studies, renewable energy, and smart grid.

Prof. Mishra is a recipient of many accolades throughout his academic tenure of over 25 years. He has been a recipient of Indian National Science Academy Medal for Young Scientist (2002), Indian National Academy of Engineering (INAE) Young Engineer Award (2009), INAE Silver Jubilee Young Engineer Award (2012), and has recently won the Samanta Chandra Shekhar Award (2016). He has been granted fellowship from many prestigious technical societies like IET (U.K.), NASI (India), INAE (India), IETE (India), and IE (India) and is also recognized as the INAE Industry Academic Distinguish Professor. He is currently serving as an Editor for the IEEE TRANSACTIONS ON SUSTAINABLE ENERGY, IEEE TRANSACTIONS ON SMART GRID, and an Associate Editor for the IET Generation, Transmission, & Distribution journal. Currently, he is the Vice Chair of Intelligent System Subcommittee of Power and Energy society (PES) of IEEE. Apart from all research and academic collaborations, he is very actively involved in industrial collaborations. He is currently acting as INAE Chair Professor and was previously delegated as the NTPC and Power Grid Chair Professor. He is also serving as an Independent Director of Cross Border Power Transmission Company Ltd., Gurugram, India, and River Engineering Pvt. Ltd., Greater Noida, India.

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Exeter. Downloaded on May 05,2020 at 15:09:36 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

An Embedding-based Deep Learning Approach for Movie Recommendation

Ram Murti Rawat
Assistant Professor, Department of COE, Delhi Technological University, Delhi, India.

Vikrant Tomar, Vinay Kumar
Undergraduate Student, Department of COE, Delhi Technological University, Delhi, India.

Abstract - **Recommender systems are employed either as tools or algorithms, whose key task is to efficiently predict the ratings for items and recommend items, using the data generated by users. It assists users in finding items that they will like. Hence, recommendation systems are becoming an essential part of some applications, e-commerce websites, and online streaming services, etc. This paper emphasizes on recommendation system for movies whose main objective is to propose a movie recommendation system through a deep learning technique. The size and complication of websites have increased due to the rapid growth of the internet. On these websites, it has become time-consuming and extremely difficult for the users to find the information that they are searching for. Therefore, a collaborative filtering-based movie recommendation system using deep learning and embedding is proposed. The proposed system is evaluated by calculating the RMSE and MAE values. The proposed method is compared with other machine learning algorithms and some state-of-the-art methods. Our model gave an MAE score of 0.7305 and an RMSE score of 0.9311. It performs better than some previous approaches to collaborative-filtering based movie recommendation.**

Keywords – Recommendation system, artificial neural network, collaborative filtering, embedding, machine learning

I. INTRODUCTION

A recommendation system is now a central part of all e-commerce applications and online entertainment services. The recommendation system gains data about the user's preferences for a particular item (e.g. songs, movies, games, music, etc.) in two different ways. An implicit way of obtaining user data involves observing the user's past behavior such as order history, watched movies, and search queries. Another way of obtaining data is done by gathering the user's previous ratings and likes. The size and complication of websites have increased due to the exponential growth of the internet. On these websites, it has become time-consuming and extremely difficult for the users to find the information that they are searching for. Recommendation systems assist users in finding the information that is within their interests by some interaction with the user. A Recommendation system predicts the rating for items that the user has not rated yet based on his previous ratings and ratings by similar users. Recommendation systems have been a field of research for decades now, but the curiosity remains high because of the plenty of practical applications and the problem abundant domain. Recommender systems deliver tailored information by the familiarity of the user's interests by interaction or experience with that user. As online streaming services are becoming a part of our lives, the recommendation systems have gained importance. With this in mind, our objective is to work on a collaborative filtering-based deep learning strategy for the movie recommendation system. Collaborative filtering is an approach for recommendation systems that depends on the ratings given by a particular user along with the ratings for items given by other similar users. New items of interest can be recommended to the users if the ratings are precisely predicted for an item that the user has not yet rated. In this way, movies are recommended that are different than the user's previously watched movies. Whereas, a Content-Based Recommender only recommends movies based on the similarity between items. The basic idea behind content-based filtering is that if someone likes an item, then he or they will also like a "similar" item. It usually works well when it's easy to determine the properties of each item. Collaborative filtering helps users find diverse items if there are users with similar preferences. Different ways of performing collaborative filtering are present and have been implemented. One way is to create a user profile based on user's ratings and watch history and then find similar users using the k-nearest neighbor algorithm. Similarity can be defined in various ways between the users. The user profile can be defined as a vector of continuous values and then use some comparison measures like cosine similarity. Matrix factorization is one other known method for performing collaborative filtering. In this algorithm, a user-item matrix is composed, then it is factorized into two different matrices such that the inner dimension of the matrices denotes some latent factors. The representation of these factors in the above factorization helps recommend new unique items to the users. Now, deep learning is proving to be quite accurate in recommendation tasks. The popularity increase of deep learning in recommendation systems is due to its state-of-the-art performance and highly accurate predictions. Deep learning better represents the user's interest, features of the item, and past interactions between them as compared to other previous approaches to recommendation systems. A movie recommendation system using the deep learning technique and embedding layers are used.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:56:24 UTC from IEEE Xplore. Restrictions apply.

II. Related Work

Recommender systems are classified into three different techniques that are content-based [5], [6], collaborative filtering [7], [9]-[10], [20], hybrid [13], [14]. Also, clustering algorithms are used for recommendation purposes as it helps cluster similar users into groups. In a movie recommendation model, the system learns the previous user ratings then these ratings help the system to recommend news movies to users using different approaches. Collaborative filtering [7], [9]-[10], [20] is becoming extremely important as this technique is used by many of the recommender systems. Hybrid approaches use the combination of both the content and collaborative-based filtering techniques. Memory-based and model-based are the two main types of collaborative filtering techniques. Memory-based collaborative filtering [4], [16] discoveries similar users in the user space corresponding to a specific user and then recommend movies to that user dynamically. There are two problems related to this approach that are high computational complexity and sparsity of data. An item-based method for collaborative filtering was proposed by different authors to resolve these issues. They calculated the similarity amongst the nearby region items about a fixed entity. This usually uses the k-nearest neighbor algorithm to find groups of similar items. This technique reduced the time complexity to some extent and made accurate predictions. Model-based collaborative filtering [3], [8] approach prepares a model in advance based on the previous item ratings by the user and then the model delivers item recommendation. The model is prepared using different machine learning algorithms. This approach is scalable and solves data sparsity, but time complexity is higher than other techniques. The most general method of achieving collaborative filtering is to use a k-nearest-neighbor approach between users [7]. Clustering also helps in finding similar groups in userspace. K-Means clustering is the most common clustering algorithm. An effective collaborative movie recommendation system was proposed which uses a hybrid K-Means clustering with a Cuckoo search algorithm for optimization [1]. There is another popularly known technique matrix factorization for performing collaborative filtering [19]. Deep learning has reformed various fields of computer engineering, and now it is being extensively used in recommendation systems [2], [15], [20]. Wang et al. [20] built a collaborative model for recommendation using deep learning which works both on content data and rating matrix. A recommendation system is proposed by Elkahky et al. [15] using deep learning which uses search queries and web browsing history provided by users. But there is a limitation of this method, as it needs history and search queries which are not accessible all the time. Wei et al. [21] proposed a movie recommendation model using deep neural networks for prediction of movie ratings by mining the characteristics of items and is particularly used to cold start items. A deep learning technique for collaborative filtering based on autoencoders precisely predicts the movie ratings [2]. There is another extraordinary approach to applying collaborative filtering known as matrix completion (MC). In this technique, the rating matrix is approximated to a low-rank rating matrix. This is a mathematically complex approach. In 2014, Kalofolias et al [23] introduced a regularization of the MC model known as the geometric matrix completion (GMC) model by adding additional information in the form of user and item graphs. In [11], a more efficient graph-regularized method for matrix completion was proposed using alternating least squares optimization method (GRALS). More understandable and simpler approach with Embedding layers and deep neural networks is proposed.

III. Proposed Work

A. Dataset Description

MovieLens 100K data set is used in the proposed model. The MovieLens data sets were collected by the GroupLens Research Project at the University of Minnesota [22]. This data set contains 100,000 ratings in the range of 1 to 5 for 1682 movies from a user base of 943. The data set has been preprocessed and includes only those users who have rated more than 20 different movies. It also contains simple demographic information of users (age, gender, occupation, zip) and information of movie genres. 80% of the dataset will be used for training and rest 20% will be used for testing. The data of users who did not have complete information were removed. **u.data** file is the main file used in building the model. It contains the rating data from a user for a particular movie. It has 4 columns UserID, MovieID, Rating, Timestamp.

The distribution of ratings in range 1 to 5 is shown in Fig 1.

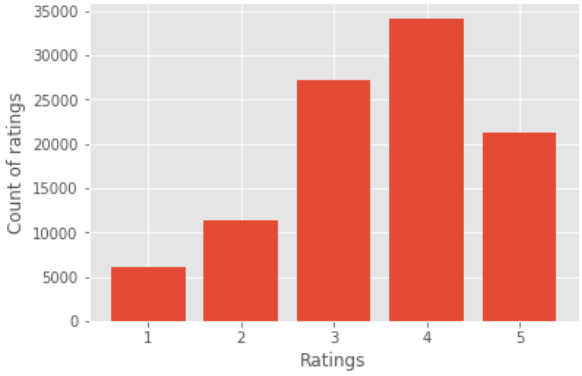


Fig. 1. Distribution of ratings

B. The architecture of the Model

After observing and pre-processing the dataset, 943 unique users and 1,682 unique movies is detected in the MovieLens dataset. However, the users and movie IDs are passed directly into our deep learning network because the model will try to find the relation between IDs that are not present. Our algorithm uses the embeddings to convert these IDs to a continuous vector of fixed size as machine learning algorithms

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:56:24 UTC from IEEE Xplore. Restrictions apply.

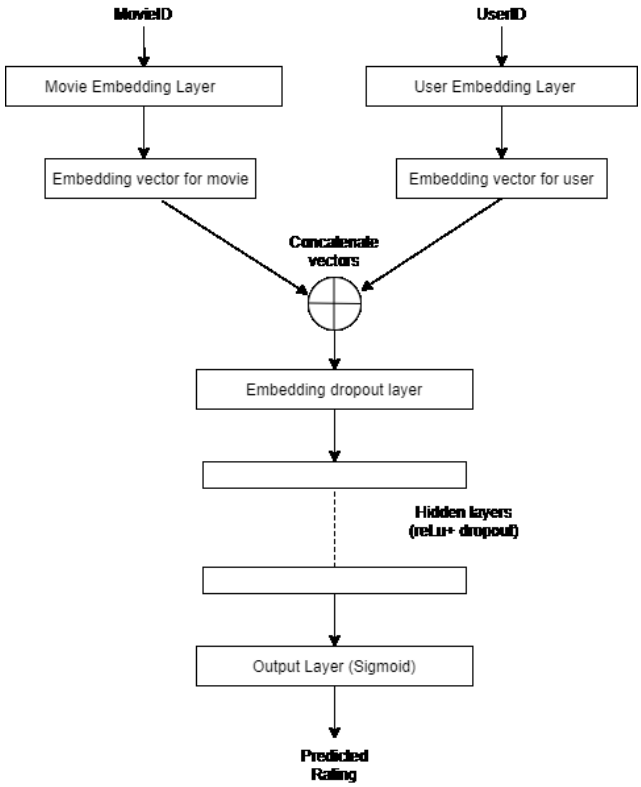


Fig. 2. Structure of network

expects numerical values as input. An embedding vector will be produced for a userID and one embedding vector for a movieID from the embedding space of 943 and 1682 respectively. Then, the vectors are concatenated into one and then passed to the dropout layer. The dropout layer is used to avoid the over fitting of a model. In each training epoch, dropout randomly sets outgoing edges of the layer to 0 with some predefined probability. Next, several fully-connected hidden layers are added with a dropout probability and reLU activation function. Each hidden layer can have a different value of output features and dropout probability. The input features are decided based on the output features of the previous layer. Finally, the predicted rating should be sent as an output for the input user and model IDs. For this, an output layer with a sigmoid activation function is employed and later rescale in the range of 1 to 5. The weights of the neurons in each layer are updated in each training epoch with a back propagation algorithm. The structure of our proposed network used as our model for predicting ratings is as shown in Fig.2.

C. Multi-Layer Perceptron (MLP)

Multi-Layer Perceptron (MLP) is a category of artificial neural networks (ANN). ANN a technique encouraged by the biological neuron network that makes up the human brain. This network learns to accomplish tasks by examining examples, without being explicitly programmed. The MLP is made up of several layers (an input layer, an output layer, and one or more hidden layers) that contains nodes with different activation functions. MLPs are fully connected, which means each node in a layer is connected to every node of the immediate next layer. The input layer receives the input data. No information processing is done on this layer. The received input is conveyed to the following layer, the hidden layer. The data directed from the input layer is processed in the hidden layers and passed to the next layer i.e. output layer. In MLP, many artificial neurons and hidden layers can be obtained. MLP can solve more complex tasks by adding a large number of neurons and layers for processing. However, it will increase the computation time. The output layer is the final layer, every node of the last hidden layer is connected to the final output layer, and hence, the result of the processed data from the hidden layer is produced at this layer. Fig. 3 shows a simple multi-layer perceptron. The advantages of MLP are:

1) It can solve highly complex and complicated problems.
2) It can handle missing data.
3) It learns the weights on its own using the backpropagation algorithm. Thus, it solves a wider range of problems than the algorithms.

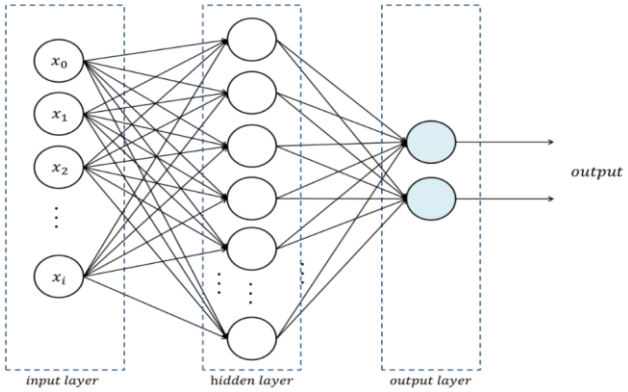


Fig. 3. Simple Multi-Layer Perceptron Model

D. Embeddings

An embedding is a conversion of discrete or categorical variables into a vector of continuous numerical values. In the view of deep learning, embeddings are multi-dimensional vector with learned continuous values that represent discrete or categorical variables. The embeddings are used in reducing the dimension of categorical variables and conveniently represent categories in a transformed space that can be used as input to neural networks and other machine learning algorithms. Embeddings have 3 basic purposes:

1. Finding nearest neighbors in the embedding space. It can be used in making recommendations based on item properties and user interests.
2. It can be used as input to various machine learning algorithms for a supervised learning problem.
3. Used in visualizing of relations between categorical variables and their characteristics.

Neural network embeddings overcome the problems of another simple method for transforming categorical variables, that is, one-hot encoding.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:56:24 UTC from IEEE Xplore. Restrictions apply.

E. Root Mena Square Error

Root Mean Square Error (RMSE) is a common evaluation metric used in calculating the accuracy of a system. RMSE is calculated by taking the square root of the mean squared errors. The mean value is calculated by dividing the sum of the squared difference between actual and predicted output by the number of samples. Equation (1) represents the RMSE value.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=0}^{N}(predicted_i - actual_i)^2}{N}} \quad (1)$$

N is the total number of instances. The smaller RMSE value means better performance for the system. Thus, a system which produces smaller RMSE value will provide better recommendations.

F. MAE

Mean Absolute error (MAE) is a similar evaluation metrics as RMSE. It is calculated by the mean of absolute error values. Absolute error is the positive value of the difference between the predicted output and the actual output. The equation for MAE is mentioned in (2).

$$\text{MAE} = \frac{\sum_{i=0}^{N}|predicted_i - actual_i|}{N} \quad (2)$$

IV. Experimental Result & Analysis

First, some pre-processing of the data is done on the dataset for the desired input for machine learning algorithms. The dataset is divided into two parts in which 80% is for training data and the rest 20% is test data for verification of the result. Our network has 4 basic parameters. To check the performance of our recommender system model, the RMSE and MAE is calculated corresponding to various values of these parameters.

Table 1. RMSE and MAE value for different parameter values.

k	hidden	emb_dropout	Dropouts	MAE	RMSE
100	[300, 200, 100]	0.02	[0.25, 0.25, 0.25]	0.7375	0.9371
100	[200, 100]	0.01	[0.15, 0.15]	0.7324	0.9321
75	[100, 200]	0.02	[0.25, 0.25]	0.7349	0.9353
50	[200, 300, 400]	0.02	[0.20, 0.20, 0.20]	0.7374	0.9355
50	[400, 300]	0.02	[0,25, 0.25, 0.25]	0.7319	0.9315
125	[100, 200]	0.02	[0.20, 0.25]	0.7337	0.9340
125	**[500, 400, 300]**	**0.01**	**[0.15, 0.15, 0.15]**	**0.7305**	**0.9311**
200	[300, 200, 300]	0.02	[0.25, 0.25, 0.20]	0.7353	0.9356
150	[200, 300, 200]	0.02	[0.25, 0.25, 0.25]	0.7402	0.9345

The size of embedding vectors for each unique user and movie is denoted by k. The hidden parameter is a list of values each corresponds to the number of output features of the hidden layers. The emp_dropout represents the dropout probability of the embedding dropout layer. The dropout parameter is a list of dropout probabilities for each hidden layer. Table 1 shows the RMSE and MAE for different parameter values.

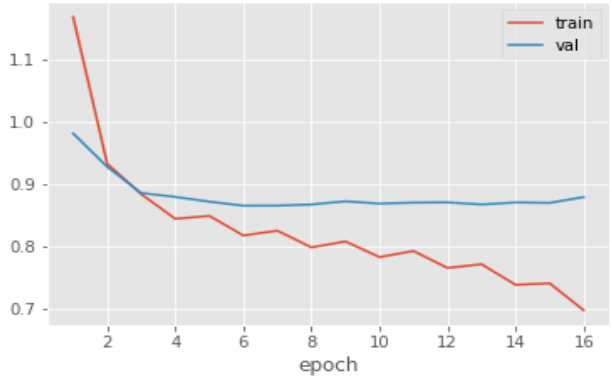


Fig 4. Training and Validation Error vs the number of Epoch

Fig. 4 shows how the training and validation errors are changing with increasing training epochs of the model. The training error is decreasing continuously but the validation error first decreases till epoch 6 but almost stays the same after that. Table 2 shows the corresponding RMSE and MAE values for a single parameter configuration but only changing k value, i.e. the size of output embedding from the embedding layer. The parameter configuration for the comparison is two hidden layers have [400, 300] output features, emp_dropout value is 0.02, dropout values are [0.25, 0.25].

Table 2. Evaluation metric values for different k.

k	*MAE*	*RMSE*
50	0.7359	0.9315
75	0.7330	0.9313
100	0.7370	0.9339
125	0.7315	0.9308
150	0.7338	0.9313
175	0.7328	0.9324
200	0.7331	0.9306

The proposed algorithm is compared with different algorithms for collaborative filtering. KNN is the K-nearest neighbor algorithm, which is a supervised machine learning algorithm for classification. PCA (Principal Component Analysis) is a technique for dimensionality reduction. A genetic algorithm (GA) is a nature-inspired algorithm. SOM refers to Self-organizing Map which maps a higher dimension input into a lower dimension. Clustering is an unsupervised learning algorithm that groups together similar items. GMC (geometric matrix completion) and GRALS (Graph regularized matrix completion method) are both state-of-the-art methods used for

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:56:24 UTC from IEEE Xplore. Restrictions apply.

collaborative filtering. KMEANS CUCKOO is a combination of K-Means clustering and Cuckoo search optimization technique. EMBEDDING+ANN is our proposed approach which uses Artificial Neural networks and Embedding. Table 3 shows the MAE and RMSE values for different algorithms. Our algorithm performs better than all of the algorithms in terms of RMSE value and has a larger value of MAE than just the KMEANS CUCKOO algorithm. Fig. 5 and 6 show the plot for different algorithms vs RMSE and MAE respectively.

Table 3. RMSE and MAE values for different algorithms

ALGORITHM	MAE	RMSE
KNN	-	1.124
PCA-GAKM	0.94	-
PCA-SOM	0.98	-
SOM-CLUSTER	0.75	-
GAKM-CLUSTER	0.76	-
KMEANS CUCKOO	0.68	1.2363
GMC	-	0.996
GRALS	-	0.945
EMBEDDING + ANN	**0.7305**	**0.9311**

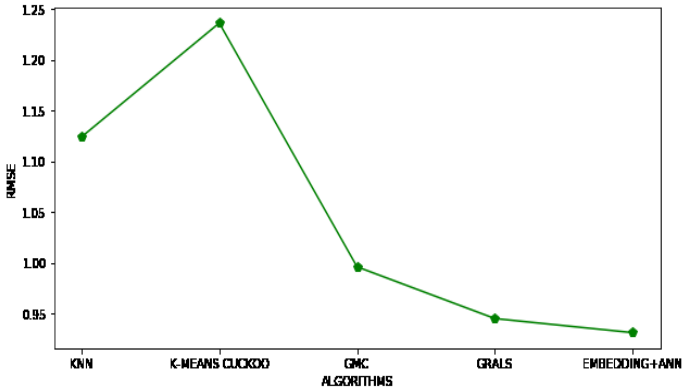


Fig 5. RMSE vs Different Algorithms

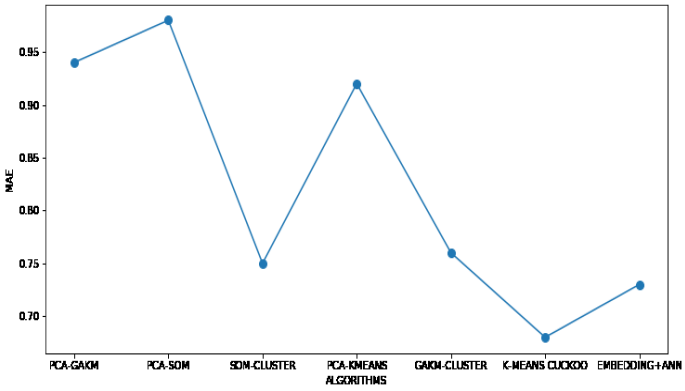


Fig 6. MAE vs Different Algorithms

V. CONCLUSION

In this paper, a collaborative recommender system using an embedding-based deep learning approach is implemented. MovieIDs and UserIDs are compared at the embedding layer to a vector of continuous values of a specific size. Then we add various layers with certain weights for processing of our input data. The hidden layers have a reLu activation function and have a dropout probability for better results. Finally, an output layer with sigmoid activation is added for the prediction of ratings. The movie ratings are predicted for the users in test data and calculated the evaluation metrics based on those predictions. Based on the evaluation metrics, our model gives better results than other machine learning algorithms in terms of MAE and RMSE values. Also, the proposed technique is compared with some state-of-the-art methods and provided more accurate results. While recommending new movies to a user, the rating based on the model is predicted and recommend the movie with higher ratings to the user. The deep learning technique with embedding layers has an additional advantage of scalability. In the future, a hybrid model will be built to give better recommendations. A combination of various machine learning algorithms can help make the recommendation system more efficient and accurate. Ensemble learning is also one of the techniques that will exploit in movie recommendation systems.

REFERENCES

[1] Katarya R, Verma OP. An effective collaborative movie recommender system with cuckoo Search. Egyptian Informatics Journal 18 (2017) 105–112.
[2] Jeffrey Lund, Yiu-Kai Ng. Movie Recommendation using deep learning approach. IEEE Computer Society, 2018.
[3] Lu J, Mao M, Wu D, Zhang G, Wang W. Recommender system application developments: a survey. Decision Support Sys 2015; 74:12–32.
[4] Gutiérrez a, Bobadilla J, Hernando a, Ortega F. Recommender systems survey. Knowledge Based System 2013; 46:109–32.
[5] Semeraro G, De Gemmis M, Lops P. Content-based recommender systems: state of the art and trends. Recommender System Handbook 2011:1–33.
[6] Billsus D, Pazzani MJ. Content-based recommendation systems 2007:325–41.
[7] W. Croft, T. Strohman, and D. Metzler. Search Engines: Information Retrieval in Practice. Addison Wesley, 2010.
[8] Zhang R, Wang Y, Bao H, Liu X, Sun H. Recommender systems based on ranking performance optimization. Front Computer Sci China 2015;1–11.
[9] Su X, Khoshgoftaar TM. A survey of collaborative filtering techniques. Adv Artificial Intelligence 2009; 2009:1–19.
[10] Zeng D, Huang Z, Chen H. A comparison of collaborative-filtering algorithms for e-commerce. IEEE Intelligent Systems 2007.
[11] Nikhil Rao, Inderjit S. Dhillon, Pradeep K. Ravikumar, Hsiang-Fu Yu. Collaborative filtering with graph information: Consistency and scalable methods. In C. Cortes, R. Garnett, N. D. Lawrence, M. Sugiyama, D. D. Lee, and editors, Advances in Neural Info Processing Syst 28, pages 2107–2115. Curran Associates, Inc., 2015.
[12] Ekstrand MD. Collaborative filtering recommender systems. Found Trends_Human-Comput Interact 2010; 4:81–173.
[13] Burke R. Hybrid web recommender systems. Adapt Web 2007.
[14] Kim BM. Clustering approach for hybrid recommender system. In: Proc IEEE/WIC Int Conf Web Intell (WI 2003). p. 33–8.
[15] A. Y. Song, Elkahky, and X. He. A Multi-View Deep Learning Approach for Cross Domain User Modeling in Recommendation Systems. In WWW, pages 278–288, 2015.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:56:24 UTC from IEEE Xplore. Restrictions apply.

[16] Chen G, Chen L, Wang F. Recommender systems based on user reviews: the state of the art. User Model. User-Adapt Interact 2015; 25:99–154.
[17] Jannach D, Nilashi M, Ithnin N, Bin Ibrahim O. Clustering and regression based multi-criteria collaborative filtering with incremental updates. Inf Sci(Ny) 2014;293:235–50.
[18] Tsapatsoulis N, Georgiou O. Improving the scalability of recommender systems by clustering using genetic algorithms. Lecture Notes Comput Sci 2010; 6352:442–9.
[19] C. Volinsky, R. Bell, Y. Koren. Matrix Factorization Techniques for Recommender Systems. Computer, 42(8):30–37, 2009.
[20] N. Wang, D.-Y. Yeung, H.Wang. Collaborative Deep Learning for Recommender Systems. In KDD, pages 1235–1244, 2015.
[21] J. Wei, K. Chen, J. He, Z. Tang, Y. Zhou. Collaborative Filtering and Deep Learning Based Recommendation System for Cold Start Items. Expert Systems with Applications, 69:29–39, 2017.
[22] F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets: History and Context. ACM Transactions on Interactive Intelligent Systems (TiiS).
[23] Vassilis Kalofolias, Pierre Vandergheynst, Michael Bronstein, Xavier Bresson. Matrix completion on graphs. arXiv preprint arXiv:1408.1717,2014.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:56:24 UTC from IEEE Xplore. Restrictions apply.

The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/1934-8835.htm

Analysis of demand risks for the Indian automotive sector in globally competitive environment

Girish Kumar
Department of Mechanical Engineering, Delhi Technological University, Delhi, India

Rajesh Kr. Singh
Management Development Institute, Gurgaon, India, and

Rishabh Jain, Raman Kain and Naveen
Department of Mechanical Engineering, Delhi Technological University, Delhi, India

Received 15 March 2020
Revised 23 April 2020
2 June 2020
11 June 2020
Accepted 11 June 2020

Abstract

Purpose – The purpose of this study is to understand the different types of risks affecting the demand for the automotive sector in India. The study is further trying to illustrate an approach for analyzing the relative intensities of these risks in the present uncertain business environment.

Design/methodology/approach – Risk on the overall demand is assessed by a combined Bayesian – multi-criteria decision-making approach. Data related to the different factors, affecting their product demand is collected from major automobile firms. Then, weights for these factors are evaluated by applying the analytic hierarchy process approach. Further, these weights are used in the Bayesian analysis network to evaluate the risk intensity for different subgroups, namely, political, economic, social, technological and environmental.

Findings – From the literature and experts' opinion, total 16 risk factors have been finalized and these are further grouped into 5 categories i.e. political, economic, social, technological and environmental. It is observed that the demand for organizations functioning in the automotive sector is more vulnerable to economic risk as compared to other risks considered in the study.

Practical implications – Managers and decision makers of associated organizations can use the proposed framework to assess the demand risks so as to pre-evaluate their demand corresponding to future changes. Factors can be added or removed and importance could be assigned to different risk factors according to the prevailing business environment for an organization or sector. This will also help the organizations to conduct a more effective risk management in an uncertain business environment.

Originality/value – The study will help in better understanding of the various demand risks prevalent in the Indian auto sector. The methodology used, provides a novel approach for assessing the macroeconomic demand risks and can be used by the firms working in the automotive sector. The proposed methodology could be used for assessing supply chain risk or any other business initiative risk. The suggested approach will help managers in devising flexible management techniques so as to mitigate the risk.

Keywords Bayesian analysis, Automotive industry, Demand, Risk modelling, MCDM, Macroeconomics.

Paper type Research paper

Authors would like to thank Editor of the journal and reviewers for valuable suggestions and comments to improve quality and content of this paper.

International Journal of Organizational Analysis
© Emerald Publishing Limited
1934-8835
DOI 10.1108/IJOA-03-2020-2076

1. Introduction

The automobile industry plays a significant role in the economic growth of any country (Ramey and Vine, 2004). According to data released by the Press Information Bureau (2019a, 2019b), Government of India, the automotive sector contributes around 7.1% to India's total gross domestic product (GDP) and is expected to grow to 12% by the year 2026. It also constitutes 49% of the manufacturing GDP and 27% of the industrial GDP. This industry is one of the largest employers and provides about 37 million direct and indirect jobs (Press Information Bureau, 2019a, 2019b). Thus, it could be inferred that the auto-component sector is one of the core contributors to the development of the Indian economy (Singh *et al.*, 2007). However, manufacturers in this industry are subjected to huge variations in the demand of their products, as, the demand is highly cyclical. Inaccurate presumption in the demands can lead to huge losses both directly and indirectly. Thus, manufacturers and dealers must properly assess and create a balance between the demand and supply and carry the optimal mix of products for each period in the business cycle. Therefore, it is very important to conduct a prior risk-based decision-making process so as to evaluate the forces that have an impact on the demand. Decision-making is a process involving interdependence in complex situations, which possess multiple (and generally opposing) objectives and uncertainty concerning the outcomes (Azis, 2005). According to Wu *et al.* (2015), risk-based decision-making has been always important for being competitive in global markets.

In India, the automotive demand began to experience a downfall from the last quarter of the calendar year 2018 and intensified with the passage of every month in 2019. According to the Society of Indian Automobile Manufacturers (SIAM), in 2018–2019, the overall vehicle sales fell by 8%, with total retail sales accounting for 16,82,656 units as compared to 18,21,538 units sold in March 2018. Two-wheeler sales fell by a sharp 17% in 2018-2019 from the previous year (SIAM). All categories saw a "de-growth" since March 2018, which was one of the biggest months for retail sales. This has also affected the automotive dealerships negatively, leading to the closure of about 300 of them across the country. Hundreds of thousands of jobs have also been cut citing slumping sales of cars and motorcycles, with many companies forced to shut down factories for days and axe shifts (ACMA, Automotive Components Manufacturers Association of India, 2020; SIAM, Society of Indian Automobile Manufacturers, 2020). The country's GDP too witnessed a slow-down, one of the main reasons for which is believed to be the decline in the automobile demand (Press Information Bureau, 2019a, 2019b). Through this study, we would like to analyze the factors that affect the auto demand globally and, thus, assess the risk proposed by each of those factors.

In the past, various studies have been done to relate the sales of automobiles to customer choices and vehicle attributes (Chand and Avikal, 2015; Aghdaie and Yousefi, 2011; Donatos and Kioulafas, 1990). Sales forecast of vehicles have also been studied using conventional models (Berkovec, 1985; Geurts and Kelly, 1986; Shahabuddin, 2009; Wang *et al.*, 2011). Some work has also been dedicated to automobile sales assessment using macroeconomic factors (Wetzel and Hoffer, 1982; Muhammad *et al.*, 2012; Zhang *et al.*, 2017). Gupta *et al.* (2019), Chirra and Kumar (2018), Zahraee *et al.* (2018), Mangla *et al.* (2015), are amongst a few researchers, who have carried out work on the supply chain in the auto sector. Their studies were limited to the management of production and the working efficiency of an organization by choosing the best supply chain network. However, they were unable to address the risks, which control the demand for its products, which are external to the direct control of an organization and sometimes even unknown to it. Nauhria *et al.* (2011), have identified the competitive priorities based on the manufacturing system and their relative importance for the Indian automobile industry. Chuang and Zhao (2019), have studied the effect of sales demand on the level of inventories for better dealership. Marx *et al.* (2020) have shown that

how governmental decision and policies have led to booming of the Brazilian automobile manufacturing industry. Study has also been done to improve the productivity of Indian automobile firms using knowledge management (Patil and Warkhedkar, 2016).

It has been observed from the above discussion and the literature review in Section 2, that enough research has not been carried out in the past to relate auto demand with the factors that lie outside the control of an organization (Chand and Avikal, 2015; Nauhria *et al.*, 2011, Patil and Warkhedkar, 2016). Auto sales are either attributed to vehicle features or to a few major macroeconomic factors or to any other practice followed by the organization explicitly, such as the supply chain network (Gupta *et al.*, 2019; Chirra and Kumar, 2018; Mangla *et al.*, 2015). Sales forecasting has been done using previously available data by applying conventional models (Wang *et al.*, 2011; Shahabuddin, 2009; Geurts and Kelly, 1986). In this study, it has been tried to understand and analyze all the factors and the risk emerging from these factors that may either explicitly or implicitly impact the demand for the entire auto-industry rather than just an organization.

This study is trying to answer the following research questions:

RQ1. What are the factors that affect the demand in automotive sector?

RQ2. How to determine the relative weights of these factors?

RQ3. How to assess the impact of overall risk in net demand?

The structure of this paper is as follows: Section 2 i.e. literature review deals with the detailed description of different risk factors affecting the demand in the auto sector and also examines the model to be used in the study. In Section 3, the research methodology is discussed. Section 4 deals with the results and numerical illustration. Section 5 discusses the obtained results, managerial implications and theoretical implications of the study. Finally, Section 6 deals with the limitations of the study and scope of future work.

2. Literature review

In this section, review analysis has been done to extract the risk factors affecting the demand and to identify the most appropriate tool for analyzing these risks. Risk factors have been finalized through literature review and experts' opinion. In the Section 2.1, each of these factors have been described briefly along with appropriate references and examples as applicable. Section 2.2 examines the model used for factor ranking. The significance of using multi-criteria decision-making (MCDM) is also discussed. Moreover, past studies on the use of the analytic hierarchy process (AHP) for ranking of criteria have been also reviewed. Section 2.3 describes the various terms regarding risk assessment and risk management. Consequently, the model to be used for risk assessment in the present study is derived.

2.1 Risk factors for automotive demand

The automotive demand is dependent on various factors. In this section, we have described these factors along with a brief discussion on the same. From the literature on automobile demand factors, and by experts' opinion, 16 risk factors have been finalized. These 16 factors have been grouped into 5 categories, namely, political, economic, social, technological and environmental. These risk factors are summarized in Table A1 given in the Appendix.

2.1.1 Political risk.

2.1.1.1 Tax policies. These policies refer to the tax rates and incentives that governments provide so as to take the control of variation in demands and consumption rates. Taxes generally include the corporate taxes and the indirect taxes that consumers have to pay

while purchasing (e.g. Goods and Services Tax). Tax policies may also include rebates and other forms of incentives that government might provide so as to increase or decrease the consumption rates. It is generally seen that demand rises as soon as the consumer starts to benefit from the policy and vice versa. Study by Pritchard and DeBoer (1995), Alberini and Bareit (2016) and Ott and Andrus (2000), show the change in consumer behaviour for car purchase with tax variation. Customers are always attracted towards the vehicles and ownership schemes that levy lower taxation rates.

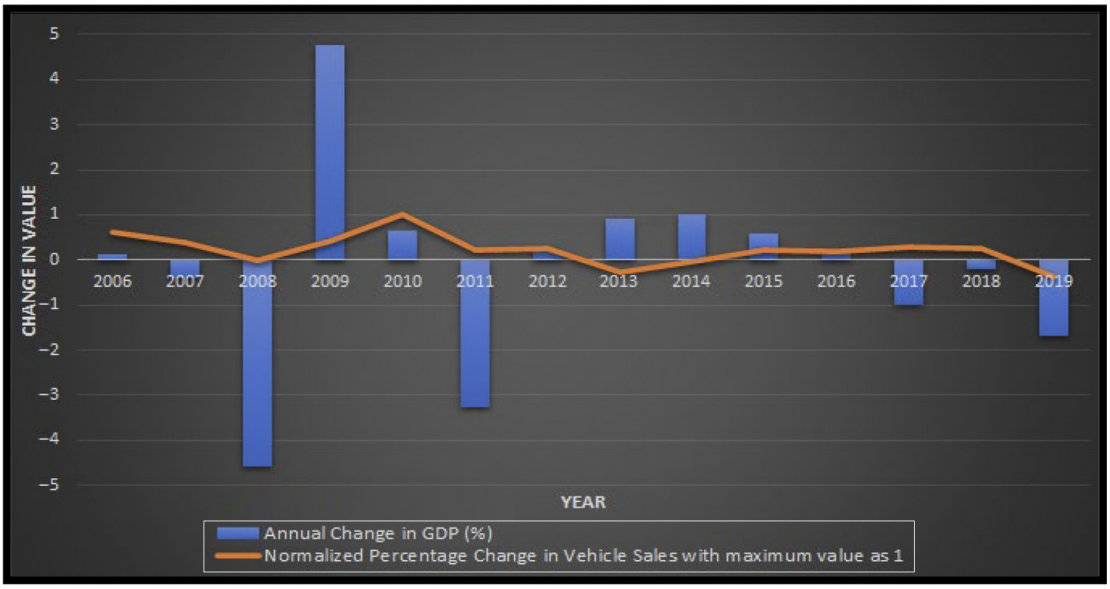
2.1.1.2 Environmental laws. These are the laws, which are implemented by the governments so as to address the problem of rapidly increasing environmental pollution. These may be different in different countries but generally, exist everywhere. These may include rules pertaining to use of vehicles that run on electricity or energy sources other than oil or shifting to modified grades of fuel. Various vehicle emission standards are also set up by the nations. According to the expert reviews demand of vehicles, generally has a negative correlation with the frequency of changes in the environmental laws as, the consumers are confused whether it is the right time to purchase or not. For example, in India one of the major reasons for stagnation in the demand of ic-engine vehicles is attributed to the government's policy to shift to Bharat Stage (BS)-VI (cleaner grade of fuel) from BS-IV by April, 2020 and to completely switch to electric vehicles by 2030 so as to reduce the air pollution.

2.1.1.3 Vehicle regulations. These are the rules and regulations set up by the governments for the manufacture and sale of new vehicles and to ply on the roads. These include some pre-requisites that the vehicle must satisfy such as the axle norms, safety standard norms and emission norms, and those which need to be followed while in use such as, the motor vehicle act. By expert reviews, it is known that sales drop with frequent changes in these rules. According to SIAM, the sales of commercial vehicles dropped by nearly 10% in May 2019, at 68,847 units as compared to 76,517 units in May 2018, which is mainly attributed to the new axle norms. Effect of government policies on the performance of automobile manufacturers was studies by Yadav and Sagar (2015).

2.1.2 Economic risk.

2.1.2.1 Economic growth and development. This imply to the rate at which the increase of per capita income takes place. It also includes, the inflation rate and the tax rates, which indirectly determine the net disposable income. New cars have a high-income elasticity of demand i.e. their demand is highly sensitive to a change in the real income of consumers, keeping all other things constant. When real incomes rise, the demand for new vehicles is expected to expand. Muhammad *et al.* (2012), Shahabuddin (2009) and Dargay (2001), had shown the impact of income rates on the sales of cars. Tambade *et al.* (2019), found that in case of Indian automobile industry business environment plays a very crucial role amongst the other factors. The graph shown in Figure 1 is plotted to simultaneously analyze the percentage change in vehicle sales with respect to the annual change in the GDP of India (Ministry of Road Transport and Highways, Government of India; Press Information Bureau, 2019a, 2019b; The World Bank). It could be clearly noticed that in most cases there exists a strong direct correlation between the two i.e. vehicle sales show a positive sign with an increase in GDP and vice versa.

2.1.2.2 Acquisition and running costs. These include the cost of the vehicle and the costs incurred while running it such as the fuel costs, insurance and maintenance. Hike in these costs lead to a decrease in the number of buyers. High fuel price, though, does not always bring a decline in demand, as a new vehicle would be technologically more advanced and fuel efficient than the older one, thus offering the buyer a chance to save money. Research by Hamilton (1983), Mork (1989), Lee and Ni (2002), Barber *et al.* (1999) and Duncan (1980), show that how significant increase in fuel price could impact car sales.



Source: Ministry of Road Transport and Highways, Government of India; PIB 2019; The World Bank

Figure 1.
Percentage change in vehicle sales w.r.t. annual change in the GDP

2.1.2.3 Inflation rate. Many economists, businessmen and other experts maintain that moderate inflation levels i.e. between 2%-3% are needed to drive the consumption, assuming that higher levels of spending are crucial for economic growth. Inflation, less than this might lead to under development and higher inflation rate may lead to significant rise in price of commodities. As auto industry makes up a good portion of a country's GDP it is necessary to maintain a low inflation rate (Chifurira *et al.*, 2014).

2.1.2.4 Availability of credit or finance. Some of the new customers may purchase a vehicle in one single payment but many mainly depend on securing a finance agreement. Even though interest rates may decrease according to government monetary policy, it still remains hard to get a new loan for a car and also, the rate of interest on unsecured credit keeps on rising. Thus, the high real rate of interest on car loans does not allow new buyers to purchase. Ludvigson (1998), had found that an increase in the rate of interest on loans poses a significant negative impact on car sales. A similar observation was also made by Muhammad *et al.* (2012), in their research on the analyzing the impact of economic variables on automobile sales.

2.1.2.5 Employment rate. Muhammad *et al.* (2012) and Shahabuddin (2009) have studied the effect of rate of employment on the sales of new vehicles. Better rate of employment means, more money in the hands of consumers leading to better consumption rates, and thus, increased demand. This process is cyclical in itself, i.e. decline in employment rate leads to lower demands and lower demand leads to again decline in the employment. Thus, the reason becomes the consequence of the problem. This is valid for the auto sector as well.

2.1.3 Social risk.

2.1.3.1 Mass transit systems. According to a few experts, less availability of good public transport system leads to increased demand of personal vehicles. Similarly, a good transport facility leads to less demand of automobiles. Improvement of public transport and reduction in distance to the nearest railway station leads to less dependence on private vehicles (Rith *et al.*, 2019). A well-planned metro network leads to decrease in private vehicle ownership by 2% ∼ 3% (Mulalic *et al.*, 2016).

2.1.3.2 Cab ridership services. Similar to the mass transit system, better cab facility leads to decreased demand of private vehicles. Consumers, generally prefer riding a cab, rather than driving themselves. Cab services are growing better and better by providing comfortable rides at competitive prices. This has led to a risk for the automobile industry. Rawal and Sisodia (2017), have shown that how the acceptance of cab services have increased in India.

2.1.3.3 Vicinity of roads and availability of parking. Ample of driving and parking space leads to comfort and helps in saving a lot of time. Thus, it encourages consumers to purchase new vehicles. On the other hand, congestion and lack of parking, forces users to avail mass transit and cab services. With rapidly increasing population and speedy urbanization, a well-developed transport system is required as individual vehicles cannot be copped up in such a large volume (Barbero and Galeota, 2019).

2.1.4 Technological risk.

2.1.4.1 Introduction of new vehicles. With introduction of vehicles with new designs and features, it is generally expected that the sales would also rise. Consumers who are pioneer innovators and early adopters are attracted towards new and modular looks. If the vehicles have an upper edge above the past running models, it would generate good reviews and demands would shoot up. Automobile industry is fast changing dynamic market, the growth of which depends upon the competitive technology (Nag and De, 2020; Yadav and Sagar, 2015).

2.1.4.2 Automation. Other than the looks and design, consumers also look for sophisticated and user-friendly functions and automatic features. According to Andrea and Smith (1994) and Aghdaie and Yousefi (2011), customers look for specifications, performance and automation while selecting amongst vehicles.

2.1.4.3 Engine efficiencies. It has been seen in the past that with the use of research and development, new technologies to enhance the fuel efficiency (such as modified fuel supply systems, integrated starter generator, start-stop technology and modified gear ratios) have evolved. Such features act as a huge mode of attraction for the consumers. Customers are easily encouraged to purchase new vehicles if they have better efficiencies than their older ones. Andrea and Smith (1994) and Aghdaie and Yousefi (2011), also showed that customers have higher demands for vehicles with higher fuel efficiencies.

2.1.5 Environmental risk.

2.1.5.1 Pollution levels. With the rapidly changing climate patterns all over the world whose root cause is pollution, it becomes a big risk factor for the automotive business. An automobile, from the start of its manufacturing, throughout its use and till the end of its scrapping, leaves a giant footprint of metals, carbon and other chemically toxic wastes. It leads to rise in all forms of pollution. Manufacturers need to think of ways to mitigate this risk by pre-planning the alternative risk barriers. More movement of people and goods lead to more pollution and emission of greenhouse gases (Barbero and Galeota, 2019).

2.1.5.2 Season and climate. During a certain period of the year the demand may remain very low (such as the rainy season) while during some other time the demand suddenly shoots up (like during the festive season). Vehicle manufactures should be ready to bear such fluctuations in the demand by prior mitigation strategies. Busse *et al.* (2015), found that the purchase of an automobile is highly dependent on the weather, which is controlled by psychological effects, rather than by classical utility theory. Bloesch and Gourio (2015), too suggest that though weather has a temporary, but a significant impact on the economic activities.

2.2 Ranking of the risk factors

Demand pattern of automobiles, as observed from discussion in the previous sub-section, depends upon many different factors. Sorting of these factors according to their relative importance requires the consideration of various evaluation notions. Therefore, it is known as MCDM

problem i.e. decision-making in the presence of multiple and generally conflicting criteria (Rao, 2007). MCDM techniques have been widely used in the studies of risk analysis and management (Chand *et al.*, 2015a, 2015b, 2017). The AHP can be used for the assessment of relative weights for the given factors in a complex problem involving many criteria (Singh, 2013; Chauhan, 2016). The essence of the AHP is that human judgements, and not just the underlying information, can be used in performing the evaluations. The AHP converts these evaluations to numerical values that can be processed and compared over the entire range of the problem. A numerical weight or priority is derived for each element of the hierarchy, allowing diverse and often incommensurable elements to be compared to one another in a rational and consistent way. This capability distinguishes the AHP from other decision-making techniques (Saaty, 1980). Chhabra *et al.* (2017) have used it for analyzing green logistics alternatives in context to an automotive organization. Apak *et al.* (2012) have used AHP to research preferences of a customer for features of luxury cars. Chand and Avikal (2015), have also used AHP in their research to select medium range cars on the basis of different criteria such as cost, performance and fuel consumption Kumar and Singh (2012) have used fuzzy AHP for evaluation of third-party logistic provider. AHP has been used to analyze productivity and wastage reduction for an Indian manufacturing firm (Chauhan, 2016). Very recently, AHP has been used for factor weight calculation of internet of things and big data in customer relationship management (Abu Ghazaleh and Zabadi, 2020). Al Hammadi and Hussain (2019), have used AHP to help decision makers in public health care sector to sort out the factors that affect sustainable organizational performance. It has also been used to prioritize the indicators of managerial success (Ghosh, 2013).

In the classical AHP, a decision maker is asked to supply exact pairwise comparison ratios for each attribute in each level of the hierarchy (Saaty, 1980). However, these decisions might be imprecise due to uncertainty in human opinions in statistical context (Saaty and Vargas, 1987; Moreno-Jimenez and Vargas, 1993; Basak and Saaty, 1993; MacKay *et al.*, 1996). Outcomes of real-life problems are usually not bivalent i.e. there is no clear-cut solution such as "yes" or "no" (Tong and Bonissone, 1980). Zadeh (1965), had introduced the concept of fuzzy set theory (FST) in decision-making, to show the linguistic variables on a numbered scale. The extended form of AHP method was proposed by Chang (1996). It combines the use of FST with the AHP method, i.e. the data used for AHP is a fuzzy number. Thus, when uncertainty exists, FST could be used (Leung and Cao, 2000; Taylan, 2014). Accuracy in assigning weights to some given alternatives could be increased by using fuzzy-MCDM methodology (Bellman and Zadeh, 1970). As the preferences in AHP are usually judgements of human beings based their perception, Leung and Cao (2000) believe that the fuzzy approach allows a more accurate description of the decision-making process. Torfi *et al.* (2010), in their research based on calculating weight using fuzzy AHP have found that AHP is one of the most outstanding MCDM approaches, used to obtain the relative weights. Fuzzy AHP, has also been used in studies regarding supply chain problems (Gupta *et al.*, 2019; Kumar and Rahman, 2017; Yadav and Sharma, 2016; Mangla *et al.*, 2015; Singh and Sharma, 2014). Recently, it has also been used to rank the criteria that determine the stress amongst hospitality and tourism employees (Vatankhah *et al.*, 2020). Thus, in comparison with other MCDM methods, the AHP method has widely been used in similar scenarios and the fuzzy AHP technique, could provide with more accurate results in many practical decision-making problems.

2.3 Risk assessment

Risk implies any form of consequence or the probability of a major loss (Tipu, 2017). Risk analysis is the "systematic use of available information to identify hazards and to estimate the risk to individuals, property and the environment" (IEC 60300-3-9, 1995). Risk evaluation is the:

> [. . .] process in which judgments are made on the tolerability of the risk on the basis of a risk analysis and taking into account, the factors such as, socioeconomic and environmental aspects (IEC 60300-3-9, 1995).

Risk assessment is the "overall process of risk analysis and risk evaluation" (IEC 60300-3-9, 1995). Risk management is a continuous management process with the objective to identify, analyze and assess the potential uncertainty related to an activity and to identify and introduce risk control measures to reduce or completely eliminate the potential harms to the system, as well as the environment (Rausand, 2013; Chand *et al.*, 2014). Framework for risk analysis, evaluation and assessment, given by Rausand (2013), is shown in Figure 2.

Risk analysis has three main parts, namely, to identify the possible undesired events, the likelihood of these events and the consequences of these events (Kaplan and Garrick, 1981). Singh *et al.* (2008) have observed that organizations should develop competencies to evaluate their risks for being competitive in global market. Quantitative risk assessment requires two quantities, namely, the probability of the occurrence of a loss and the magnitude of the potential loss (Ayyub, 2014). According to Rausand (2013), there are three main approaches to probability:

(1) The classical approach;

(2) The frequentist approach; and

(3) The Bayesian or subjective approach.

The first two approaches i.e. the classical and the frequentist approach are used when there is either a fixed sample space of outcomes or when the experiment can be repeated again and again. The Bayesian or the subjective approach is used when these conditions do not exist. In this, the probability is considered to be subjective i.e. there is a numerical value in the interval [0, 1] representing an individuals' degree of belief about the probability of an event to occur.

Thus, the Bayesian network analysis is a good risk assessment methodology that may be used for analyzing qualitatively, quantitatively or both, depending on the scope of the analysis. One of the earliest researches for risks assessment by using Bayesian analysis was done by Pai *et al.* (2003). Bayesian analysis has also been used for optimization of the cost (Narayanan *et al.*, 2005), to conduct diagnostics (Kao *et al.*, 2005; Kauffmann *et al.*, 2002) and to analyze the flexibility in supply chains (Milner and Kouvelis, 2005; Wu, 2005). Nielsen and Jensen (2009) and Alexander (2003), have used the Bayesian network analysis in the business field, for multi-variate and integrated analysis of the risks, for monitoring and evaluating intervention strategies. Bayesian network analysis has also been used to

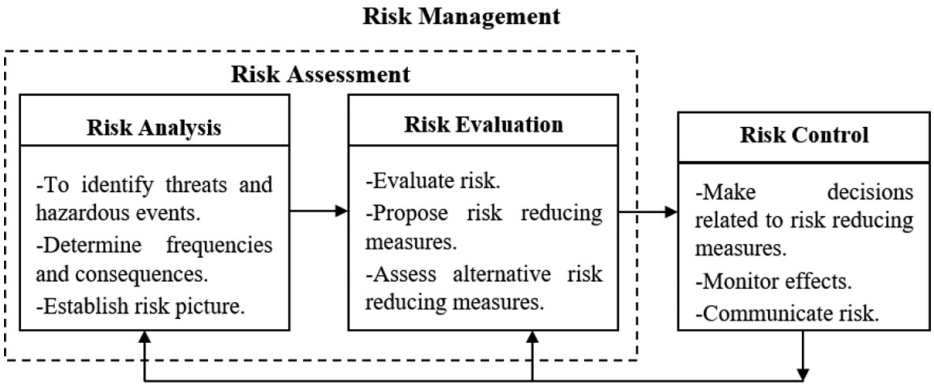


Figure 2.
Risk analysis, evaluation, assessment and management

Source: Rausand (2013)

examine the fluctuations in exchange rates caused by monetary fundamentals (Charfi *et al.*, 2019). Dalla Valle and Giudici (2008), used the Bayesian methods to integrate the scarce and inaccurate quantitative data with prior information provided by the experts to estimate operational loss distributions. In a different research, to assess the enterprise risk, Bonafede and Giudici (2007), had shown that the Bayesian networks are a useful tool to integrate different information and to study risk distribution by using the data collected from the experts. Salehi *et al.* (2020), have used the Bayesian approach to analyze the predictability of the firm level data for determining macroeconomic factors, and suggest that it is a good method, which provides results similar to the classical theory.

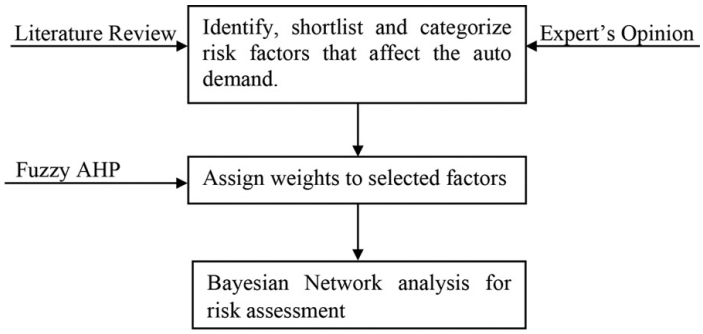
Through the above literature analysis, three important aspects regarding the significance of the present study have been drawn. Firstly, various studies have been presented to illustrate the different factors on which demand of the automotive sector depends. Different authors have worked upon different risk factors as presented in Table A2. Secondly, the application of fuzzy AHP technique has been also reviewed. Finally, to assess the probability of the risk factors under such a scenario, the Bayesian network analysis application has been studied. In an industry working under practical situations, the risk for demand may not always depend only on a few factors in a selected field, rather it is often diversified and interconnected over a vast network of factors (Kumar *et al.*, 2019). In this study, the risk through all the factors has been assessed by a combined Bayesian- MCDM model. Factors have been collected through Section 2.1, factor weights have assigned by using the fuzzy AHP technique and risk probability has been assessed using the Bayesian network analysis.

3. Research methodology

The factors affecting the sales of automotive have been identified and discussed in the final section. The methodology includes, application of fuzzy AHP and Bayesian analysis. For applying AHP, level of importance for each of the demand factor is required. This was achieved by expert survey of the decision makers of four different firms. Profile of experts is given in Table A3 in the Appendix. Thus, the obtained level is based on personal opinions and thoughts supported with some historical data, which makes the use of FST necessary. Relative weights for each of the factor is obtained by the importance level in fuzzy AHP, further the weights obtained by fuzzy AHP are used to assess the relative risk for each factor and its subgroup through the Bayesian analysis. Figure 3 shows the flowchart of the applied research methodology in this study.

3.1 Fuzzy set theory

Definition 1. A fuzzy number ($\tilde{\tilde{n}}$), which is triangular in nature, is denoted by triplet Tr (j,k,l), where [j-lower, k-middle and l-upper]. In terms of the membership function ($\mu_{\tilde{\tilde{n}}}$), the graphical representation is shown in Figure 4.



Source: Self-created

Figure 3.
Flowchart of the applied methodology

$$\mu_{\tilde{\text{n}}} = \begin{cases} \dfrac{x-j}{k-j} & ;\text{for } \text{j} \le \text{x} \le \text{k} \\ \dfrac{x-l}{l-k} & ;\text{for } \text{k} \le \text{x} \le \text{l} \\ 0 & ;\text{otherwise} \end{cases} \tag{1}$$

Definition 2. Let two fuzzy-triangular numbers be, $\tilde{\tilde{\text{x}}}_1 = (\text{j}_1, \text{k}_1, \text{l}_1)$ and $\tilde{\tilde{\text{x}}}_2 = (\text{j}_2, \text{k}_2, \text{l}_2)$, then the mathematical operations associated with these fuzzy numbers are given as following:

- "adding" two Fuzzy Tr. Numbers:

$$\begin{aligned} \tilde{\tilde{\text{x}}}_1 \oplus \tilde{\tilde{\text{x}}}_2 &= (\text{j}_1, \text{k}_1, \text{l}_1) \oplus (\text{j}_2, \text{k}_2, \text{l}_2) \\ &= \left[(\text{j}_1 + \text{j}_2), (\text{k}_1 + \text{k}_2), (\text{l}_1 + \text{l}_2)\right. \end{aligned} \tag{2}$$

- "subtracting" two Fuzzy Tr. Numbers:

$$\begin{aligned} \tilde{\tilde{\text{x}}}_1 \ominus \tilde{\tilde{\text{x}}}_2 &= (\text{j}_1, \text{k}_1, \text{l}_1) \ominus (\text{j}_2, \text{k}_2, \text{l}_2) \\ &= \left[(\text{j}_1 - \text{j}_2), (\text{k}_1 - \text{k}_2), (\text{l}_1 - \text{l}_2)\right. \end{aligned} \tag{3}$$

- "multiplication" of two Fuzzy Tr. Numbers:

$$\begin{aligned} \tilde{\tilde{\text{x}}}_1 \otimes \tilde{\tilde{\text{x}}}_2 &= (\text{j}_1, \text{k}_1, \text{l}_1) \otimes (\text{j}_2, \text{k}_2, \text{l}_2) \\ &= \left[(\text{j}_1 \times \text{j}_2), (\text{k}_1 \times \text{k}_2), (\text{l}_1 \times \text{l}_2)\right. \end{aligned} \tag{4}$$

- "dividing" two Fuzzy Tr. Numbers:

$$\begin{aligned} \tilde{\tilde{\text{x}}}_1 / \tilde{\tilde{\text{x}}}_2 &= (\text{j}_1, \text{k}_1, \text{l}_1) / (\text{j}_2, \text{k}_2, \text{l}_2) \\ &= \left[(\text{j}_1 / \text{j}_2), (\text{k}_1 / \text{k}_2), (\text{l}_1 / \text{l}_2)\right. \end{aligned} \tag{5}$$

Definition 3. Distance between two fuzzy numbers:

Chen (2000), gave the vertex method for finding the distance between to fuzzy Tr. Numbers.

$$\text{d}\left(\tilde{\tilde{\text{x}}}_1, \tilde{\tilde{\text{x}}}_2\right) = \left\{\frac{1}{3}\left[(\text{j}_1 - \text{j}_2)^2 + (\text{k}_1 - \text{k}_2)^2 + (\text{l}_1 - \text{l}_2)^2\right]\right\}^{1/2} \tag{6}$$

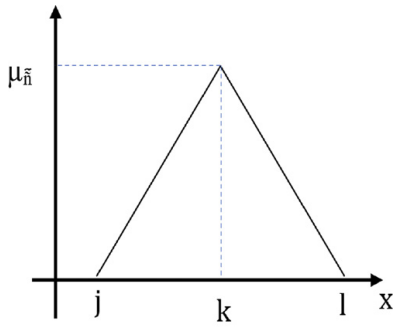


Figure 4.
Depiction of Fuzzy-Triangular number $\tilde{\tilde{x}} = (\text{j}, \text{k}, \text{l})$

Source: Rao (2007)

Definition 4. De-fuzzifying the fuzzified values:

Let $\tilde{\tilde{x}} = (j, k, l)$ represent a fuzzy number. The fuzzy number can defuzzied by the centre of gravity method given in equation (6) Jager *et al.* (1992).

$$\text{De fuzzify}\left(\tilde{\tilde{y}}_i\right) = \frac{1}{3}[j + k + l] \tag{7}$$

3.2 Fuzzy analytic hierarchy process method

As discussed in the literature review, in this study, the AHP technique is applied in its extended form, to determine the weights of the concerned factors. Thus, the geometric mean method of fuzzy AHP is used and the detailed procedure for the same is given in this section. The conventional basic AHP method is not capable to handle the imprecision of human judgments. While, fuzzy AHP, an improved form of AHP is able to handle this issue. The scale used for this analysis is given in Table 1.

In the AHP method, factor importance is first obtained in linguistic terms from decision makers from different firms or organizations based on their level of importance. Then, an aggregated pairwise comparison matrix is constructed amongst the considered factors for their weight calculation. The level of importance obtained from the expert review are converted into triangular fuzzy numbers (TFNs) as per the Table 1.

The complete process of fuzzy AHP, used in this study is explained in five steps:

Let the TFN is represented by:

$$\tilde{\tilde{x}}_i^r = (j_r, k_r, l_r), \text{ where, } r = (1, 2, 3, \ldots., m)$$

Step 1: Calculate the fuzzy geometric mean value, $\tilde{\tilde{r}}$,

$$\tilde{\tilde{r}}_i = \left((j_{i1} \times j_{i2} \times \ldots. \times j_{im})^{1/m}, (k_{i1} \times k_{i2} \times \ldots. \times k_{im})^{1/m}, (l_{i1} \times l_{i2} \times \ldots. \times l_{im})^{1/m}\right. \tag{8}$$

Step 2: Calculate the sum of all the fuzzy geometric mean values,

$$\sum_{i=1}^{m} \tilde{\tilde{r}}_I = [r_1 \oplus r_2 \oplus r_b \oplus \ldots\ldots. \oplus r_m] \tag{9}$$

Step 3: Calculate the fuzzy weights, $\tilde{\tilde{w}}_I$, for all the fuzzy geometric mean values,

$$\tilde{\tilde{w}}_i = r_i \otimes [r_1 \oplus r_2 \oplus r_b \oplus \ldots\ldots. \oplus r_m]^{-1} \tag{10}$$

Step 4: Defuzzify the fuzzy weights to obtain weights, w_i,

$$w_i = \frac{1}{3}[j_{w1} + k_{w2} + l_{w3}] \tag{11}$$

Level of importance	Triangular fuzzy number
1	(1, 1, 1)
2	(1, 2, 3)
3	(2, 3, 4)
4	(3, 4, 5)
5	(4, 5, 6)

Source: Kannan *et al.* (2013)

Table 1. Triangular fuzzy number of a given level of importance

Step 5: Obtain the normalized weights, (w_{ni}) by following,

$$w_{ni} = w_i / \sum w \tag{12}$$

3.3 Bayesian analysis

A Bayesian network graphically represents the probability distribution for a large set of variables (Pai *et al.*, 2003). In a Bayesian network, nodes depict random variables. Nodes that are not connected imply that those variables are independent of each other (Heckerman, 1996). Each node, associated with a probability function takes a particular set of values as input and gives the probability of the variable represented by the node as output (Heckerman, 1996; Nielsen and Jensen, 2009).

The Bayes' Theorem was given by Thomas Bayes in the late 1700*s* who worked on the concept of conditional probability. It states that:

$$P(H|E,r) = \frac{P(H\ |r) \times P(E|H,r)}{P(E|r)} \tag{13}$$

The equation, i.e. P(H|E,*r*) represents the probability of any hypothesis H after considering the occurrence of event E on any past event r. The term P (H|*r*) is the probability of occurrence of H given that only *r* occurs. The likelihood is represented by the term P (E|H,*r*), which gives the probability of the E assuming H and *r* are true. The term, P(E|*r*) is independent of H and is regarded as a normalizing or scaling factor (Niedermayer, 2008). Thus, we see that the Bayesian network analysis could be used for combining subjective a hypothesis with the available evidence. A combined diagram for the AHP and Bayesian Network Analysis is presented in Figure 5.

4. Results and analysis

Results are analyzed in two parts. First, we obtain the relative weights for each of the factors by applying the fuzzy AHP MCDM technique on the data received from industry experts.

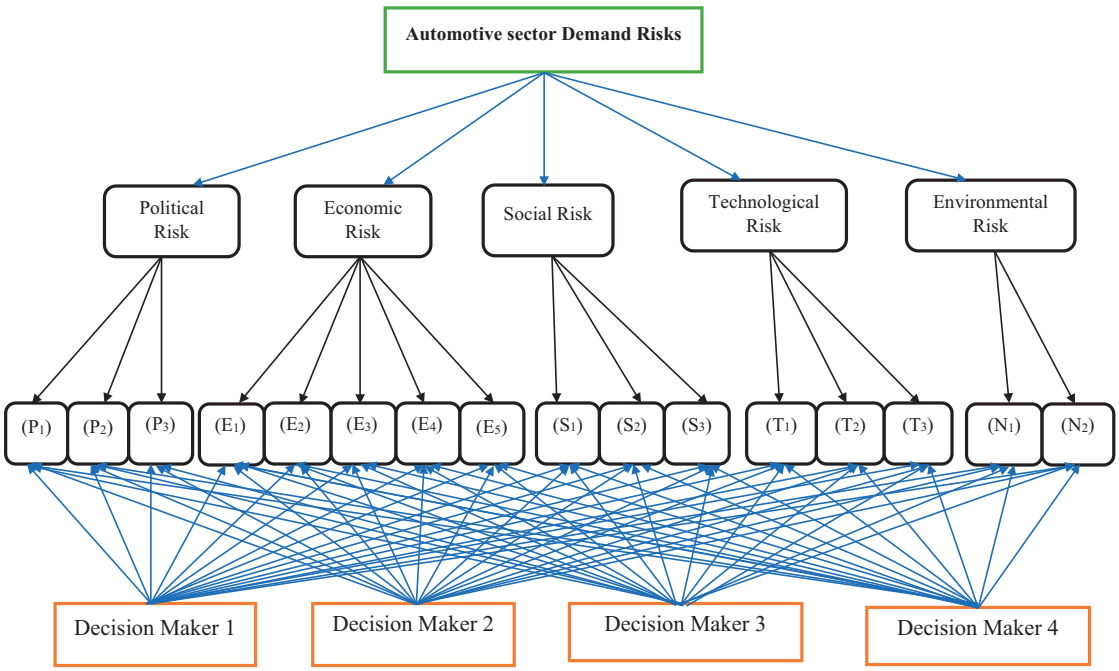


Figure 5.
Integrated AHP and Bayesian Network for automotive risk assessment

Secondly, these weights are used in the Bayesian network to assess the risk for each of the node, and then on each category of risk. The numerical illustration for the same is shown in the following section.

4.1 Numerical illustration

First, the weights for all the risk factors based on their relative importance were collected from four decision makers. Importance of different factors as obtained from the decision makers is given in Table A4 in the Appendix.

4.1.1 Analysis based on fuzzy analytic hierarchy process. Steps to obtain the final ranking of factors affecting the demand of automotive have been illustrated:

Step 1: Following the procedure explained in the research methodology, an integrated pairwise comparison matrix for all the factors (based on expert reviews) is obtained and is presented in Table A5 in the Appendix.

Step 2: By using equation (8), the Fuzzy Geometric Mean ($\tilde{r}$) for each fuzzy number with respect to each of the factors are obtained. For example:

$$\tilde{r}_i = (1 \times 1 \times \ldots \times 1)^{1/16}, (1 \times 1 \times \ldots \times 2)^{1/16}, (1 \times 1 \times \ldots \times 3)^{1/16}$$
$$\Rightarrow \tilde{r}_i = (1.000, 1.500, 2.059)$$

Step 3: Using equation (10), the fuzzy weights, ($\tilde{\tilde{w}}_i$), for all the fuzzy geometric mean values are calculated. For example:

$$\tilde{\tilde{w}}_1 = (1.000, 1.500, 2.059) \otimes [(1.000, 1.500, 2.059) \otimes (1.000, 1.500, 2.059) \otimes \ldots\ldots \ldots \otimes (0.642, 0.760, 1.000)]^{-1}$$
$$\Rightarrow \tilde{\tilde{w}}_1 = (0.04317, 0.08823, 0.16268)$$

Step 4: Using equation (11), each of the fuzzy weights ($\tilde{\tilde{w}}_i$), are defuzzified, to obtain the crisp weights (w_i). For example,

$$w_1 = \frac{1}{3}[0.04317 + 0.08823 + 0.16268]$$
$$\Rightarrow w_1 = 0.0980266$$

Step 5: Final value of normalized weights, (w_{ni}) are obtained using equation (12),

$$w_{n1} = 0.0980266/(0.0980266 + 0.0980266 + \ldots\ldots + 0.0504733)$$
$$\Rightarrow w_{n1} = 0.0870986$$

Calculated values of normalized weights are given in Table A6 in the Appendix.

4.1.2 Analysis based on bayesian network. Normalized weights obtained by fuzzy AHP are given in Table A6 in Appendix. These are used to calculate their respective probability of impact on the demand by solving them through Bayesian network analysis.

Using the formula given in equation (13), the risk for each of the category is calculated,

$$P(\text{Political Risk}) = \frac{\Sigma(\textit{Probability of political risk factors}) \times (\textit{Probability of event occurence})}{\Sigma(\textit{Probability of event occurence})}$$

$$\Rightarrow \mathrm{P(Political\,Risk)} = (P_1 + P_2 + P_3)/(1 + 1 + 1)$$
$$= (0.0732081 + 0.0448465 + 0.0732081)/3$$
$$= 0.06375423$$

Similarly, the risk for all the categories is calculated. The results have been tabulated in Table A7 in the Appendix.

Using the results obtained after applying the Bayesian analysis, different categories of risk are arranged in decreasing order of their magnitude in Table 2. It is observed that economic risk is the most important risk category followed by political, social, technological and environmental risks, respectively.

5. Discussion of findings

This study identifies the different types of risk factors impacting the demand in the automotive industry and presents an approach for modelling and evaluating risk factors for the demand in this sector. By the help of literature review and experts' opinion, total 16 risk factors are identified and these are further grouped into 5 categories i.e. political, economic, social, technological and environmental. Risk profile on the overall demand was assessed by a combined technique of fuzzy AHP and Bayesian network analysis approach. Relative weights for these factors are evaluated by applying the fuzzy AHP approach on the basis of the inputs received from expert interviews. Further, these weights are used in the Bayesian analysis to evaluate the risk for the different subgroups and its impact on the final demand is assessed. After the application of Bayesian analysis, it is observed that organizations are more vulnerable to economic risk. It means, organizations should try to control major dimensions of economic risks such as acquisition and running cost, inflation rate and availability of credit or finance carefully. Thus, from this study, it can be inferred that the changes in demand are more driven by economic factors.

Economic risks include- the economic growth and development, acquisition and running costs, inflation rate, availability of credit or finance and the employment rate. Figure 1 validates the finding that vehicle sales largely depend on the GDP. Study by Muhammad *et al.* (2012) also shows that in the Association of Southeast Asian Nations countries, these economic factors have a noteworthy deep-rooted correlation with the automobile sales. Further, Dargay and Gately (1999), have found that in India, being a developing country, the percentage change in vehicle sales was even more than twice the percentage change in GDP, in between 1966 and 1992. Results by Patra and Rao (2017), in their study regarding economic factors using co-integration and vector models suggest that GDP has a direct relation with the automobile demand, while, the prices and lending rates have an inverse relationship.

Economic risk is followed by political risk, which contain the tax policies, environmental laws and vehicle regulations. This type of risk is again outside the direct control of the firms and possess a huge amount of risk for the demand. Studies suggest that these factors too

Table 2. Risk categories in descending order of risk

Rank	Risk category
1	Economic risk
2	Political risk
3	Social risk
4	Technological risk
5	Environmental risk

contribute greatly to the demand of the automotive (Fridstrøm and Østli, 2017; Clibborn *et al.*, 2016; Alberini and Bareit, 2016; Ott and Andrus, 2000; Pritchard and DeBoer, 1995). Further, from the studies itself, it could be seen that though social and technological factors also impact the overall demand, the intent of these risks is not that high and often lies between 2%-3% (Mulalic *et al.*, 2016; Rith *et al.*, 2019; Barbero and Galeota, 2019; Rawal and Sisodia, 2017; Aghdaie and Yousefi, 2011; Andrea and Smith, 1994). Amongst these factors, the technological ones could be directly or indirectly controlled by an organization itself. Chatterjee *et al.* (2020), suggest that competency in these factors is important for an organization to expand its customer base. At last, the seasonal factors, may have a control on the demand only for a short range of time. These could easily be anticipated by the organizations and its effects could simply be mitigated. Team of experts considered for this study were again consulted with final results of the study. All the experts agreed with final findings and its significance for the industry.

5.1 Managerial implications

The analysis of different types of risks and assessment of the overall risk to the demand in automotive sector will help the industries in this sector in devising their strategies for better gains. The model used in this study would help the associated practitioners to easily assess the risk through each factor at any moment by using appropriate support data. This will, in turn, assist them in a better preparation for facing the future circumstances (for e.g. a sharp surge or decline in the demand) and to deal with the risk situation with accurate mitigation plans by the anticipation of the future demand. For a better efficiency and increased profits, the factors and their categories having higher risk should be given priority to be reduced earlier. The organizations need to devise and implement their policies accordingly. This study reveals that economic risk has the highest magnitude, therefore organizations need to modify their plans more often so as to move hand in hand with the economic changes. These changes majorly include- the economic growth, costs, inflation rate, liquidity and employment. As discussed earlier, the demand increases with economic growth and decreases with an increase in costs, inflation and unemployment. Thus, the rate of production should go hand in hand with the data regarding these aspects. Deviation from the appropriate rate of production may eventually lead to huge losses for the organization. This is valid for all other categories of risk as well. Hence, prioritizing risk factors becomes an essentially important step for risk assessment and its management.

5.2 Theoretical implications

This study has a threefold theoretical contribution to the literature regarding automobile demand and risk assessment. The research contributes:

- to understand the risk factors affecting demand of automotive;
- to study a model that prioritizes these factors by assigning relative weights to them; and
- to assess the risk from each of these factors and correspondingly the overall risk on automotive demand.

Till present there is no single study available where all the external factors affecting the automotive demand have been discussed. In the current study, these factors have been collected through thorough literature and expert reviews. Relative weights have been assigned by applying the fuzzy AHP MCDM technique by using human judgements supported with historical data, and eventually risk probability of each factor has been assessed using the Bayesian network analysis. Thus, it lays the foundation for development

of a new Bayesian MCDM hybrid model for calculation of probability of occurrence of different phenomenon in other areas of study as well. In any sector or industry, it is often required to analyze different factors affecting the outcome of a phenomena, to prioritize the factors and to manipulate their depth of occurrence on the event. The applied methodology provides immense help to the user in assessing and prioritizing factors that lie outside the limits of his direct control. This model would also be extremely useful when an experiment cannot be conducted again and again or when a large sample space of data cannot be gathered. It would also be helpful, when the list of factors varies for organizations or when an organization wants to vary its own importance for the different factors. A successful combination of the fuzzy AHP and Bayesian network is, hence, found to be a very good tool for risk or probability evaluation. Other researchers can use this integrated approach in context to other sectors for analyzing the risks.

6. Limitations and future scope

While the study is very helpful in analyzing the demand risk for the automobile industry, it has a few limitations too. One of the limitations of this study is that the risk factors taken into consideration for assessing are based on literature review and experts' opinions. Therefore, these factors or their extent, could change in the future depending upon the sector and business environment. If a new factor, identical to a non-optimal factor, is introduced, then the ranking of the existing factor may change. Belton and Gear (1983), suggest that the reason for this inconsistency in ranking is that the relative performance measures of all alternatives in terms of each attribute is summed to one. Instead of having the relative performance values sum up to one, dividing each relative performance value by the maximum value in the corresponding vector of relative values was proposed. This avoids the rank reversals when a new non-optimal factor is added. Sometimes it has also been observed that the AHP is not able to include the effects resulting from inter-reliant factors (Gurumurthy *et al.*, 2013). So, there is a necessity to present the interdependencies amongst the factors. In this study, a large sample space of data is not available, as only a limited number of experts could be contacted to gather the data. Though, it has been tried to be resolved by using FST, but there may still be some scope of error, which could be easily removed if it is applied to an industry practically. Other limitation is that it does not provide the mitigation plans for the assessed risks. In the future these mitigation plans can also be devised so as to negate the risk from each of the factors. The suggested approach may also help managers in devising risk mitigation plans. Continuing with this, processes for lean and agile systems can also be implemented to overcome these risks. Future studies may be carried out for finding new risks and possible mitigation strategies in emerging business environment of circular economy and Industry 4.0. Findings may be further validated through case and empirical studies.

The applied methodology can be used to evaluate the factors that affect the demand and the overall risk imposed by them on the auto industry. Organizations need to maintain balance between the production and change in forces of factors affecting demand. Managers and decision makers of associated organizations can use the proposed framework to pre-evaluate their demands corresponding to the future changes. Anticipated demands will help significantly in cost-saving and also in maximizing the profits and to increase the efficiency. This study provides a novel approach for assessing the demand risks. Similar approach can be also applied for other manufacturing sectors. This study can also be extended to assess the physical factors that affect the production rate in an industry. Subjectivity caused due to limited number of experts can be solved by empirical or case-based studies.

References

Abu Ghazaleh, M. and Zabadi, A. (2020), "Promoting a revamped CRM through Internet of Things and Big Data: an AHP-based evaluation", *International Journal of Organizational Analysis*, Vol. 28 No. 1, pp. 66-91.

ACMA, Automotive Components Manufacturers Association of India (2020), "Message", available at: www.acma.in/message.php?id=262# (accessed 11 April 2020).

Aghdaie, S.F.A. and Yousefi, E. (2011), "The comparative analysis of affecting factors on purchasing domestic and imported cars in Iran market-using AHP technique", *International Journal of Marketing Studies*, Vol. 3 No. 2, pp. 142.

Akbari, F., Salehi, M. and Vlashani, M.A.B. (2018), "The relationship between tax avoidance and firm value with income smoothing – a comparison between classical and Bayesian econometric in multilevel models", *International Journal of Organizational Analysis*, Vol. 27 No. 1, pp. 125-148.

Al Hammadi, F. and Hussain, M. (2019), "Sustainable organizational performance: a study of health-care organizations in the United Arab Emirates", *International Journal of Organizational Analysis*, Vol. 27 No. 1, pp. 169-186.

Alberini, A. and Bareit, M. (2016), "The Effect of Registration Taxes on New Car Sales and Emissions: Evidence from Switzerland. CER-ETH Economics Working Paper, Vol. 16, p. 245.

Alexander, C. (Ed.). (2003), *Operational Risk: regulation, Analysis and Management*, Pearson Education, Great Britain.

Andrea, D.J. and Smith, B.C. (1994), "Automotive demand, markets, and material selection processes", *SAE Transactions*, pp. 435-443.

Apak, S., Göğüş, G.G. and Karakadılar, İS. (2012), "An analytic hierarchy process approach with a novel framework for luxury car selection", *Procedia – Social and Behavioral Sciences*, Vol. 58, pp. 1301-1308.

Ayyub, B.M. (2014), *Risk Analysis in Engineering and Economics*, Chapman and Hall/CRC, Taylor and Francis Group, FL.

Azis, I.J. (2005), "A regional cooperation to support financial crisis management and prevention: an application of a model with uncertainty and feedback influences", *International Journal of Organizational Analysis*, Vol. 13 No. 3, pp. 196-215.

Barber, B.M., Click, R.W. and Darrough, M.N. (1999), "The impact of shocks to exchange rates and oil prices on US sales of American and Japanese automakers", *Japan and the World Economy*, Vol. 11 No. 1, pp. 57-93.

Barbero, J.A. and Galeota, C. (2019), "Urban transportation", *The Wiley Blackwell Encyclopedia of Urban and Regional Studies*, pp. 1-12.

Basak, I. and Saaty, T. (1993), "Group decision making using the analytic hierarchy process", *Mathematical and Computer Modelling*, Vol. 17 Nos 4/5, pp. 101-109.

Bellman, R.E. and Zadeh, L.A. (1970), "Decision-making in a fuzzy environment", *Management Science*, Vol. 17 No. 4, pp. B-141.

Belton, V. and Gear, T. (1983), "On a short-coming of saaty's method of analytic hierarchies", *Omega*, Vol. 11 No. 3, pp. 228-230.

Berkovec, J. (1985), "Forecasting automobile demand using disaggregate choice models", *Transportation Research Part B: Methodological*, Vol. 19 No. 4, pp. 315-329.

Bloesch, J. and Gourio, F. (2015), "The effect of winter weather on US economic activity", *Economic Perspectives*, Vol. 39 No. 1.

Bonafede, C.E. and Giudici, P. (2007), "Bayesian networks for enterprise risk assessment", *Physica A: Statistical Mechanics and Its Applications*, Vol. 382 No. 1, pp. 22-28.

Busse, M.R., Pope, D.G., Pope, J.C. and Silva-Risso, J. (2015), "The psychological effect of weather on car purchases", *The Quarterly Journal of Economics*, Vol. 130 No. 1, pp. 371-414.

Chand, M. and Avikal, S. (2015), "An MCDM based approach for purchasing a car from indian car market", in *2015 IEEE Students Conference on Engineering and Systems (SCES)*, IEEE, pp. 1-4.

Chand, M., Raj, T. and Shankar, R. (2014), "Analysing the operational risks in supply chain by using weighted interpretive structure modelling technique", *International Journal of Services and Operations Management*, Vol. 18 No. 4, pp. 378-403.

Chand, M., Raj, T. and Shankar, R. (2015a), "A comparative study of multi criteria decision making approaches for risks assessment in supply chain", *International Journal of Business Information Systems*, Vol. 18 No. 1, pp. 67-84.

Chand, M., Raj, T. and Shankar, R. (2015b), "Risk mitigations strategy in supply chain planning and control: an ANP approach", *International Journal of Productivity and Quality Management*, Vol. 16 No. 1, pp. 92-113.

Chand, M., Raj, T., Shankar, R. and Agarwal, A. (2017), "Select the best supply chain by risk analysis for Indian industries environment using MCDM approaches", *Benchmarking: An International Journal*, Vol. 24 No. 5, pp. 1400-1413.

Chang, D.Y. (1996), "Applications of the extent analysis method on fuzzy AHP", *European Journal of Operational Research*, Vol. 95 No. 3, pp. 649-655.

Charfi, S., BenHamad, S. and Masmoudi, A. (2019), "Assessing the impact of monetary fundamentals on exchange rate fluctuations a Bayesian network approach", *Journal of Modelling in Management*, Vol. 15 No. 1, pp. 166-181.

Chatterjee, S., Chaudhuri, R., Vrontis, D., Thrassou, A., Ghosh, S.K. and Chaudhuri, S. (2020), "Social customer relationship management factors and business benefits", *International Journal of Organizational Analysis*, doi: 10.1108/IJOA-11-2019-1933.

Chauhan, G. (2016), "An analysis of the status of resource flexibility and lean manufacturing in a textile machinery manufacturing company", *International Journal of Organizational Analysis*, Vol. 24 No. 1, pp. 107-122.

Chen, C.T. (2000), "Extensions of the TOPSIS for group decision-making under fuzzy environment", *Fuzzy Sets and Systems*, Vol. 114 No. 1, pp. 1-9.

Chhabra, D., Garg, S.K. and Singh, R.K. (2017), "Analysing alternatives for green logistics in an Indian automotive organization: a case study", *Journal of Cleaner Production*, Vol. 167 No. 20, pp. 962-969.

Chifurira, R., Mudhombo, I., Chikobvu, M. and Dubihlela, D. (2014), "The impact of inflation on the automobile sales in South Africa", *Mediterranean Journal of Social Sciences*, Vol. 5 No. 7, pp. 200-200.

Chirra, S. and Kumar, D. (2018), "Evaluation of supply chain flexibility in automobile industry with fuzzy dematel approach", *Global Journal of Flexible Systems Management*, Vol. 19 No. 4, pp. 305-319.

Chuang, C.H. and Zhao, Y. (2019), "Demand stimulation in finished-goods inventory management: Empirical evidence from general motors dealerships", *International Journal of Production Economics*, Vol. 208, pp. 208-220.

Clibborn, S., Lansbury, R.D. and Wright, C.F. (2016), "Who killed the Australian automotive industry: the employers, government or trade unions?", *Economic Papers: A Journal of Applied Economics and Policy*, Vol. 35 No. 1, pp. 2-15.

Dalla Valle, L. and Giudici, P. (2008), "A Bayesian approach to estimate the marginal loss distributions in operational risk management", *Computational Statistics and Data Analysis*, Vol. 52 No. 6, pp. 3107-3127.

Dargay, J.M. (2001), "The effect of income on car ownership: evidence of asymmetry", *Transportation Research Part A: Policy and Practice*, Vol. 35 No. 9, pp. 807-821.

Dargay, J. and Gately, D. (1999), "Income's effect on car and vehicle ownership, worldwide: 1960-2015", *Transportation Research Part A: Policy and Practice*, Vol. 33 No. 2, pp. 101-138. No.

Donatos, G.S. and Kioulafas, K.E. (1990), "A quantitative analysis of new car sales and advertising in Greece", *European Journal of Operational Research*, Vol. 48 No. 3, pp. 311-317.

Duncan, R.S. (1980), "The effect of gasoline prices on automobile sales", *The American Economist*, Vol. 24 No. 1, pp. 62-66.

Fridstrøm, L. and Østli, V. (2017), "The vehicle purchase tax as a climate policy instrument", *Transportation Research Part A: Policy and Practice*, Vol. 96, pp. 168-189.

Geurts, M.D. and Kelly, J.P. (1986), "Forecasting retail sales using alternative models", *International Journal of Forecasting*, Vol. 2 No. 3, pp. 261-272.

Ghosh, K. (2013), "Prioritization of individual success indicators for managerial success", *International Journal of Organizational Analysis*, Vol. 21 No. 3, pp. 260-287.

Gupta, S., Soni, U. and Kumar, G. (2019), "Green supplier selection using multi-criterion decision making under fuzzy environment: a case study in automotive industry", *Computers and Industrial Engineering*, Vol. 136, pp. 663-680.

Gurumurthy, A., Mazumdar, P. and Muthusubramanian, S. (2013), "Graph theoretic approach for analysing the readiness of an organisation for adapting lean thinking: a case study", *International Journal of Organizational Analysis*, Vol. 21 No. 3, pp. 396-427.

Hamilton, J.D. (1983), "Oil and the macroeconomy since world war II", *Journal of Political Economy*, Vol. 91 No. 2, pp. 228-248.

Heckerman, D. (1996), "A tutorial on learning with bayesian networks (tech. Rep. No. MSR-TR-95-06)", *One Microsoft Way, Redmond, WA, 98052*.

IEC 60300-3-9 (1995), "IEC 60300-3-9, international electrotechnical commission. (1995), "IEC 60300-3-9: 1995 risk management: Part. 3 guide to risk analysis of technological systems", IEC, Geneva.

Jager, R., Verbruggen, H.B. and Bruijn, P.M. (1992), "The role of defuzzification methods in the application of fuzzy control", *IFAC Proceedings Volumes*, Vol. 25 No. 6, pp. 75-80.

Kannan, D., Khodaverdi, R., Olfat, L., Jafarian, A. and Diabat, A. (2013), "Integrated fuzzy multi criteria decision making method and multi-objective programming approach for supplier selection and order allocation in a green supply chain", *Journal of Cleaner Production*, Vol. 47, pp. 355-367.

Kao, H.Y., Huang, C.H. and Li, H.L. (2005), "Supply chain diagnostics with dynamic Bayesian networks", *Computers and Industrial Engineering*, Vol. 49 No. 2, pp. 339-347.

Kaplan, S. and Garrick, B.J. (1981), "On the quantitative definition of risk", *Risk Analysis*, Vol. 1 No. 1, pp. 11-27.

Kauffmann, P.J., Jacobs, D.A. and Fernandez, A.A. (2002), "Use of bayesian probabilities to identify and improve distribution center error rates", *Production and Inventory Management Journal*, pp. 1-6.

Kumar, D. and Rahman, Z. (2017), "Analyzing enablers of sustainable supply chain: ISM and fuzzy AHP approach", *Journal of Modelling in Management*, Vol. 12 No. 3, pp. 498-524.

Kumar, P. and Singh, R.K. (2012), "A fuzzy AHP and TOPSIS methodology to evaluate global 3PL", *Journal of Modelling in Management*, Vol. 7 No. 3, pp. 287-303. VolNo.

Kumar, P., Kumar, D. and Chand, M. (2019), "Analysing the interactions of operational issues in sustainable supply chain: a framework", *International Journal of Services and Operations Management*, Vol. 32 No. 4, pp. 507-521.

Lee, K. and Ni, S. (2002), "On the dynamic effects of oil price shocks: a study using industry level data", *Journal of Monetary Economics*, Vol. 49 No. 4, pp. 823-852.

Leung, L.C. and Cao, D. (2000), "On consistency and ranking of alternatives in fuzzy AHP", *European Journal of Operational Research*, Vol. 124 No. 1, pp. 102-113.

Ludvigson, S. (1998), "The channel of monetary transmission to demand: evidence from the market for automobile credit", *Journal of Money, Credit and Banking*, Vol. 30 No. 3, pp. 365-383.

MacKay, D.B., Bowen, W.M. and Zinnes, J.L. (1996), "A thurstonian view of the analytic hierarchy process", *European Journal of Operational Research*, Vol. 89 No. 2, pp. 427-444.

Mangla, S.K., Kumar, P. and Barua, M.K. (2015), "Flexible decision modeling for evaluating the risks in green supply chain using fuzzy AHP and IRP methodologies", *Global Journal of Flexible Systems Management*, Vol. 16 No. 1, pp. 19-35.

Marx, R., de Mello, A.M. and de Lara, F.F. (2020), "The new geography of the automobile industry: trends and challenges in Brazil", *In New Frontiers of the Automobile Industry*, Palgrave Macmillan, Cham, pp. 349-375.

Milner, J.M. and Kouvelis, P. (2005), "Order quantity and timing flexibility in supply chains: the role of demand characteristics", *Management Science*, Vol. 51 No. 6, pp. 970-985.

Ministry of Road Transport and Highways, Government of India (2020), "Annual report 2018-2019", "Total Number of Registered Motor Vehicles in India: 2003-2017", p. 115, available at: https://morth.nic.in/sites/default/files/Annual_Report_English_2018-19.pdf" (accessed 11 April 2020).

Moreno-Jimenez, J.M. and Vargas, L.G. (1993), "A probabilistic study of preference structures in the analytic hierarchy process with interval judgments", *Mathematical and Computer Modelling*, Vol. 17 Nos 4/5, pp. 73-81.

Mork, K.A. (1989), "Oil and the macroeconomy when prices go up and down: an extension of hamilton's results", *Journal of Political Economy*, Vol. 97 No. 3, pp. 740-744.

Muhammad, F., Hussin, M.Y.M. and Razak, A.A. (2012), "Automobile sales and macroeconomic variables: a pooled mean group analysis for asean countries", *IOSR Journal of Business and Management*, Vol. 2 No. 1, pp. 15-21.

Mulalic, I., Pilegaard, N. and and Rouwendal, J. (2016), "Does improving public transport decrease car ownership? Evidence from the copenhagen metropolitan area", Tinbergen Institute Discussion Paper 15-139/VIII. 10.2139/ssrn.2710547.

Nag, B. and De, D. (2020), "The Indian automobile industry: technology enablers preparing for the future", *In New Frontiers of the Automobile Industry*, Palgrave Macmillan, Cham, pp. 301-321.

Narayanan, V.G., Raman, A. and Singh, J. (2005), "Agency costs in a supply chain with demand uncertainty and price competition", *Management Science*, Vol. 51 No. 1, pp. 120-132.

Nauhria, Y., Pandey, S. and Kulkarni, M.S. (2011), "Competitive priorities for Indian car manufacturing industry (2011-2020) for global competitiveness", *Global Journal of Flexible Systems Management*, Vol. 12 Nos 3/4, pp. 9-20.

Niedermayer, D. (2008), "An introduction to bayesian networks and their contemporary applications", *In Innovations in Bayesian Networks*, Springer, Berlin, Heidelberg, pp. 117-130.

Nielsen, T.D. and Jensen, F.V. (2009), *Bayesian Networks and Decision Graphs*, Springer Science and Business Media, New York, NY.

Ott, R.L. and Andrus, D.M. (2000), "The effect of personal property taxes on consumer vehicle-purchasing decisions: a partitioned price/mental accounting theory analysis", *Public Finance Review*, Vol. 28 No. 2, pp. 134-152.

Pai, R.R., Kallepalli, V.R., Caudill, R.J. and Zhou, M. (2003), "Methods toward supply chain risk analysis", in *SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme-System Security and Assurance (Cat. No. 03CH37483)*, IEEE, Vol. 5, pp. 4560-4565.

Patil, N. and Warkhedkar, R. (2016), "Knowledge management implementation in Indian automobile ancillary industries: an interpretive structural model for productivity", *Journal of Modelling in Management*, Vol. 11 No. 3, pp. 802-810.

Patra, T. and Rao, M.J. (2017), "Impact of macroeconomic factors on automobile demand in India", *Journal of International Economics*, Vol. 8 No. 1, pp. 97-113.

Press Information Bureau (2019a), "Press information bureau, government of India, ministry of heavy industries and public enterprises, (2019 November 26), "slow down in automobile sector", available at: https://pib.gov.in/newsite/PrintRelease.aspx?relid=194999

Press Information Bureau (2019b), "Press information bureau, government of India, ministry of heavy industries and public enterprises, (2019 December 20), "year ender 2019 ministry of heavy industry", available at: https://pib.gov.in/newsite/PrintRelease.aspx?relid=196019

Pritchard, T. and DeBoer, L. (1995), "The effect of taxes and insurance costs on automobile registrations in the United States", *Public Finance Quarterly*, Vol. 23 No. 3, pp. 283-304.

Ramey, V.A. and Vine, D.J. (2004), "Tracking the source of the decline in GDP volatility: an analysis of the automobile industry", (No. w10384), National Bureau of Economic Research.

Rao, R.V. (2007), *Decision Making in the Manufacturing Environment: using Graph Theory and Fuzzy Multiple Attribute Decision Making Methods*, Springer Science and Business Media, London.

Rausand, M. (2013), *Risk Assessment: theory, Methods, and Applications*, John Wiley and Sons, NJ, Vol. 115.

Rawal, P. and Sisodia, D.R. (2017), "Growing acceptance of cab services in urban India", *International Journal in Management and Social Science*, Vol. 5 No. 6, pp. 270-280.

Rith, M., Fillone, A. and Biona, J.B.M. (2019), "The impact of socioeconomic characteristics and land use patterns on household vehicle ownership and energy consumption in an urban area with insufficient public transport service – a case study of metro manila", *Journal of Transport Geography*, Vol. 79, pp. 102484.

Saaty, T.L. (1980), *The Analytic Hierarchy Process*, Mcgraw Hill, New York, NY.

Saaty, T.L. and Vargas, L.G. (1987), "Uncertainty and rank order in the analytic hierarchy process", *European Journal of Operational Research*, Vol. 32 No. 1, pp. 107-117.

Salehi, M., Gah, A.D., Akbari, F. and Naghshbandi, N. (2020), "Does accounting details play an allocative role in predicting macroeconomic indicators? Evidence of Bayesian and classical econometrics in Iran", *International Journal of Organizational Analysis*, doi: 10.1108/IJOA-10-2019-1902.

Shahabuddin, S. (2009), "Forecasting automobile sales", *Management Research News*, Vol. 32 No. 7.

SIAM, Society of Indian Automobile Manufacturers (2020), "Domestic sales trends", available at: www.siamindia.com/statistics.aspx?mpgid=8&pgidtrail=14/ (assessed 10 January 2020).

Singh, R.K. and Sharma, M.K. (2014), "Selecting competitive supply chain using fuzzy AHP and extent analysis", *Journal of Industrial and Production Engineering*, Vol. 31 No. 8, pp. 524-538.

Singh, R.K., Garg, S.K. and Deshmukh, S.G. (2007), "Strategy development for competitiveness: a study on indian auto component sector", *International Journal of Productivity and Performance Management*, Vol. 56 No. 4, pp. 285-304.

Singh, R.K. (2013), "Prioritizing the factors for coordinated supply chain using analytic hierarchy process (AHP)", *Measuring Business Excellence*, Vol. 17 No. 1, pp. 80-98.

Singh, R.K., Garg, S.K. and Deshmukh, S.G. (2008), "Competency and performance analysis of indian SMEs and large organizations: an exploratory study", *Competitiveness Review*, Vol. 18 No. 4, pp. 308-321.

Tambade, H., Singh, R. and Modgil, S. (2019), "Identification and evaluation of determinants of competitiveness in the indian auto-component industry", *Benchmarking: An International Journal*, Vol. 26 No. 3, pp. 922-950.

Taylan, O. (2014), "IT project risk assessment of learning organizations by fuzzy set and systems", *International Journal of Organizational Analysis*, Vol. 22 No. 2, pp. 161-180.

Tipu, S.A.A. (2017), "Entrepreneurial risk taking: themes from the literature and pointers for future research", *International Journal of Organizational Analysis*, Vol. 25 No. 3, pp. 432-455.

Tong, R.M. and Bonissone, P.P. (1980), "A linguistic approach to decision making with fuzzy sets", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 10 No. 11, pp. 716-723.

Torfi, F., Farahani, R.Z. and Rezapour, S. (2010), "Fuzzy AHP to determine the relative weights of evaluation criteria and fuzzy TOPSIS to rank the alternatives", *Applied Soft Computing*, Vol. 10 No. 2, pp. 520-528.

Vatankhah, S., Bouzari, M. and Safavi, H.P. (2020), "Unraveling the fuzzy predictors of stress at work", *International Journal of Organizational Analysis*, doi: 10.1108/IJOA-01-2020-2005.

Wang, F.K., Chang, K.K. and Tzeng, C.W. (2011), "Using adaptive network-based fuzzy inference system to forecast automobile sales", *Expert Systems with Applications*, Vol. 38 No. 8, pp. 10587-10593.

Wetzel, J. and Hoffer, G. (1982), "Consumer demand for automobiles: a disaggregated market approach", *Journal of Consumer Research*, Vol. 9 No. 2, pp. 195-199.

Wu, D., Olson, D.L. and Dolgui, A. (2015), *Decision Making in Enterprise Risk Management: A Review and Introduction to Special Issue*, Omega, Elsevier Ltd.

Wu, J. (2005), "Quantity flexibility contracts under Bayesian updating", *Computers and Operations Research*, Vol. 32 No. 5, pp. 1267-1288.

Yadav, N.S. and Sagar, M. (2015), "Modeling strategic performance management of automobile manufacturing enterprises: an Indian context", *Journal of Modelling in Management*, Vol. 10 No. 2, pp. 198-225.

Yadav, V. and Sharma, M. (2016), "Multi-criteria supplier selection model using the analytic hierarchy process approach", *Journal of Modelling in Management*, Vol. 11 No. 1, pp. 326-354.

Zadeh, L.A. (1965), "Fuzzy sets", *Information and Control*, Vol. 8 No. 3, pp. 338-353.

Zahraee, S.M., Mamizadeh, F. and Vafaei, S.A. (2018), "Greening assessment of suppliers in automotive supply chain: an empirical survey of the automotive industry in Iran", *Global Journal of Flexible Systems Management*, Vol. 19 No. 3, pp. 225-238.

Zhang, Q., Zhan, H. and Yu, J. (2017), "Car sales analysis based on the application of big data", *Procedia Computer Science*, Vol. 107, pp. 436-441.

Further reading

Bhawan, V. and Marg, S.J.S. (2005), "Interpretive matrix: a tool to aid interpretation of management and social research", *Global Journal of Flexible Systems Management*, Vol. 6 No. 2, pp. 27-30.

Dargay, J. (2007), "The effect of prices and income on car travel in the UK", *Transportation Research Part A: Policy and Practice*, Vol. 41 No. 10, pp. 949-960.

Jagathy Raj, V.P. and Balakrishnan, M. (2012), "Dominant partial least square factors of consumer purchase behaviour of passenger cars", *Asian Journal of Management Research*, Vol. 3 No. 1, pp. 83-99.

Press Information Bureau (2017), "Press information bureau, government of India, ministry of heavy industries and public enterprises, (2017, august 2), "automobile sector's contribution to GDP", available at: https://pib.gov.in/newsite/PrintRelease.aspx?relid=169436

Shalender, K. and Yadav, R.K. (2019), "Strategic flexibility, manager personality, and firm performance: the case of Indian automobile industry", *Global Journal of Flexible Systems Management*, Vol. 20 No. 1, pp. 77-90.

The World Bank, Data (2020), "GDP growth rate (annual %), India", available at: https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?locations=IN (accessed 11 April 2020).

Appendix

Political risk	Economic risk	Social risk	Technological risk	Environmental risk
Tax policies (P_1) Environmental laws(P_2) Vehicle regulations (P_3)	Economic growth and development (E_1) Acquisition and running cost (E_2) Inflation rate (E_3) Availability of credit or finance (E_4) Employment rate (E_5)	Mass transit system (S_1) Cab ridership services (S_2) Vicinity of Roads and availability of parking (S_3)	Introduction of new vehicles (T_1) Automation in vehicles (T_2) Engine efficiencies (T_3)	Pollution levels (N_1) Season and climate (N_2)

Table A1. Classification of factors affecting the risk of automotive

Risk	Risk category	References
Economic growth and development	Economic	Muhammad *et al.* (2012), Shahabuddin (2009), Dargay (2001), Tambade *et al.* (2019)
Introduction of new vehicles	Technological	Nag and De (2020), Yadav and Sagar (2015)
Acquisition and running costs	Economic	Hamilton (1983), Mork (1989), Lee and Ni (2002), Barber *et al.* (1999), Duncan (1980), Chand and Avikal (2015)
Inflation rate	Economic	Chifurira *et al.* (2014)
Availability of credit or finance	Economic	Ludvigson (1998), Muhammad *et al.* (2012)
Tax policies	Political	Pritchard and DeBoer (1995), Alberini and Bareit (2016), Ott and Andrus (2000), Clibborn *et al.* (2016)
Employment rate	Economic	Muhammad *et al.* (2012), Shahabuddin (2009)
Mass transit systems	Social	Rith *et al.* (2019), Mulalic *et al.* (2016)
Cab ridership services	Social	Rawal and Sisodia (2017), Akbari *et al.* (2018)
Vicinity of roads and availability of parking	Social	Barbero and Galeota (2019)
Automation	Technological	Andrea and Smith (1994), Aghdaie and Yousefi (2011)
Engine efficiencies	Technological	Andrea and Smith (1994), Aghdaie and Yousefi (2011), Chand and Avikal (2015)
Pollution levels	Environmental	Barbero and Galeota (2019)
Season and climate	Environmental	Busse *et al.* (2015), Bloesch and Gourio (2015)

Table A2. Risk factors affecting product demand

Experts	Designation	Education qualification	Company where employed
Expert 1	Senior manager	MBA	Omax Auto
Expert 2	Manager	MBA	Maruti Suzuki
Expert 3	Assistant manager	MBA	Honda
Expert 4	Technical expert	PGDBA, M.Tech.	Mercedes

Table A3. Experts profile

Table A4. Level of factor importance according to decision makers

S. no.	Factors	Level of importance DM1	DM2	DM3	DM4
1.	Economic growth and Dev. (e.g. disposable income, etc), E_1	4	4	4	2
2.	Introduction of new vehicles with newer technology, T_1	5	4	4	3
3.	Acquisition and running costs, E_2	2	4	4	4
4.	Inflation rate, E_3	2	2	5	5
5.	Availability of credit/ finance, E_4	1	3	3	5
6.	Tax policies (taxes on purchase, incentives, etc), P_1	4	2	3	4
7.	Environmental laws (EVs, cleaner fuels, etc), P_2	1	2	2	4
8.	Vehicle regulations (Axle, safety, emission, etc), P_3	3	1	3	3
9.	Employment rate, E_5	2	1	1	4
10.	Public transport, S_1	2	1	3	2
11.	Cab services, S_2	3	4	4	3
12.	Vicinity of roads/ availability of parking spots, S_3	2	2	2	3
13.	Automation in vehicles, T_2	2	4	1	2
14.	Engine efficiencies, T_3	4	1	2	1
15.	Pollution levels, N_1	1	3	3	2
16.	Season and climate, N_2	1	2	3	1

Table A5.
Fuzzy integrated pairwise comparison matrix

Factors	E_1	T_1	E_2	E_3	E_4	P_1	P_2	P_3
E_1	(1, 1, 1)	(1, 1, 1)	(1,1.34,2)	(1, 1, 1)	(1,1.34,2)	(1,1.34,2)	(1, 2, 3)	(1,1.34,2)
T_1	(1, 1, 1)	(1, 1, 1)	(1,1.34,2)	(1, 1, 1)	(1,1.34,2)	(1,1.34,2)	(1, 2, 3)	(1,1.34,2)
E_2	(0.6,0.75,1)	(0.6,0.75,1)	(1, 1, 1)	(0.6,0.75,1)	(1, 1, 1)	(1, 1, 1)	(1,1.5,3)	(1, 1, 1)
E_3	(1, 1, 1)	(1, 1, 1)	(1,1.34,2)	(1, 1, 1)	(1,1.34,2)	(1,1.34,2)	(1, 2, 3)	(1,1.34,2)
E_4	(0.6,0.75,1)	(0.6,0.75,1)	(1, 1, 1)	(0.6,0.75,1)	(1, 1, 1)	(1, 1, 1)	(1,1.5,3)	(1, 1, 1)
P_1	(0.6,0.75,1)	(0.6,0.75,1)	(1, 1, 1)	(0.6,0.75,1)	(1, 1, 1)	(1, 1, 1)	(1,1.5,3)	(1, 1, 1)
P_2	(0.34,0.5,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.5,0.67,1)	(1, 1, 1)	(0.5,0.67,1)
P_3	(0.6,0.75,1)	(0.6,0.75,1)	(1, 1, 1)	(0.6,0.75,1)	(1, 1, 1)	(1, 1, 1)	(1,1.5,3)	(1, 1, 1)
E_5	(0.34,0.5,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.5,0.67,1)	(1, 1, 1)	(0.5,0.67,1)
S_1	(0.34,0.5,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.5,0.67,1)	(1, 1, 1)	(0.5,0.67,1)
S_2	(1, 1, 1)	(1, 1, 1)	(1,1.34,2)	(1, 1, 1)	(1,1.34,2)	(1,1.34,2)	(1, 2, 3)	(1,1.34,2)
S_3	(0.34,0.5,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.5,0.67,1)	(1, 1, 1)	(0.5,0.67,1)
T_2	(0.34,0.5,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.5,0.67,1)	(1, 1, 1)	(0.5,0.67,1)
T_3	(0.34,0.5,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.5,0.67,1)	(1, 1, 1)	(0.5,0.67,1)
N_1	(0.34,0.5,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.5,0.67,1)	(1, 1, 1)	(0.5,0.67,1)
N_2	(0.34,0.5,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.5,0.67,1)	(1, 1, 1)	(0.5,0.67,1)

(*continued*)

Table A5.

Factors	E_5	S_1	S_2	S_3	T_2	T_3	N_1	N_2
E_1	(1, 2, 3)	(1, 2, 3)	(1, 1, 1)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)
T_1	(1, 2, 3)	(1, 2, 3)	(1, 1, 1)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)
E_2	(1,1.5,3)	(1,1.5,3)	(0.6,0.75,1)	(1,1.5,3)	(1,1.5,3)	(1,1.5,3)	(1,1.5,3)	(1,1.5,3)
E_3	(1, 2, 3)	(1, 2, 3)	(1, 1, 1)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)
E_4	(1,1.5,3)	(1,1.5,3)	(0.6,0.75,1)	(1,1.5,3)	(1,1.5,3)	(1,1.5,3)	(1,1.5,3)	(1,1.5,3)
P_1	(1,1.5,3)	(1,1.5,3)	(0.6,0.75,1)	(1,1.5,3)	(1,1.5,3)	(1,1.5,3)	(1,1.5,3)	(1,1.5,3)
P_2	(1, 1, 1)	(1, 1, 1)	(0.34,0.5,1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)
P_3	(1,1.5,3)	(1,1.5,3)	(0.6,0.75,1)	(1,1.5,3)	(1,1.5,3)	(1,1.5,3)	(1,1.5,3)	(1,1.5,3)
E_5	(1, 1, 1)	(1, 1, 1)	(0.34,0.5,1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)
S_1	(1, 1, 1)	(1, 1, 1)	(0.34,0.5,1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)
S_2	(1, 2, 3)	(1, 2, 3)	(1, 1, 1)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)
S_3	(1, 1, 1)	(1, 1, 1)	(0.34,0.5,1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)
T_2	(1, 1, 1)	(1, 1, 1)	(0.34,0.5,1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)
T_3	(1, 1, 1)	(1, 1, 1)	(0.34,0.5,1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)
N_1	(1, 1, 1)	(1, 1, 1)	(0.34,0.5,1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)
N_2	(1, 1, 1)	(1, 1, 1)	(0.34,0.5,1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)

Factors	Fuzzy geometric mean ($\tilde{r}$)	Fuzzy weights ($\tilde{w}_i$)	Weights (w_i)	Normalized weights(w_{ni})
E_1	(1.000, 1.500, 2.059)	(0.04317, 0.08823, 0.16268)	0.0980266	0.0870986
T_1	(1.000, 1.500, 2.059)	(0.04317, 0.08823, 0.16268)	0.0980266	0.0870986
E_2	(0.880, 1.230, 1.732)	(0.03798, 0.07235, 0.13685)	0.0823933	0.0732081
E_3	(1.000, 1.500, 2.059)	(0.04317, 0.08823, 0.16268)	0.0980266	0.0870986
E_4	(0.880, 1.230, 1.732)	(0.03798, 0.07235, 0.13685)	0.0823933	0.0732081
P_1	(0.880, 1.230, 1.732)	(0.03798, 0.07235, 0.13685)	0.0823933	0.0732081
P_2	(0.642, 0.760, 1.000)	(0.02771, 0.04470, 0.07901)	0.0504733	0.0448465
P_3	(0.880, 1.230, 1.732)	(0.03798, 0.07235, 0.13685)	0.0823933	0.0732081
E_5	(0.642, 0.760, 1.000)	(0.02771, 0.04470, 0.07901)	0.0504733	0.0448465
S_1	(0.642, 0.760, 1.000)	(0.02771, 0.04470, 0.07901)	0.0504733	0.0448465
S_2	(1.000, 1.500, 2.059)	(0.04317, 0.08823, 0.16268)	0.0980266	0.0870986
S_3	(0.642, 0.760, 1.000)	(0.02771, 0.04470, 0.07901)	0.0504733	0.0448465
T_2	(0.642, 0.760, 1.000)	(0.02771, 0.04470, 0.07901)	0.0504733	0.0448465
T_3	(0.642, 0.760, 1.000)	(0.02771, 0.04470, 0.07901)	0.0504733	0.0448465
N_1	(0.642, 0.760, 1.000)	(0.02771, 0.04470, 0.07901)	0.0504733	0.0448465
N_2	(0.642, 0.760, 1.000)	(0.02771, 0.04470, 0.07901)	0.0504733	0.0448465
Σ	(12.656, 17.000, 23.164)		1.125466	1.000000

Table A6. Calculated values of normalized weights

Risk categories	Risk factors	Factor probability	Risk
Political risk	Tax policies	0.0732081	0.06375423
	Environmental laws	0.0448465	
	Vehicle regulations	0.0732081	
Economic risk	Economic growth and development	0.0870986	0.07309198
	Acquisition and running cost	0.0732081	
	Inflation rate	0.0870986	
	Availability of credit or finance	0.0732081	
	Employment rate	0.0448465	
Social risk	Mass transit system	0.0448465	0.05893053
	Cab ridership services	0.0870986	
	Vicinity of roads and availability of parking	0.0448465	
Technological risk	Introduction of new vehicles	0.0870986	0.05893053
	Automation in vehicles	0.0448465	
	Engine efficiencies	0.0448465	
Environmental risk	Pollution levels	0.0448465	0.04484650
	Season and climate	0.0448465	

Table A7. Risk for each of the risk category

About the authors

Dr Girish Kumar is Associate Professor in Mechanical Engineering Department at Delhi Technological University, Delhi, India. He has more than 22 years of experience in industry, teaching and research. His teaching and research areas include Reliability Engineering, Maintenance Management, Quality engineering, Stochastic Modeling, etc. He holds a PhD degree in Reliability Engineering from Indian Institute of Technology Delhi, India. Dr Kumar has published 25 papers in International Journals and conference proceedings.

Dr Rajesh Kr. Singh is a Professor at the Management Development Institute, Gurgaon, India. His areas of interest include supply chain management, Industry 4.0, quality management and circular economy. He has about 170 research papers published in reputed international/national journals and

conferences. He has published papers in journals such as *Resources, Conservation and Recycling, Journal of Cleaner Production, Industrial Management and Data Systems, Production planning and Control, Singapore Management Review, International Journal of Productivity and Performance Management, Benchmarking: An International Journal, Journal of Modelling in Management, Competitiveness Review: An International Business Journal, Global Journal of Flexible Systems and Management, International Journal of Productivity and Quality Management, International Journal of Services and Operations Management, International Journal of Automotive Industry and Management, International Journal of Logistic Systems Management, IIMB Management Review and Productivity Promotion.* Rajesh Kr. Singh is the corresponding author and can be contacted at: rksdce@yahoo.com

Mr Rishabh Jain is an undergraduate student at Delhi Technological University, (formerly known as Delhi College of Engineering), New Delhi, India. He is pursuing his Bachelor in Technology with Mechanical Engineering as major. His areas of interest include- decision-making, risk-modelling and other concepts of industrial engineering management.

Mr Raman Kain is a final year student at Delhi Technological University, New Delhi, India. He is pursuing his B.Tech with Mechanical Engineering as major. His areas of interest include- creative problem solving, risk analysis and concepts of quality management.

Mr Naveen is a final year student at Delhi Technological University, New Delhi, India. He is pursuing his Bachelor in Technology with Mechanical Engineering as major. His specialization is automotive engineering.

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/336668692

Analysis of Total Quality Management (TQM) Barriers on its Performance using TOC

Article · August 2019

DOI: 10.22214/ijraset.2019.8097

CITATIONS
0

READS
32

1 author:



Rakesh Kumar
Delhi Technological University
7 PUBLICATIONS **0** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Performance Measurement View project

Population Statistics View project

All content following this page was uploaded by Rakesh Kumar on 12 July 2020.

The user has requested enhancement of the downloaded file.

iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** VIII **Month of publication:** August 2019

DOI: http://doi.org/10.22214/ijraset.2019.8097

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VIII, Aug 2019- Available at www.ijraset.com

Analysis of Total Quality Management (TQM) Barriers on its Performance using TOC

Rakesh Kumar[1], Prof. R.S. Mishra[2]
[1]*Ph.D Scholar,* [2]*Professor, Department of Mechanical Engineering, Delhi Technological University, Delhi, India-110042*

Abstract: ***This study examines and analyzes the TQM factors which act as barrier and affects its performance. TQM view the organization as a collection of processes that must be continuously improved through utilization of the knowledge and experience of associates in all functions and at all levels. The industries must be more focused on understanding their own structure in terms of production/ service processes. Theory of Constraints (TOC) focuses on the weakest ring(s) in the chain to improve the performance of process systems. In this situation, TOC becomes an important problem structuring and solving methodology which changes the way of thinking of decision makers. This paper includes the extensive literature review on Theory of Constraints and suggested that how the same approach to analyze the performance of the Total Quality Management system.***

Keywords: Performance measurement, Theory of constraints (TOC), TQM.

I. INTRODUCTION

In today scenario of the industrial increasing competitiveness and cutthroat competition to capture wide market has led to the introduction of unconventional manufacturing management strategies such as just-in-time (JIT), material resources planning (MRP), Six-sigma quality control (SSQC), Lean Manufacturing and the theory of constraints (TOC). The basic concept of Theory of Constraints, shortly known as TOC assumes that the performance of the organization or say organizational system is adversely affected by constraints present in the system. Within the organizational system process, various functions are performed that are part of aiding the increase within production. The TOC then develops an explicit approach to manage these constraints to maximize throughput with continuous improvement. That's why it was also called that TOC is a multi-faceted systems methodology that has been progressively developed to assist people and organizations to think about problems, develop breakthrough solutions and implement those solutions successfully. The TOC defines a constraint as, *"Anything that prevents the organisation from achieving higher performance versus its goal"*. So, in the case of a profit-making organisation this centers on what limits. The TOC claim is that there are few constraints to any system preventing it from achieving its goal. Although TOC originally focused on physical resource constraints the underlying constraints are commonly policy, or deeper paradigm constraints. TOC uses cognitive mapping to verify this assumption where necessary, but there are many other authors who acknowledge the importance of underlying core problems that constrain system improvement. The main objective of this study is to identify the critical factors which behaves like bottleneck is required for TQM implementation in the organization, and these factors act as the critical chain. TOC is a complex methodology requiring skill and cooperation to implement. The past literature about the link between TQM practices and organization performance gives contradictory results. So the purpose of this study is to develop a conceptual framework and a research model showing the relationship between TQM implementation and performance measures of organization. Most of the previous works show that TQM has significant relationship with firm's performance. However, the examining of moderators is less given in previous work, which mediators are known generally as general tools and techniques without specific focus on types of improvement. And also to determine the effects of moderators TPM, SPC, Lean such as which improve the relationship between TQM practices and firm's performance.

II. LITERATURE REVIEW

In past TOC has developed rapidly in terms of both methodology and area of applications. Dettmer (1995) in his literature compared the system approach taken by TQM and TOC. He argued that TQM views the system in terms of discrete processes and then optimizes the quality in each process, whereas TOC improves performance by concentrating on the weakest link in the total system. It is apparent, however, that without subsequent identification of the weakest links, and application of the philosophy to these groups or processes within the organization, major improvements cannot be achieved. Dettmer further described the CRT technique of TOC is as functional rather than organizational and as such is blind to internal and external system boundaries.

©IJRASET: All Rights are Reserved

The TOC is providing a method of focusing improvement on the critical areas, and thus has the potential for much faster improvement towards a more global goal. The TOC can also contribute by emphasizing methods of dealing with change. The main obstacle for broader implementation of TQM in many industrial organizations, the need to win over those who do not believe in the method and who find it difficult to wait for TQM payoffs when they are battling daily problems.

The TOC provides some answers to these problems by contributing the Socratic thinking process for dealing with change, and potentially creating ownership and commitment throughout the organization. The focus on the weakest links helps also in hastening the improvement and generating more apparent results at a quicker pace. To address the policy constraints and effectively implement the process of on-going improvement, Goldratt (1990, 1994) develop a generic approach called the "thinking process" (TP).

Literature concerning TOC in different manufacturing system, like, JIT and TOC buffering philosophies are compared and suggested that improved system performance stems from the strategic placement of buffers in DBR(Drum-Buffer-Rope), which maximizes protection of the constraint from variation rather than attempting to protect each individual station is given by Watson, K.J.et.al (2008). Spencer, M.S. et.al (1995) also supported through his paper that DBR technique of TOC can be used to improve current production operations.

Gupta, S., (1997), Perez, J.L. (1997) and Watson, K.J. et.al (2003) advocated for the same technique of TOC for managing the supply chains systems. Boyd and Gupta (2004) established TOC as a theory by identifying the underlying construct "throughput orientation" along with its three dimensions: Mindset, Measures and Methodology. The thinking process tools of TOC provide techniques for identifying and solving root problems – which is a stated aim of TQM practices. Cox et al. (2003) provided some real-life examples of the applications of the TOC thinking process tools to quality problems.

A. TOC Method to Address TQM

A number of authors and quality experts consider that TQM principles and practices are universally applicable to any organization (Deming, 1982) regardless of the organisational contextual factors (e.g. size, technology, culture or external environment) (Zhao et al., 2004). Despite the claimed benefits of TQM, the relevant literature points to many research studies that indicate a high rate of failures in the process of implementing TQM practices, due to barriers that hinder its implementation (Sila, 2007; Soltani and Wilkinson, 2010).

TQM, Six-sigma, Lean Manufacturing, TOC all these strategic tools aiming similar manner, but with distinct flavor. These quality improvement tools first create its own virtual system in the existing system where it has to be implemented with active participation of the employee. In our further analysis of the relationship between TQM and TOC we will use the following definition of TQM (Dahlgaard et al., (1998): TQM is a company culture characterized by increased customer satisfaction through continuous improvements, in which all employees actively participate.

Researchers exploited these strategic tools far away from its actual usages. Like Six-sigma was initially developed by Motorola in later 80's for the product centric defect reduction and quality improvement tool, later it was experimented and successfully implemented in various areas.

Schonberger (2008) points out that the objective of Six Sigma programs is to create a higher perceived value of the company's products and services in the eyes of the customer. Six Sigma is a well-established approach that seeks to identify and eliminate defects, mistakes or failures in business processes or systems by focusing on those process performance characteristics that are of critical importance to customers (Snee, 2004).

Ehie and Sheu (2005) were able to integrate the theory of constraints with the DMAIC concept for achieving continuous quality improvement. Similarly, Lean manufacturing is designed for the process improvement as well as reduction/elimination of the existing waste, and further it also applied in the field which were substantial.

TQM is a wide concept since it embraces the whole organization and its processes instead of focusing on the product. It is also considered by many to be a holistic approach, which seeks to convert the culture and structure of the organization into a total commitment to quality (Barad, 1996). TQM is generally considered to be based on a number of core values such as customer focus, decisions based on facts, process orientation, continuous improvement, everybody's commitment and leadership; see Hellsten and Klefsjo (2000).

©IJRASET: All Rights are Reserved

Table I: Linkage of Barriers to TQM

Factors	Barriers
Support Quality	• Lack of Leadership commitment • High turnover of management level • Company work culture is not synergetic with company goal and objectives • Difficult to change employee mindset about quality • Lack of information about quality
Support Internal Communication	Lack of coordination amongst departments
Provide quality resources	Lack of sufficient funds to mobilize TQM driven activities
Use competent personal	• Poor education levels of workers • Low morale (absenteeism, industrial action etc.) • Lack of skill of workers • High worker turnover • Indiscipline (non-conformances with procedures)
Support Competence	• Insufficient Training about quality • Management related training is not achieving organizational training targets
Provide quality Infrastructure	• Poor condition of machine • Ineffective maintenance programs • Poor condition of equipment spare part procurement
Control Purchasing Process	• Raw material does not confirm with the specification • Spontaneous delivery of raw materials • Difficulty in procurement of raw material

The core idea of TOC is that every system has at least one constraint that prevents from achieving the goal to a larger degree. Constraints can be physical resources or policies. TOC develops a set of procedures and methodologies to identify and optimize such constraints. In this paper TQM is considered as a system under the organizational system and the barriers are considered as the constraints, which hinders the quality goal. And the well-known fact is that the success of the system depends on everybody's participation (one of the core principles of TQM).

The Theory of Constraints is based on five steps.

1) Identify the system's constraints(s)
2) Decide how to exploit the system's constraint(s)
3) Subordinate everything else to the above decision
4) Elevate the system's constraint(s)
5) If, in the previous steps, a constraint has been broken, go back to step 1, and do not allow inertia to cause a system's constraint.

The system's constraint is that part of the system that constrains the objective or goal of the system. The issues with TOC is that the constraints must be kept operating at its full capacity, if not then entire process slows further. The organization where TQM is implemented, and the barriers/hurdles/constraints even if identified must be kept operating to maintain the pace of the production. Upstream operations must provide only what the constraint can handle and downstream operations will only receive what the constraint can put out. Total Quality Management System is system of any organizational systems. It is known that System of systems is a collection of task-oriented or dedicated systems that pool their resources and capabilities together to create a new, more complex system which offers more functionality and performance than simply the sum of the constituent systems.

The failures in implementing TQM are typically attributed to organisational contingent, situational or contextual factors such as unsupportive organisational culture, resistance to change, lack of resources, or uncommitted leadership (Zhao et al., 2004; Sadikoqlu and Zehir, 2008). It is also said about TQM that it is system that starts with consumers and ends with them. So successful implementation of TQM improves organizational performance and leads to high degree of satisfaction to customers and employees.

©IJRASET: All Rights are Reserved

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VIII, Aug 2019- Available at www.ijraset.com

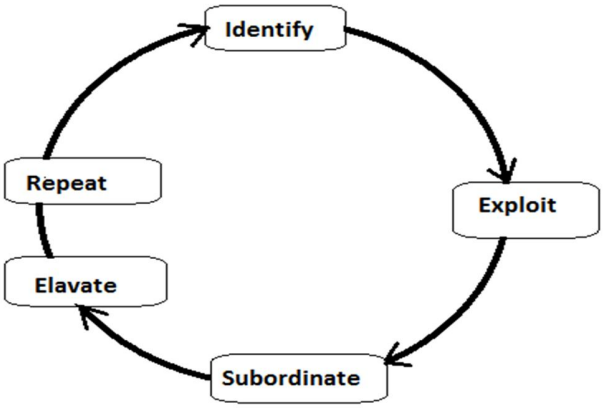


Figure1. Five aiming steps of the TOC

TQM systems include the procedures for quality planning strategy and operations, for setting capital and operating budgets, for measuring and rewarding performance, and for reporting progress and conducting meetings. many companies introduced total quality management as a new management system. But while TQM enabled firms to focus more effectively on process improvements, the ability to implement strategy across organizational units remained elusive. Companies' management systems were still tactical and operational, not strategic. In terms of the TQM system's constraint is the existing barriers. The whole performance of TQM system depends on successful run of the TQM drivers and enablers and the TQM driver depends on the TQM capability. Consequently, it is essential to keeping in view the quality goals. In Step 1 this needs to be identified. By identifying these barriers, company management can increase their understanding of implementation failure and the subsequent removal of these barriers can lead to the possibility of successful implementation. Thus, any company that was rated as high performing was evaluated as moving toward successful implementation of TQM practices. Activities to remove barriers for consensus encompasses a wide range of activities including leadership, decision-making tools, and other tools or tactics to restructure culture and/or to re-motivate employees to the TQM paradigm. As it is a system, then the maximum possible utilization needs to be achieved. This may mean to eliminate those factors which were identified as obstacle for the accomplish the quality improvement procedure as soon as possible or reducing the number of barriers. It will mean ensuring that there is always analysis of the system to do. Thus the system's constraint is exploited (Step 2). If this is the constraint, then there is no point running the enablers: so every other TQM tools needs to be subordinated to remove the bottleneck if still present (Step 3). To achieve the goal/ objective, the system's constraint may need to be `eliminated' or minimized. The enablers, the capacity of the system, has been increased, or `elevated' (Step 4). The application of Step 4 may have changed the system's constraint. With its increased drivers, the original bottleneck may no longer be constraining the system, so the new bottleneck needs to be identified, and the process repeated (Step 5). Thus, this is a process of continual improvement.

B. TOC Strategies

The core premise of TOC is that any organization or system's performance is limited most by a "constraint" — the thing that most blocks throughput. Identifying the nature and location of the system's constraint is the key to continuous system improvement, stated by Srikanth et al. (1997). TOC is a data-driven approach like other continuous improvement processes uses Identify the system's constraints, decide how to exploit the constraint(s), Subordinate everything else to the above decisions, Elevate the constraint(s) and Don't let inertia become the new constraint - go back to step 1, but do not allow previous decisions made in steps 2 to 4 to become constraints because TOC views constraints as positive, not negative. Because constraints determine the performance of a system, a gradual elevation of the system's constraints will improve its performance. According to Dettmer (1995), he compared the system approach taken by TQM and TOC. He argued that TQM views the system in terms of discrete processes and then optimizes the quality in each process, whereas TOC improves performance by concentrating on the weakest link in the total system. Whether organization manages stand-alone or multiple projects, whether those projects are small or large, whether customers are internal or external, or whether the nature of the work performed is product development, construction, design, IT, or service; most projects are difficult to manage because of two things:

©IJRASET: All Rights are Reserved

1) They involve uncertainty, and
2) They involve three different and opposing commitments: Due date, budget, and content

The main strategies of TOC along with tasks and its descriptions summarized in Table II.

Table II

TASK	Description
Identifying "What to Change?"	Organization must effectively address the underlying root causes that lead to the problems.
Identifying "To What to Change To?"	To address these root causes and coping mechanisms, a comprehensive solution must be provided. The solution includes: 1) a robust planning process, 2) a more effective scheduling process, 3) a methodology for introducing work that actually leads to increased capacity, 4) execution processes that provide excellent project control, visibility and decision support, and 5) work behaviors that are more conducive to good project performance.
Identifying "How to cause the change?"	The right people must be brought into the picture at the right time, in just the right way. The process must move slowly enough to permit identification of essential changes, yet fast enough that it does not lose the momentum that is necessary to sustain continued progress.

III. CONCLUSION AND FUTURE RESEARCH

The performance of the Total Quality Management (TQM) System is the responsibility of every person involved in all activities related to the company. As in this work TQM considered as a management system for a customer-focused organization that involves all employees in continual improvement. This paper combines the powerful and practical management theories – theory of constraints over TQM. When applied this concept enable managers to analyse and control the most important factors in their firm – the constraints – in new ways.

TOC may be also applied for the lean manufacturing, six-sigma, Business process reengineering (BPR), supply chain management (SCM) and logistic system by considering them as a system inside the system. Cellular manufacturing system and group technology is also one of the untouched field where TOC strategic management tool may be implemented in the similar fashion for the better result. All the above mentioned the defined process or system functions as intended and produces intended results under normal operating conditions. So, as a future work, the barriers may be ranked and weighted to get the precise result is recommended. Choosing the right TOC tool for the desired solution is also the one of the most challenging job always.

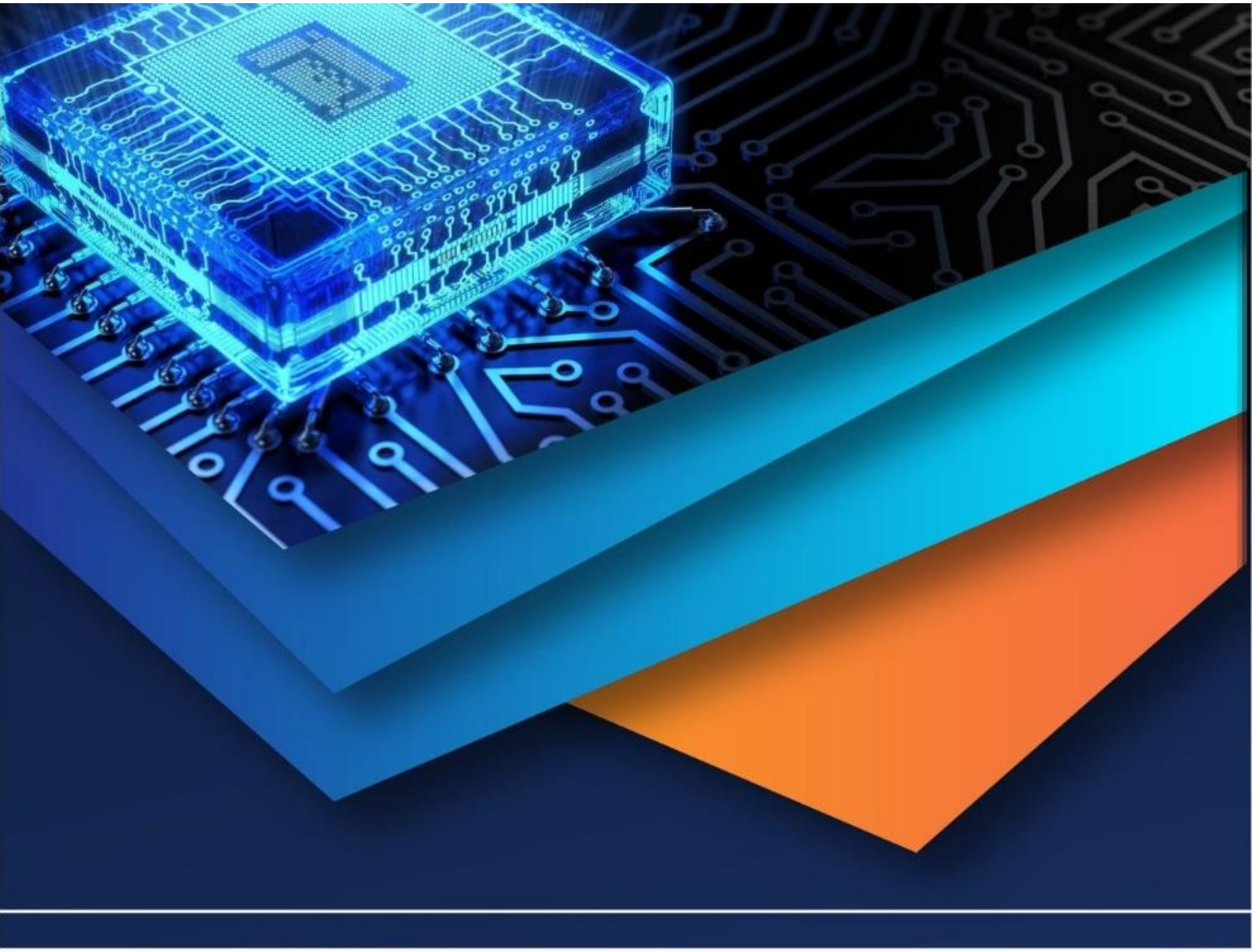
REFERENCES

[1] Boyd, L.H., Gupta, M.C. (2004), "Constraints management: what is the theory?", International Journal of Operations & Production Management, Vol. 24 No. 4, pp. 350-371.
[2] BARAD M,. (1996) Total quality management, International Encyclopedia of Business and Management, 5.
[3] Baladhandayutham, R., Devadasan, S. R., Selladurai, V., & Senthil, V. (2001). Integration of BPR and TQM: Past, present and future trends. Production Planning Control: The Management of Operations, 12(7), 680–688.
[4] Catalin, S. H., Bogdan, B., & Dimitrie, G. R. (2014). The existing barriers in implementing total quality management. Annals of the University of Oradea, Economic Science Series, 23(1), 1234–1240.
[5] Cox, J.F., Blackstone, J.H. and Schleier, J.G. (2003), Managing Operations: A Focus on Excellence, North River Press, Great Barrington, MA.
[6] C. Carl Pegels, Craig Watrous, (2005) "Application of the theory of constraints to a bottleneck operation in a manufacturing plant", Journal of Manufacturing Technology Management, Vol. 16 Issue: 3, pp.302-311,
[7] Chakravorty, S.S., Atwater, J.B., (2006) Bottleneck management: theory and practice. Production & Planning Control 17 (5), 441–447.
[8] Dahlgaard, J.J., Kristensen, K. and Kanji, G.K. (1998), Fundamentals of TQM, Carfax, London.
[9] Deming, W.E., (1982) "Quality Productivity and Competitive Position" Massachusetts Institute of Technology, Cambridge.
[10] Dettmer, H.W., (1995) "Quality and the theory of constraints", Quality Progress, Vol. 28 No. 4, pp. 77-81.
[11] Ehie, I. and Sheu, C. (2005), "Integrating Six-Sigma and theory of constraints for continuous improvement: a case study", Journal of Manufacturing Technology Management, Vol. 16 No. 5, pp. 542-53.

©IJRASET: All Rights are Reserved

[12] Goldratt, E.M., (1990), "What is This Thing Called Theory of Constraints and How Should it be Implemented?", North River Press, New York, NY.
[13] Goldratt, E.M., (1994), "It's Not Luck", Gower, England.
[14] Gupta, S., (1997) "Supply Chain Management in Complex Manufacturing", IIE Solutions, vol. 29(3), 18-21.
[15] Hellstenu & Klefsjob,. (2000) TQh.1 as a management system consisting of values, techniques and tools, The TQM Magazine, 12, No. 4, pp. 238-244.
[16] Hakes, C. (1991), Total Quality Management: The Key to Business Improvement, Chapman and Hall, London, UK.
[17] Ho, S.K.M. and Fung, C.K.H. (1994), "Developing a TQM excellence model", The TQM Magazine, Vol. 6 No. 6, pp. 24-30.
[18] Issac, G., Rajendran, C. and Anantharamanan, R.N. (2004), "A conceptual framework for total quality management in software organizations", Total Quality Management, Vol. 15 No. 3, pp. 307-44.
[19] Jun, M., Shaohan, C., & Peterson, R. (2004). Obstacles to TQM implementation in Mexico's Maquiladora industry. Total Quality Management & Business Excellence, 15(1), 59–72.
[20] Koh, S. C. L., Demirbag, M., Bayraktar, E., Tatoglu, E., & Zaim, S. (2007). The impact of supply chain management practices on performance of SMEs. Industrial Management & Data Systems, 107(1), 103–124.
[21] Mabin, V.J. and Balderstone, S.J., (2000) The World of the Theory of Constraints: A Review of the International Literature, (CRC Press LLC: Boca Raton, FL).
[22] Moorthy, M. K., Tan, A., Choo, C., Wei, C. S., Tan, J., Ping, Y., & Leong, T. K. (2012). A Study on Factors Affecting the Performance of SMEs in Malaysia. Academic Research in Business and Social Sciences, 2(4), 224–239.
[23] Meredith, J., Mantel, S., 2003. Project Management: A Managerial Approach. Wiley, New York, NY.
[24] Perez, J.L. (1997), "TOC for World Class Global Supply Chain Management", Computers and Industrial Engineering, vol 33(1-2), 289-293.
[25] Nair A (2006). Meta-analysis of the relationship between quality management practices and firm performance – implications for quality management theory development. Journal. Operations. Management. 24: 948-975.
[26] Pyzdek, T., (2003), The Six Sigma Handbook: A Complete Guide for Green Belts, Black Belts, and Managers at All Levels. McGraw-Hill, New York, NY.
[27] Rahman, S. (1998), "Theory of constraints: a review of the philosophy and its applications", International Journal of Operations & Production Management, Vol. 18 No. 4, pp. 336-56.
[28] Snee, R.D. (2004), "Six Sigma: the evolution of 100 years of business improvement methodology", International Journal of Six Sigma and Competitive Advantage, Vol. 1 No. 1, pp. 4-20.
[29] Sila, I., (2007) "Examining the effects of contextual factors on TQM and performance through the lens of organizational theories: an empirical study", J. Oper. Manag. 25, 83–109.
[30] Salenga, G. and Fazel, F, (2000), "Obstacles to implementing TQM", Quality Progress, Vol. 33 No. 7, Milwaukee, July,pp 53-64
[31] Salegna, G., & Fazel, F. (2000). Obstacles to implementing quality. Quality Progress, 33(7), 53–57.
[32] Schonberger, R.J., (2008). Best Practices in Lean and Six Sigma Process Improvement. John Wiley & Sons, Hobokon, NJ.
[33] Soltani, E., Wilkinson, A., (2010) "Stuck in the middle with you: the effects of incongruency of senior and middle managers' orientations on TQM programmes". Int. J.Oper.Prod.Manag, Vol. 30,365–397.
[34] Srikanth, M.L., and Umble, M.M., (1997). Synchronous Management: Profit Based Manufacturing for the 21st Century Volume 1, The Spectrum Publishing Company: Guilford.
[35] Salaheldin, S. I. (2009). Critical success factors for TQM implementation and their impact on performance of SMEs. International Journal of Productivity and Performance Management, 58(3), 215–237.
[36] Spencer, M.S. & Cox, III, J.F., (1995) "Optimum Production Technology (OPT) and the Theory of Constraints (TOC): Analysis and Genealogy", International Journal of Production Research, vol 33(6), 1495- 1504.
[37] Sadikoqlu, E., Zehir, C., (2008) "The impact of contextual factors on total quality management practices". J.Glob.Strateg.Manag.4,88–97.
[38] Watson, K.J. & Polito, T., (2003) "Comparison of DRP and TOC Financial Performance within a Multi-Product, Multi-Echelon Physical Distribution Environment", International Journal of Production Research, vol 41(4), 741-765.
[39] Watson, K.J. & Patti, A., (2008) "A comparison of JIT and TOC buffering philosophies on system performance with unplanned M/C downtime", International Journal of Production Research, vol. 46(7), 1869- 1885.
[40] Zhao, X., Yeung, A.C.L., Lee,T.S.,(2004) "Quality management and organisational context in selected service industries of China.", J. Oper. Manag. 22,575–587.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)

View publication stats

Anomalous Object Tracking in Distributed Camera Network

MD Shahbaz Khan, Dr Indu Sreedevi
Delhi Technological University
New Delhi, India
mdshahbaz_bt2k16@dtu.ac.in, s.indu@dtu.ac.in

***Abstract*—The paper introduces a novel framework for real-time tracking of an object with higher precision in a Pan-Tilt-Zoom (PTZ) camera network. In the above-mentioned framework, the object which behaves anomalously is picked for tracking. An SVM classifier has been used to pick anomaly behaving object among all the objects present in a frame. It is a self-initializing algorithm as it does not require user intervention for object detection. Target detection is followed by its autonomous tracking in the distributed camera network. Hence it does not make the system bulky and takes less time to execute. The paper implements multiple object tracking using Discriminative Correlation Filtering with Channel and Spatial Reliability(CSR-DCF) tracking algorithm. Existing works are based on Particle filtering or Kalman filtering algorithms which are computationally complex and not self-initializing.**

I. INTRODUCTION

An efficient and fast algorithm is required for tracking a specified object from a frame having many objects using distributed PTZ camera based network. For each frame, some objects are at a higher priority for tracking than the rest. Although it is possible that none of the objects needs to be tracked in a particular sequence of frames. In such cases, tracking is not initialized till an anomaly behaving object is detected. Act of running, bending or picking was considered as anomalous behaviour. For detecting an anomaly behaving object, SVM based classifying model is trained. The distance of centroid of an object from its contour was used as a feature for developing the SVM model. This leads to the enhancement of the computational efficiency of the proposed algorithm. Sink frame of a particular camera in the network serves as the source frame for the other. For the mentioned decision-making task, the paper proposes a real-time, computationally efficient algorithm in which the camera having sink frame generates a histogram and is transmitted to all the adjacent cameras. Other cameras generate a histogram for their respective source frame and compare it with the received one. Camera having the same histogram starts tracking the object. Use of histogram formation and their comparison results in a smooth tracking at the time of object handover from one field of view (FOV) to another. Real-time and computationally fast algorithms are required for surveillance of densely populated public places using distributed PTZ cameras. The proposed approach assumes overlapping FOV between adjacent cameras in the network with the same resolution of the sink and the source frame of the neighbouring cameras.

II. LITERATURE LAYOUT

Several algorithms have come up to make object tracking autonomous with higher precision and real-time. However, on increasing the number of cameras in the network, their computational efficiency decreases manifold. Algorithms like [1],[2] do not consider the formation of the shadow of objects while few other algorithms like [3] are not suited for public places where the number of objects is more than an ordinary scenario. Current works in this field [3] attempts to track the object using particle filtering algorithm along with user intervention. As far as places with larger crowds are concerned, particle filtering algorithms are not efficient and do not yield real-time processing of data. CSR-DCF based tracking is a multi-object tracking algorithm and hence it can handle a large number of objects in crowd places. Paper [2] deals in Kalman filtering based object tracking, which is not efficient for real-time object tracking. In the case of Kalman filtering, noise is measured by modelling it as additive white Gaussian noise (AWGN). This makes the noise measurement in each frame as a joint Gaussian distribution. This, in turn, makes the object tracking vague when there is a bifurcation in the path of the object. This paper implements CSR-DCF tracking algorithm which is one of the best available real-time object tracking algorithm as discussed in table II. In papers like [1], object tracking is initialized by detecting the object in the very first frame by using the user interface. This, in turn, binds the algorithm from being fully autonomous. To eliminate this bottleneck, our approach proposes the concept of self-initialization of the object. Self initialization can be defined as the detection of an object of interest by the system itself. This paper implements self-initialization by choosing the object of interest as per its anomalous behaviours among all the objects present in that frame. The anomaly behaving object is picked using the SVM model. The model has been trained by using the distance of centroid of the object from its contour. No sooner did the object is detected, then its tracking gets started in all the cameras present in that distributed camera network. This algorithm proves to be efficient for the crowded region as it can track the movement of a suspicious object by itself. Each camera in the network tracks the object and communicates with adjacent cameras when the object of interest leaves its FOV. This novel method of object handover eliminates the chance of object missing during tracking in a

multi-camera system and hence yields a higher accurate tracking.

III. Problem Statement

The paper aims to find the solution for the detection of anomaly behaving object followed by its tracking in a distributed camera network. The intent is to improve the tracking accuracy and precision in comparison with probability and Gaussian distribution-based tracking algorithms[3] and [2] like particle filtering and Kalman filtering. The proposed methodology also eliminates random id generation technique[3] used to track objects in multiple cameras. Moreover, the proposed algorithm of object handover reduces the probability of object missing during its tracking in a multi-camera network system.

IV. Anomalous Behaviour Detection

The prerequisite for this algorithm is the minimum number of objects to be present in the camera frame which is achieved by classifying objects as normal and abnormal behaving groups. Activities like running, bending, or the act of picking are classified as anomalous behaviour in this paper as shown in figure 5 & 8. The paper proposes Principal Component Analysis (PCA) based feature extraction followed by its classification as anomalous behaviour using Support Vector Machine (SVM) classifier as discussed in [4] and the result has been collected in the table I. To prepare an SVM model, the maximum distance of any point on the contour from the centroid of the object was calculated and was used as features for SVM classier.

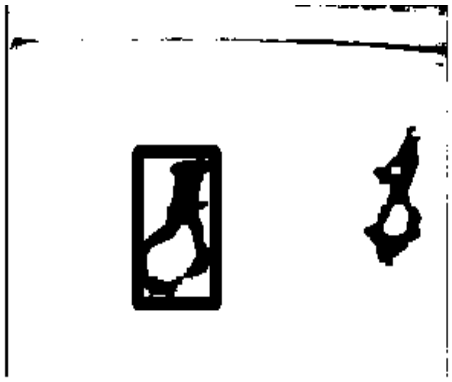


Fig. 1. A person running classified as anomalous behaviour

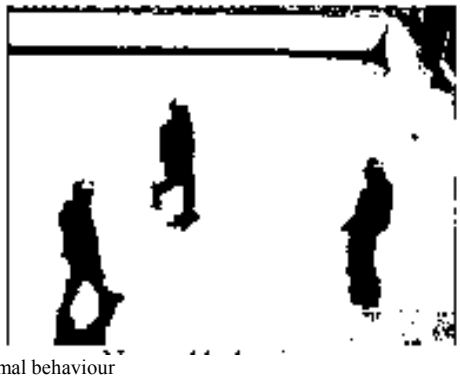


Fig. 2. Normal behaviour



Fig. 3. A person bending in the pathway – classified as anomalous behaviour.

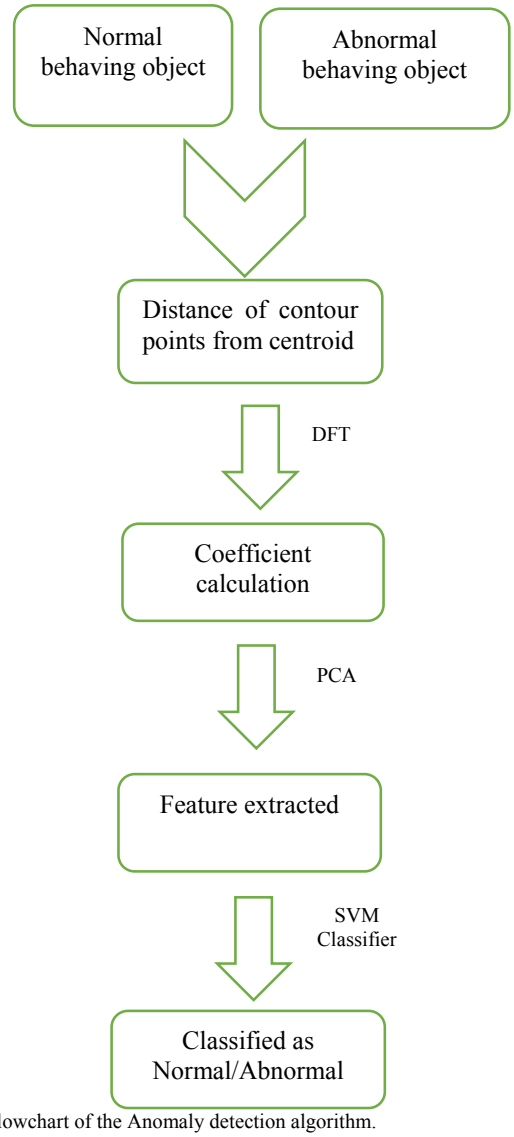


Fig. 4. Flowchart of the Anomaly detection algorithm.

First of all, the contour of every object was formed. Then using the below-mentioned formula, the centroid of that object was obtained.

$$x_c = \frac{1}{N}\sum_1^N x_i \quad (1)$$

$$y_c = \frac{1}{N}\sum_1^N y_i \quad (2)$$

where x_c, y_c are the coordinates of the centroid of contour.

For each point on contour, Euclidean distance was calculated and was transformed to coefficients by using DFT (Discrete Fourier Transform), which was later used for feature extraction using PCA (Principal Component Analysis).

A. PCA Based Feature Extraction

To extract features from each frame of camera input, PCA algorithm was implemented. The training dataset was divided into two groups namely G1, G2 each having 'n' training samples. Eigenvector of features was generated as follows: -

1) For each group of the dataset, a matrix '*X*' was formed having 'n' rows of samples and 'c' columns of features.

2) From the feature matrix '*X*', covariance matrix 'C' was computed as

$$C = (X - X')^T (X - X') \quad (3)$$

3) For the computed covariance matrix, eigenvector and eigenvalues were determined using the following relation

$$CV = V\Lambda \quad (4)$$

Where V is eigenvector and Λ is an eigenvalue.
Out of 'c' columns of Λ, we picked 'k' number of columns thus reducing the feature vector size to n x k.

B. SVM Classifier

As per the statistical learning theory [5],[6] & [7], SVM is one of the most precise classifying algorithms. The paper implements SVM for classifying the extracted features into normal and abnormal behaviour by using separating hyperplane. The constructed hyperplane was localized by maximizing the separating margins and consequently, it led to the formation of nonlinear decision boundary in the training dataset.
By implementing SVM algorithm, an optimal hyperplane was found such that it divides the dataset into two groups completely. In the course of building the SVM model, a hyperplane is adjusted by maximizing the margin between data values and hyperplane. Margin maximizes by minimizing the cost function.
Where cost function $J(x, y, f(x))$ is defined as,

$$J(x, y, f(x)) = \begin{cases} 0, \text{if } y * f(x) \geq 1 \\ 1 - y * f(x), \text{else} \end{cases} \quad (5)$$

On building the SVM model for training data, it was tested on the testing dataset whose labels was known. The hyperplane was further adjusted to get the maximum accuracy of classification. At the end of testing, the SVM model was used for unknown data.

Anomaly Behaviour Detection Algorithm

Prerequisites: Continuous video dataset, the dataset must contain anomaly behaviour, the dataset should contain video of the different timing of day having varying brightness.

Behaviour Detection:
1. Dividing the dataset into training and testing data group in a ratio of 1:4.
2. Extracting features from training dataset.
3. Optimizing number of features required to detect behaviour using PCA.
4. Adjusting the location of hyperplane by minimizing the cost function using SVM.
5. Applying the trained model to the testing dataset and evaluating the precision percentage.
6. If the precision is not more than 85% (as depicted in table I), then the number of data example in training dataset was increased and step 6 was repeated.
7. Else proceed to apply the trained model to an unknown dataset.

V. CSR-DCF ALGORITHM

In this paper, we have applied background subtraction to our input data before detecting anomaly behaviour. For the abovementioned task, background segmentation algorithm (based on Gaussian Mixture) was used, as it flexible for diurnal changes and for shadow detection as well. This algorithm is based on two research papers [8],[9] given in the year 2004 and 2006 respectively. On detecting an abnormal behaving object in a frame, the tracking algorithm starts executing. A rectangle bounding box is formed around that region of interest with a different colour for different objects in each frame. CSR-DCF algorithm was implemented for tracking objects which is a real-time processing algorithm with higher precision.
As discussed in [10], CSR-DCF is a tracking algorithm which provides learning with higher efficiency and better integration of filter updates and tracking. In this algorithm, spatial reliability map is used which helps to adjust filter support to enhance the accuracy of tracking. Similarly, reliability scores are calculated to check the quality of filters and thus helps in evaluating coefficients in localization.
This paper proposes a novel framework of object tracking based on discriminative correlation filtering followed by channel and spatial reliability scores[11]. It is a proximity-based algorithm which learns a set of filters to represent the object and uses them to find the highest correlated point in the adjacency of tracking an object from the previous frame.

A. Spatially Correlation Filters

With the help of the following approach, we comprehend the importance of learning correlation filters. For a given set of 'n' channels with features f = $\{f_k\}_{k=1:n}$ and target templates h = $\{h_k\}_{k=1:n}$, and the position of the object is determined by locating the highest frequency of ĝ(h),

$$\hat{g}(h) = \sum_{k=1}^{n} f_k \star h_k \quad (6)$$

where $\star$ is circular convolution operator & $\hat{g}(h)$ is correlation response.

Similarly, to find optimum correlation filter h, we have to minimize

$$\mathcal{E}(h) = \|\hat{g}(h) - g\|^2 + \lambda\|h\|^2 \quad (7)$$

where g is the expected output.

To find the efficiency of tracking, we transformed it into Frequency Domain as

$$\mathcal{E}(h) = \| \sum_{k=1}^{n} diag(\hat{f}_k)\hat{h}'_k - \hat{g}' \|^2 + \lambda\|h\|^2 \quad (8)$$

And on minimizing the above equation we get,

$$\hat{h}_k = (diag(\hat{f}_k)\hat{g}') \odot^{-1} (\sum_{k=1}^{n} diag(\hat{f}_k)\hat{f}'_k + \lambda) \quad (9)$$

where $\odot^{-1}$ is element-wise division operator and the above result can be used to design correlation filter trackers.

On considering each channel to be independent, we get the cost function as

$$\mathcal{E}(h) = \sum_{k=1}^{n} \|f_k \star h_k - g\|^2 + \lambda\|h\|^2 \quad (10)$$

To scale, we introduce a factor depending upon the discriminative power of feature channels which is termed as channel weights $w = \{\hat{w}_k\}_{k=1:n}$ and whose correlation response is given by,

$$\hat{g}(h) = \sum_{k=1}^{n} f_k \star h_k \cdot \hat{w}_k \quad (11)$$

B. Correlation Filter Learning

As mentioned above, it has been assumed that the correlation filter is independent of channels, hence it allows to consider the case of one filter only. The constraints are formalized as following

$$h_c \; m \odot h \equiv 0 \quad (12)$$

where h_c is defined as the dual variable with the above constraints.

On using Lagrangian augmentation [9], we find the following outcome

$$s\mathcal{L}(\hat{h}_c, h, \hat{I}|m) = \|diag(\hat{f}_k)\hat{h}'_k - \hat{g}\|^2 + \lambda\|h_m\|^2 + [\hat{I}^H(\hat{h}_c - \hat{h}_m) + \hat{I}^H(\hat{h}_c - \hat{h}_m)'] + \mu\|\hat{h}_c - \hat{h}_m\|^2 \quad (13)$$

Here in the above equation, we have introduced a new term μ which is the constraint penalty value and its value is updated by the following equation.

$$\mu^{i+1} = \beta\mu^I \quad (14)$$

C. Spatial Reliability Map

On locating the target, the region of interest is extracted from the frame and is used for updating the filter. The constrained filter uses spatial reliability map m_p, which is the collection of pixels of training model matching with that of the testing mode. To determine m_p, we have designed the following algorithm.

While the algorithm is getting executed for tracking, f oregr ound, and background of the object is stored in colour histogram c. Let $v_i = [y_i^c, y_i^x]$ is the observation i.e., it denotes the colour of y_i^c along with the y_i^x position for each pixel and m_p = random variable indicating the unknown foreground/background. For y_i to be observed on output, its joint probability was found as follows

$$\begin{aligned} p(y_i) &= \sum_{j=0}^{1} p(y_i | m_i = j)p(m_i = j) \\ &= \sum_{j=0}^{1} p(y_i^c | m_i = j)p(y_i^x | m_i = j)p(m_i = j) \end{aligned} \quad (15)$$

Algorithm 1 CSR-DCF

Prerequisites: Image I_i, the position of an object in previous frame pos_{i-1}, scale k_{i-1}, filter h_{i-1}, image histograms c_{i-1}, channel reliability w_{i-1}.

Determination and approximation of scale:-

1. Locating target at pos_i : finding the highest frequency in a correlation between filter (h_{i-1}) and extracted features (f) at pos_{i-1} position and adjusted by reliability scores (w).
2. Calculation of $\hat{w}$(det) (detection validation) with the help of per-channel responses.
3. Scale k_i was calculated, with the help of pos_i.
4. Colour histograms for foreground as well as the background were calculated and updated using $\hat{w}$(det) (detection validation).
5. Reliability map m_p was calculated using the above-mentioned relation.
6. Filter $\hat{h}$ was calculated and updated using m_p.
7. Using h, learning channel reliability was determined.
8. Calculate channel reliability from learning channel reliability & detection reliability.
9. Each filter hi was updated using the relation $h_i = (1-\eta)h_{i-1} + \eta\hat{h}$.
10. Channel reliability w_i was updated using the expression $w_i = (1-\eta)w_{i-1} + \eta\hat{w}$.

VI. Object Handover

In a distributed camera network, each camera is capable of processing the data and extracting information out of it. Cameras in a network can communicate with each other wirelessly in a range of 100m. The exchange of data between two cameras was done using OFDM modulation (Orthogonal

Frequency-Division Multiplexing). The advantage of using OFDM modulation is that it provides versatility in the presence of multiple channels.

The problem arises in handing over the object from one camera to the other. To eliminate the object handover problem, this paper implements the following algorithm:-

OBJECT HANDOVER ALGORITHM

Case 1: If the system has not yet detected any anomaly behaviour.
- i. In this case, each camera will keep on processing the input frames for anomaly detection.
- ii. If any anomaly behaviour is detected, then that camera will switch to case 2 else will continue to execute case1.

Case 2: if any anomaly behaviour is detected.
- i. Object tracking algorithm will be applied to the camera input in which anomaly action has been detected.
- ii. Camera tracking the anomaly behaving object will stop executing the anomaly detection algorithm for new objects.
- iii. As soon as the object becomes the sink for a camera, the camera generates its histogram from the last frame (say h1).

Case 3: when the object moves out of FOV of one camera
- i. In this case, the anomaly detection algorithm is paused and for each camera histogram of the new entering object is calculated (say h2).
- ii. Then both colour histograms (h1 and h2) are compared and the camera (say C2) whose histogram matches with h1, is now used for tracking.
- iii. The object tracking algorithm is applied to the camera (C2) input frame and case 2 is followed.

*Figures 6,7 and 9,10 illustrate the above-mentioned algorithm.

VII. EXPERIMENTAL RESULTS

The above experiment was performed on a dataset collected from Computer Vision Laboratory CVLAB ('https://cvlab.epfl.ch/data/data-pom-index-php/'). The mentioned dataset consists of inputs of 4 cameras located at different worldly points but with the same intrinsic parameters. On implementing the above-proposed algorithm, real-time object detection was done using anomaly behaviour detection technique with the following results

TABLE I. RESULT OF SVM CLASSIFIER

RESULT OF SVM CLASSIFIER		
Behaviour	***Success rate***	***No. of testing samples***
Normal	97%	297
Anaomaly	86%	328

The above-proposed algorithm was tested on the dataset at VOT2016 [12] benchmark which is comprised of 14 videos (7 training and 7 testing). On rigorous testing and comparison with various tracking algorithms, we got the following result (Table II).

TABLE II. COMPARISONS OF TRACKING ALGORITHMS

Tracking Algorithm	EAO	Aav	Fav	Total
Proposed method	0.34	0.510	0.851	-0.001
CCO Tracking	0.33	0.520	0.856	-0.006
TCN Net.	0.33	0.521	0.961	-0.110
SSA Tracking	0.32	0.531	1.044	-0.193
MLD Filter Based	0.31	0.482	0.833	-0.041
Staple	0.30	0.512	1.351	-0.539
DOC	0.29	0.513	1.233	-0.430
EB Tracking	0.29	0.441	0.902	-0.171
SRB Tracking	0.29	0.504	1.251	-0.547
STAPLEp	0.29	0.533	1.321	-0.498

*EAO – Expected Average Overlap
*A_{av} – Average Accuracy
*F_{av} – Average Failure
*Total = EAO + A_{av} - F_{av}

The above-mentioned comparison of various tracking algorithms was carried on the ground of expected average overlap, average accuracy and average failure. For an ideal tracking algorithm expected average overlap and average accuracy should be higher while lower average failure is preferred. In the case of real tracking algorithms, we need to compromise between these three parameters. To bring uniformity in comparison, a total usefulness function formed whose formula has been mentioned above. The higher total usefulness value for a tracking algorithm is, better that algorithm is to use. From the above table, we conclude that our tracking algorithm is better among the rest as its total usefulness value is higher than rest tracking algorithms.



Fig. 5. A person bending classified as anomalous behaviour by the camera C1.





Fig. 6. C1 detected object being tracked by the camera C2.





Fig. 7. Tracking object moving out of C2 camera being tracked by the camera C3



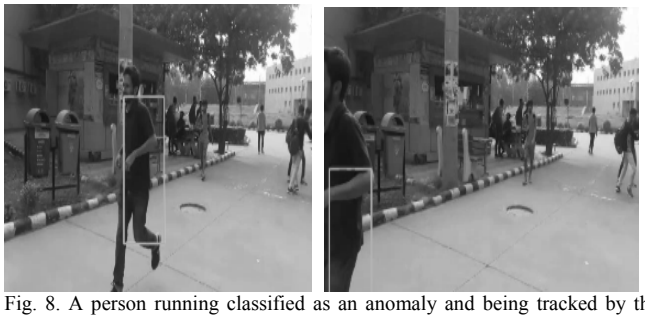


Fig. 8. A person running classified as an anomaly and being tracked by the camera C1.

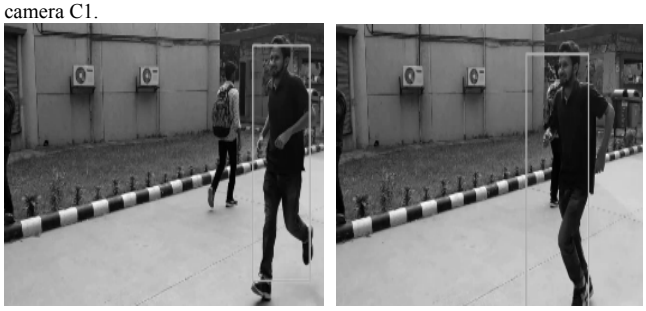




Fig. 9. C1 detected object being tracked by the camera C2





Fig. 10. Tracking object being moved out of C2 and tracked by camera C3

*Fig. 6-10 depicts the results of packing anomaly behaviour (running, bending and picking) using 3 cameras C1,C2 and C3.

VIII. Conclusion and future scope

This novel framework proposes an accurate and precise method for object tracking in distributed cameras. It is a real-time based approach with higher computational efficiency for tracking in a camera network. One of the important feature of this algorithm is self-initialization, which do not require user interventions. The proposed algorithm deals with the tracking of anomaly behaving objects using CSR-DCF based algorithm. Running, bending or picking has been classified as anomaly behaviour and based on this classification, objects are picked for tracking. For tracking the object, CSR-DCF algorithm has been implemented which is faster and more precise. Histogram comparison-based algorithm has been implemented to handover the object from one camera to the other in the network system.

Acknowledgement

I express my deepest gratitude to all those who helped me make this research complete and meaningful in all sense. First of all, I would like to thank our faculty in-charge, Prof. S. Indu & Dr Ayesha Chowdhary, whose contribution to this research work is incredible. They helped me to get through this research work by providing their valuable suggestions and sparing their precious time. Furthermore, I acknowledge the contribution of other staffs of the ECE department (DTU), who helped me by providing their valuable suggestions related to my topic. Last but not least, I am highly thankful to my parent and friends, who provided me moral support in times of distress and failure and played a vital role to make this work complete in all sense.

References

[1] G. R. Bradski, “Real time face & object tracking as a component of a perceptual user interface,” IEEE workshop on Application of Computer Vision, Princeton, NJ USA, 1998.

[2] S. Weng, C. Kuo and S. Tu, “Video Object Tracking using Adaptive Kalman Filter”, *Journal of Visual Communication & Image Representation*, 17(2006), 1190–1208.

[3] A. Choudhary, S. Sharma, I. Sreedevi and S. Choudhary, “Real-Time Distributed Multi-object Tracking in a PTZ Camera Network”, *Springer*, vol. 9127, 2015, pp. 183-192.

[4] X. Wu, Y.Ou, H. Qian, Y. Xu, “A detection system for Human Abnormal Behavior” RSJ International Conference on Intelligent Robots and Systems, Canada, 2005.

[5] K. Lee and Y. Xu, “Modeling human actions from learning” RSJ International Conference on Intelligent Robots and Systems, Japan, 2004.

[6] B. Scholkopf and A.J Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. Massachusetts: MIT Press, 2002.

[7] Y. Ou and Y. Xu, “Learning human control strategy for dynamically stable robots: Support Vector Machine approach” International Conference on Robotics and Automation, Taiwan, 2003.

[8] Z. Zivkovic, “Improved Adaptive Gaussian Mixture Model for Background Subtraction” ICPR, UK, 2004.

[9] Z. Zivkovic and F. Hejiden, “Efficient adaptive density estimation per image pixel for the task of background subtraction”, *Pattern Recognition Letters*, vol. 27, No. 7, 2006, pp. 773-780.

[10] A. Lukezic, T. Vojir, L. Zajc, J. Matasa and M. Kristan, “Discriminative Correlation Filter with Channel and Spatial Reliability”, *Internaltional Journal of Computer Vision*, 126(2017), 6309–6318.

[11] M. Kristan, R. Pflugfelder, A. Leonardis, J. Matas, F. Porikli, L. Cehovi, G., G. Fernandez and T. Vojir “The Visual Object Tracking, The Visual Object Tracking Challenge Results”, 2013, pp. 98-111.

[12] M. Kristan, et al, “The Visual Object Tracking VOT2016 Challenge Results”, Proc. European Conference Computer Vision, 2016.

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Arduino GSM based Power Theft Detection and Energy Metering System

Vivek Kumar Jaiswal
Department Of Electrical Engineering
Delhi Technological University
vivekjais01@gmail.com

Hritik Kumar Singh
Department Of Electrical Engineering
Delhi Technological University
singhhritik.singh5@gmail.com

Kuldeep Singh
Assistant Professor
Delhi Technological University
kuldeepsingh@dtu.ac.in

Abstract: Electrical energy has been considered as an essential form of energy. Whether it is Industrial need or household need, electricity is a must. Nowadays, with the increasing need for electricity, power theft is also increasing. This power theft is a severe problem like load scheduling will just become a hoax because of this Theft. Also, in developing countries like India, this power theft is 30%- 40% of total generated power and the electricity board has to provide this from their funds this results in major losses to electricity boards. This paper presents a way to detect power theft, alert the consumer, and cut off the supply accordingly. When unauthorized actions are detected, and a separate message is sent back to the microcontroller to remove the illegal supply, an SMS is sent automatically to the user via the GSM module. The GSM is interspersed with a unique system energy meter with a relay to handle the non-technical losses, problems with billing, and variations in voltage fluctuations

Keyword **– GSM, Atmega328, LM2596, Arduino GSM**

Introduction

Over the last few decades, communications technology has improved through connectivity and boundaries. The use of the "Embedded Program in Communication" provides various interesting applications. Electricity is now more than necessary. The demand for electricity is increasing continuously. With increasing electricity demand, power theft is also on the rise. Power theft is a concerning issue even in the most efficient power systems like in USA and a moderately efficient system like in Malaysia [1]. It has become important to improve the system performance and to avoid Theft. The purpose of this project is to design a system to monitor any power consumed before the power meter, and that is power pilferage. If some power theft is detected, that results in penalty or disconnection of energy supply to the respective consumer, and simultaneously an SMS is sent to the control unit. In the past, numerous non-technical and scientific methods for detecting electricity pilfering have been suggested. Non-technical procedures may involve customer inspection with a suspect load profile. Although regular inspections can help reduce theft, such action calls for the massive labour force and enormous effort. In most situations, this attempt always fails due to employee dishonesty. “The use of the main observer meter at the secondary distribution terminals are some of the technological ways to detect pilferage [3], harmonic generator [6], genetic support vector machines [7], extreme learning machine, power line impedance technique [8] [3].

However, these technical approaches can be effectively implemented only if proper communication is ensured between the central control station and the appropriate test points. Recently, energy meters based on GSM network have been proposed. These meters provide the energy metering and remote load control system facility [3]. In this paper, an energy metering scheme based on GSM is introduced to deal with various aspects of electricity theft. The proposed system avoids meter abuse and bypassing irregularities.

Existing Technology

The meter is manipulated to display low readability, bypass meter, and billing irregularities. Advanced distribution automation and dynamic price models relying on automated metering (AMR) and advanced measuring infrastructure (AMI) can overcome the billing irregularities and requirements for more employees [3]. However, this system has failed to benefit the atmosphere for energy theft and requires frequent inspection by NTL. Electrical cables, coaxial cables, optical fiber, infrared, Bluetooth are also used for current communication technologies but do not have reliability, communication speed, high installation costs, and maintenance are expensive and risky.

Methodology

For successful completion of this project, the following steps were carried out in a sequential manner

- Literature Reviewing: This includes exploring various documents and books regarding this topic.
- Problem identification: identifying the problem and its consequences.
- System modelling: Combining different elements to describe the system.
- System analysis and simulation: This indicates how the system behaves using proteus.
- Proposing Solution: designing of power theft control system to use power wisely.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:55:26 UTC from IEEE Xplore. Restrictions apply.

Proposed System

In this paper, the energy stealing system is proposed to send the SMS using a GSM Network. The most common methods of electricity theft are to bypass the phase line, cut off the neutral line, bypass the entire meter, and distort the energy meter. The current transformer is blinded to zero current, bypassing the phase line. The voltage transformer shows zero potential by disconnecting the neutral line and giving the local ground. The entire meter is bypassed, and the energy meter is measured at zero. Faulty readings are triggered by tampering the energy meter with unauthorized modification. The main contributions of this paper as follows

- The proposed system uses only a very small number of components and a simple electricity theft algorithm.
- The customer is warned initially after several warnings, the connection of the customer is being cut, and data is simultaneously sent to the control center
- Compared to other techniques discussed, total system costs are relatively low.
- It can be integrated with existing meters and new meters of production processes.

Many named and unnamed manners for theft of electricity are also available. The use of a single household meter cannot detect and control these types of theft. An observation meter is used in the proposed system to identify such unrecounted theft. Figure 1 shows the observer meter. Measure the overall household energy consumption at agreed time intervals. For example, outside an apartment building, this type of meter can be installed. In this case, all the apartments in this apartment building measure the energy consumed. The household meters communicate their measured units via SMS to the central meter. The central meter detects the theft of electricity when both values are significantly different. Since the observer meter can be read individually, households in which theft occurred can be calculated based on zero or low power consumption. The central meter informs the server of potential theft locations via SMS. The authority may send technical personnel to check the households reported and detect the dishonest consumer [3].

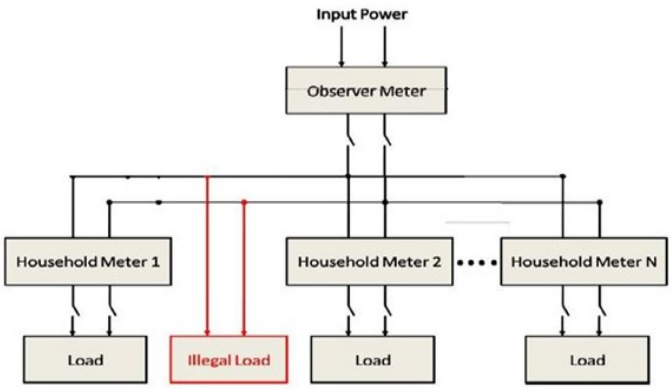


Figure 1: Block Diagram

Block Diagram

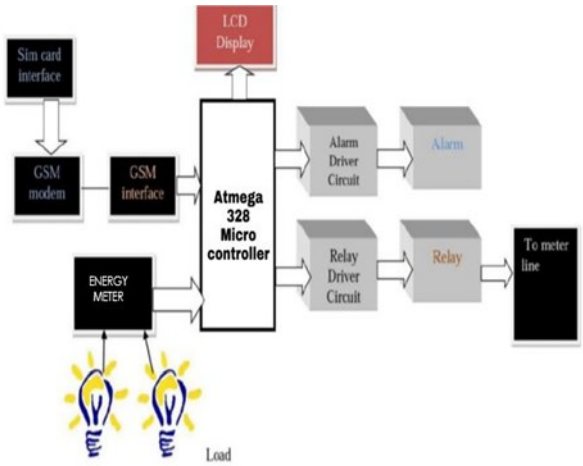


Figure 2: Proposed Architecture

The aims of the proposed program are mainly to monitor electricity theft through enhanced GSM communication features. Figure 2 shows the block diagram of the prototype. The most important components included power meter, GSM module [10], LM2596 voltage regulator, Arduino UNO [9], PC817 optocoupler, and LCD. The prototype model is a device that uses energy to link a variable load. The PC817 optocoupler and the CAL (calibration) energy meter LED is attached for the detection of consumed units. At 3200 imp/kWh the pulse rate of the power meter is. The LED blink rate is, therefore, directly proportionate to the load. To record consumed units, the output of the optocoupler is forwarded to the microcontroller.

Proposed Circuit

In the circuit first, a power supply block i.e. IC LM2596, which is made by TEXAS INSTRUMENT is introduced. This is used to get a constant 5v DC output, that is used to power 16*2 LCD, Microcontroller, Relay Module.

Data pins of the LCD display are connected to digital pins of the microcontroller (pin-6,7,8 & 9) which is used as output pins. GSM module is also interfaced with the microcontroller through digital pins 2 and 3(in the software part this is defined as transmitter and receiver pins). A buzzer is also connected to pin 11 of the microcontroller which functions in case of theft. Also, a relay is used to trip the load in case of theft, this is connected to pin 10 of the microcontroller.

Two energy meters are used one is named as master energy meter and other as a house energy meter. These meters are connected to pins A0 and A1(Analog input pins) of microcontroller through PC817(optocoupler). Optocoupler's input terminals are connected to calibration LED of Energy meter and output is connected to pins A0 & A1

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:55:26 UTC from IEEE Xplore. Restrictions apply.

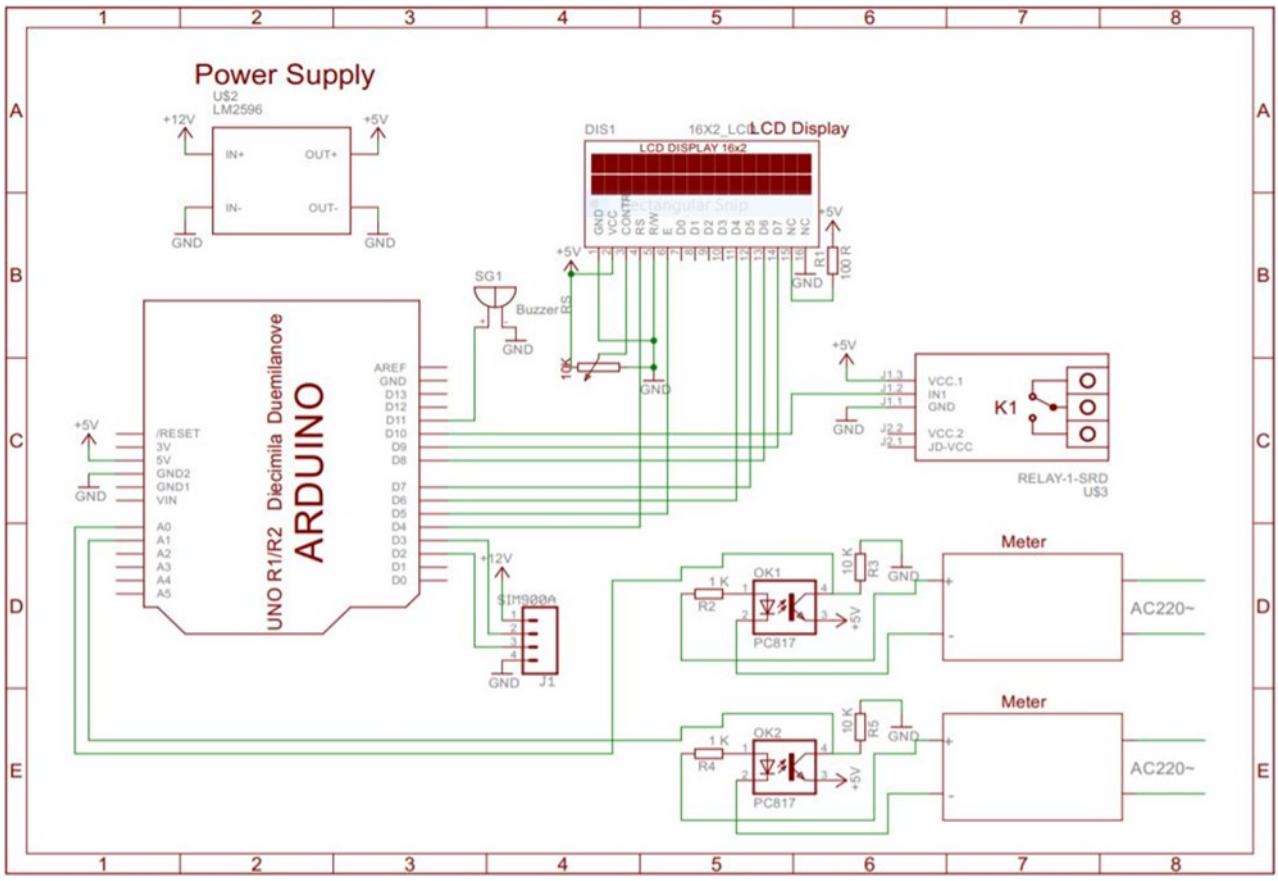


Figure 3: Circuit diagram

Hardware Design

The 230V AC power supply is given to energy meter and then connected to the load through the relay. The input of 5V for Arduino is given by a voltage regulator LM2596 from the 12V supply. The output pin of optocoupler PC817 is given to Arduino through A0 and A1. The interfacing of GSM and Arduino is made. The input 12V GSM is given through a 220 to 12V AC to DC adaptor. The corresponding message is sent through GSM.

Working

The power supply from the distributor is given to the system. This supply consists of 230 volts. The transmission lines are used to provide power supply to the consumer using electric poles at a particular distance. The system has an Arduino Atmega328, energy meter, and GSM module attached to it. Here the power supply is step down from the 230-volt power supply into 5 volts by the power supply circuit because our system uses a 5-volt power supply. The 230-volt power supply is stepped down 12 volts via a 12-volt step-down transformer so now the power supply will be 12 volts. It uses an LM2596 voltage regulator to get a 5-volt power supply which is given to our system. This 5-volt power supply given to an Atmega328 and GSM module. The system has Arduino, Energy meter, and GSM module. The master energy meter tracks the quantity of the energy units supply through it then it sends the unit consumed to Atmega328. Now Atmega328 stores this data in memory. As the supply reaches to slave energy meter it will count the number of energy units consumed. The data at the slave energy meter will be transfer to Atmega328. The Atmega328 will analyze the two data present. If the number of energy units reached varies. The Atmega328 will trigger the GSM module. The GSM module sends the SMS to the consumer about the amount of difference occurred in the transmission cable. The distributor will track the amount of loss happen and puts a penalty to the consumer. The distributer is informed about the energy loss happen to them by the system's action.

Programming

The code for the program is done in Arduino IDE and then it is burned into the Arduino. For the simulation part, Keil software is used to convert the program into .hex file. The programming logic followed is depicted in the flow chart.

978-1-7281-5371-1/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:55:26 UTC from IEEE Xplore. Restrictions apply.

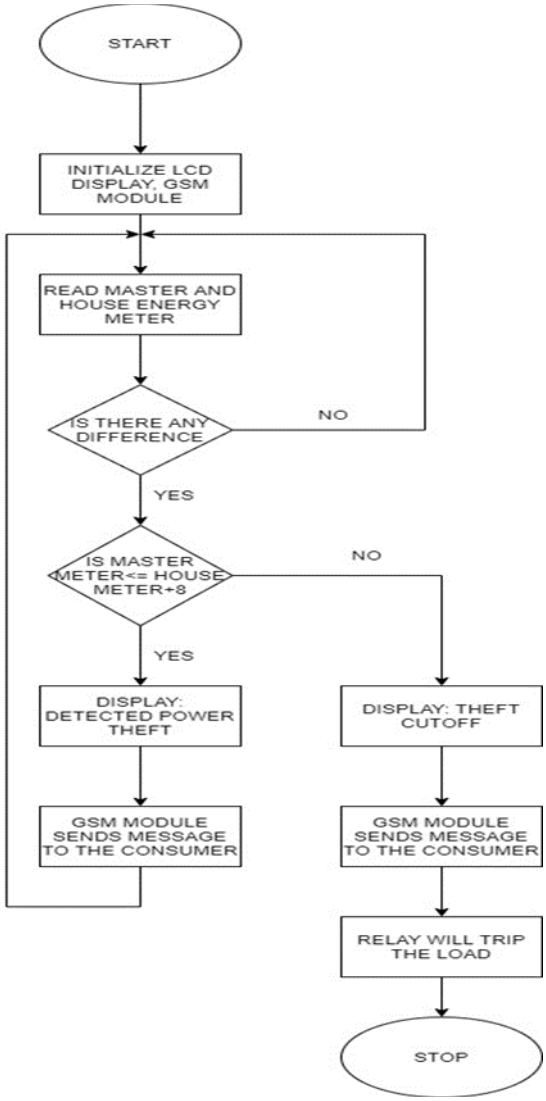


Figure 4 -programming flow chart

Simulation Results

Normal condition-Both the meter shows the same readings

NO ALERT as no theft observed.

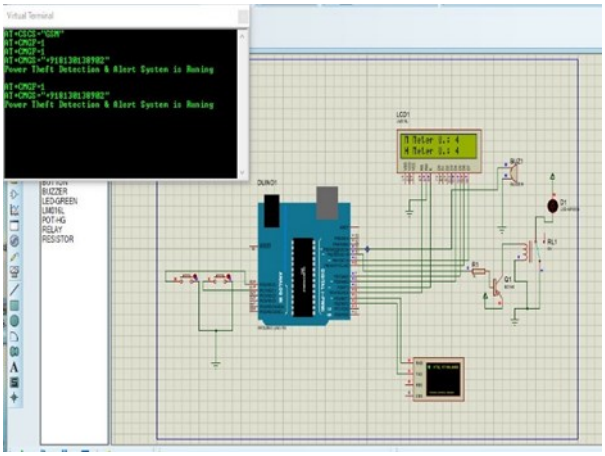


Figure 5-Normal condition

Alert stage: - When theft is observed. master meter reading and house meter readings have a difference of 3 units. alert message is sent in the message screen that power theft has been observed.

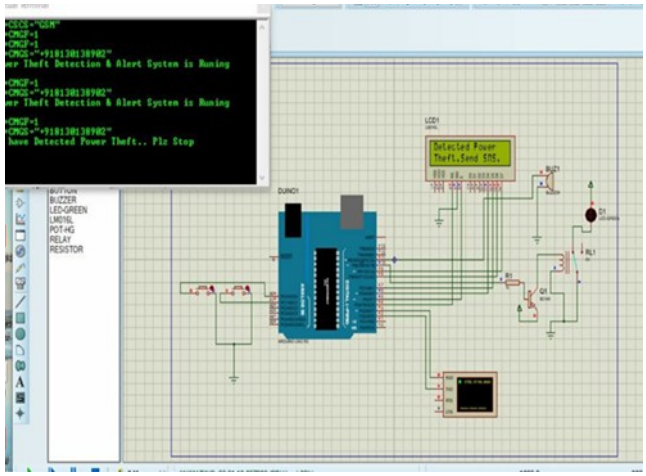


Figure 6-Alert condition

Cut off stage- when the difference between master meter reading and house meter reading is more than 8. Relay is tripped and the corresponding message of cut off is sent in the message screen.

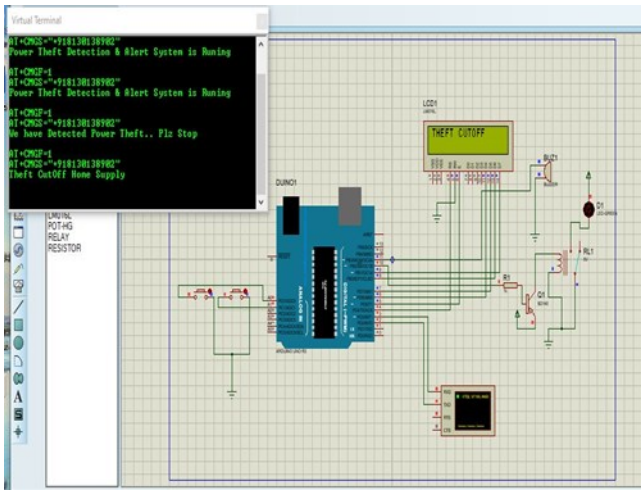


Figure 7-Cut-off condition

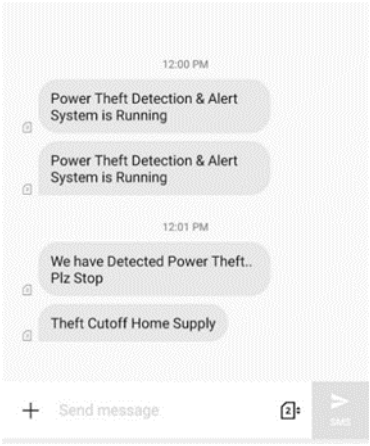


Figure 8 - Message Screenshot

Hardware Results

Under normal condition i.e. house meter reading is equal to master meter shown in figure-9. Alert stage: - when theft is observed. Master meter reading and house meter readings have a difference of 5 units, an alert message is sent inside the screen that power theft has been observed as shown in figure-10. Cut off stage: - when the difference between

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:55:26 UTC from IEEE Xplore. Restrictions apply.

master meter reading and house meter reading is more than 10. Relay is tripped and the corresponding message of cut off is send as shown in figure-11



Figure 9-Normal condition



Figure 10-Alert condition



Figure 11-Cut-off condition

Conclusion

The proposed method of power theft minimizes the heavy power and revenue loss that results from the theft of power by customers. With the completion of the project, it is concluded that power theft can be effectively reduced by detecting when does the power theft occurs and simultaneously informing the authorities to disconnect the power to the house or customer trying to steal the power supply by an integrated automatic breaker that is attached to the unit. The system is used to send or communicate data to the consumer domain as well through a GSM wireless network. The proposed system will reduce energy waste and save more for future use.

References

[1] T. B. Smith, "Electricity theft: a comparative analysis," Elsevier Journal Energy Policy, vol. 32, no. 18, pp. 2067-2076, Dec. 2004.

[2] K. Dineshkumar, Prabhu Ramanathan, Sudha Ramasamy. "Development of ARM processor based electricity theft control system using GSM network", 2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015], 2015

[3] N Mohammad,A Barua and M Arafat,"A smart prepaid energy metering system to control electricity theft", *in proceedings of International Conference on Power, Energy and Control (ICPEC)*, vol.562, no.565, pp.6-8, Feb.2013.

[4] Nitin K Mucheli, Umakanta Nanda, D Nayak, P K Rout, S K Swain, S K Das, S M Biswal. "Smart Power Theft Detection system", 2019 Devices for Integrated Circuit (DevIC), 2019

[5] M. Saad, M. F. Tariq, A. Nawaz and M. Y. Jamal, "Theft detection based GSM prepaid electricity system", *IEEE International Conference on Control Science and Systems Engineering (ICCSSE)*, pp. 435-438, 2017.

[6] S.S.S.R. Depuru, L. Wang, and V. Devabhaktuni, "A conceptual designusing harmonics to reduce pilfering of electricity," in proceedings of IEEE Power and Energy Society General Meeting, pp. 1-7, Jul. 2010.

[7] J. Nagi, K.S. Yap, S.K. Tiong, S.K. Ahmed, and A.M. Mohammad,"Detection of abnormalities and electricity theft using genetic Support Vector Machines," in proceedings of IEEE Region 10 Conference, pp. 1-6, Nov. 2008.

[8] A.Pasdar and S.Mirzakuchaki, "A solution to remote detecting of illegalelectricity usage based on smart metering," proceedings of 2nd International Workshop on Soft Computing Applications, pp. 163-167, Aug. 2007.

[9] ATMEL website.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:55:26 UTC from IEEE Xplore. Restrictions apply.

OPSEARCH
https://doi.org/10.1007/s12597-020-00464-7

THEORETICAL ARTICLE

Best criteria selection based PROMETHEE II method

Anjali Singh[1,2] · **Anjana Gupta**[2] · **Aparna Mehra**[3]

Accepted: 8 July 2020
© Operational Research Society of India 2020

Abstract

In recent years, PROMETHEE II has emerged as one of the most efficient and successfully applied outranking method in multi-criteria decision making (MCDM) problems. Conventional PROMETHEE II method presumes that the weights of the criteria are known beforehand, and this is rather a stringent assumption. Moreover, the number of preference indices to be evaluated in the method grows enormously with an increase in the number of alternatives to be ranked. In this work, we look at these concerns and propose a methodology inspired by PROMETHEE II that identifies the best criteria first and then, the preference indices are computed with respect to these criteria only. The algorithm is customized in such a way that it curtails the computational time complexity of PROMETHEE II to a great extent. Here, a large number of alternatives-criteria evaluation matrices are simulated with random numeric data to illustrate the proposed method and evaluate its performance. Statistical analysis on the stability of the rankings and the conformity measure proves that the proposed approach may be preferred over PROMETHEE II and TOPSIS methods. Rank correlation coefficients are evaluated, and they indicate the efficacy of the proposed method.

Keywords Multi-criteria decision making · PROMETHEE II · TOPSIS · Stability degree · Conformity metric · Kendall's tau-b rank correlation coefficient

Mathematics Subject Classification 91B06 · 90B50

✉ Anjana Gupta
anjanagupta@dce.ac.in

[1] Department of Mathematics and Humanities, Mahatma Gandhi Institute of Technology, JNTU, Hyderabad, Telangana 500075, India

[2] Department of Applied Mathematics, Delhi Technological University, Delhi 110042, India

[3] Department of Mathematics, Indian Institute of Technology, Delhi 110016, India

Published online: 22 July 2020

1 Introduction

MCDM problems involve analysing the subjective assessments of a finite number of alternatives under a certain predetermined, finite number of criteria by an expert/decision-maker (DM) or a group of experts/decision-makers (DMs). In the latter case, it is called a multi-criteria group decision making (MCGDM) problem, see[1]. Over the last few decades, numerous MCDM methods have been proposed under different set-ups for selecting the best alternative among the available ones. For ready reference, see[2–5]. The advancement in MCDM methods is stimulated not only by different applications to real-life problems but also by researchers' inclination towards introducing extended decision-making methodologies as discussed in[6].

1.1 Research issues and need for the present work

Conventional PROMETHEE (Preference Ranking Organization Method for Enrichment Evaluations) method is one of the most successfully practiced MCDM method. The technique was designed by Brans[7] and extended by Vincke and Brans[8]. The PROMETHEE II version is a protracted method that requires a priori knowledge of the criteria weights. It presumes that the decision-maker is a well-versed expert in the domain of the problem to decide upon these weights, reflecting the importance of criteria in ranking the alternatives. This imparts a subjective character to criteria weights varying which alter the ranking of the alternatives. Another concern is that the computational efforts in PROMETHEE II method increase with the number of alternatives and criteria. For a given decision-making problem with 'm' alternatives and 'n' criteria, one needs to evaluate '$mn(m-1)$' preference indices. This paper combines these two interrelated issues. One it noticed that these many computations are necessarily not required and another that the prior knowledge of the criteria weights can be dropped from the PROMETHEE II method.

1.2 Key differences between closely related studies and the work presented in this paper

The significant differences between the proposed work and closely related studies available in the literature are as follows:

1. In literature[9–11], the conventional PROMETHEE II method and its extensions are framed to evaluate '$mn(m-1)$' preference indices for a given decision making problem with 'm' alternatives and 'n' criteria. In this study, we have customised the existing procedure in such a manner that the reckoning process of preference indices lowers to a great extent.
2. The classical PROMETHEE II method presumes the weights of the criteria beforehand. In the available literature, various techniques like AHP, Entropy[12, 13], etc. are also used to evaluate the weight vector of the criteria. Although,

many researchers have designed various tools like weight stability intervals, walking weights, PROMETHEE VI[14, 15], etc. to study the repercussion of these values, but it brings subjectivity in the decision making process that can change the ranking of the alternatives. So, in this paper, we have first recognized the best criterion(a)/feature(s) for each alternative and then assign the uniform weight to the best criteria and no importance i.e., zero weight to the vestigial criteria. Such procedure of weight assignment to criteria is anomalous as it involves both the alternatives and criteria and vary with each alternative in an MCDM problem. Even after the withdrawal of prerequisite of the prior knowledge of weight vector, the statistical discussion given in Sect. 5 reveals that the suggested scheme performs efficiently. The ranking obtained by the proposed methodology is highly correlated as well as more stable than the classical PROMETHEE II ranking. Also, this practice dissolves the element of subjectivity from the criteria importance in an MCDM problem.

3. The computational complexity of an algorithm is also an indicator of its efficiency. The traditional PROMETHEE II has computational time complexity of $O(m^2n)$. Recently, Calders and Assche[16] suggested the method that curtails the time complexity of PROMETHEE with linear preference function to $O(mn\log(m))$. In our study, the computational time complexity of the proposed approach is of $O(mn)$ that makes the performance of the algorithm remarkable.
4. The statistical analysis is an essential and significant element of any methodology to prove its efficiency. In view of statistical studies discussed in Sect. 5, the ranking of alternatives obtained by the proposed algorithm is not merely highly correlated with the classical PROMETHEE II and TOPSIS methods' ranking but also more stable for 10 distinct pairs of numbers of alternatives and criteria. The conformity values also assert the analogous efficacy of proposed algorithm and PROMETHEE II. The conformity of designed method may be agreeing the traditional PROMETHEE II but the reduced complexity of the suggested scheme is the highlight of the work.

1.3 Focus of the present study

In this paper, we propose a methodology inspired by PROMETHEE II, which deliberates on first identifying the best criteria for each alternative and then evaluates the preference indices for these criteria only. The procedure is customised in such a way that the number of preference indices to be estimated reduces to a greater extent. Moreover, the method proposes to assign uniform weights to the best criteria and zero weight to the remaining criteria with each alternative. In this way, no single weight vector is attached to the criteria; instead, the weight vectors depend both on the alternatives and the criteria and vary with alternatives in an MCDM problem. The algorithm is contrived in such a manner that it also curtails the computational time complexity of the classical PROMETHEE II to a noteworthy extent.

The suggested method is validated by randomly generating a large number of alternatives-criteria evaluation matrices which are also used to compute stability degree and conformity measure of the method. To demonstrate the efficiency of the

designed algorithm, the study includes PROMETHEE II and TOPSIS methods in the statistical analysis which exhibits that the proposed method is comparable with the existing methods and can be opted by the practitioners for real-life decision-making problems.

1.4 Organization of the paper

The organization of remainder of this paper unfolds as follows. Section 2 briefs the literature of the two MCDM methods - PROMETHEE II and TOPSIS. Section 3 describes the proposed method. Section 4 presents an example by simulating alternatives-criteria evaluation matrix to illustrate the proposed method. Section 5 discusses the analysis and efficiency of the method via stability degree, conformity measure and rank correlation coefficient. Section 6 concludes the paper.

2 Literature survey

The outranking methods aim to build a preference relation, called an outranking relation, through a series of pairwise comparisons of alternatives to be evaluated on criteria. The study of the outranking method PROMETHEE commenced with the pioneer works of Brans and his collaborative in the early eighties. The PROMETHEE-I method for partial ranking and PROMETHEE II method for the complete ranking of the alternatives were proposed in the work of Brans[7]. The success of these two methods led to the development of more versions of PROMETHEE method, namely, PROMETHEE III, IV, and V. Human brain representation was incorporated in PROMETHEE VI method, proposed by Brans and Mareschal[15] followed by PROMETHEE GDSS for group decision-making, introduced by Macharis et al. [17] and the visual interactive module GAIA (Geometrical Analysis for Interactive Aid) for graphical representation, given by Mareschal and Brans[18, 19]. Moreover, Figueira et al.[20] introduced two extended approaches: PROMETHEE-TRI for handling sorting problem and PROMETHEE CLUSTER for nominal classification. All these versions and extensions of PROMETHEE have been applied in many fields, including environment management[9, 11], business and financial management[10, 21], chemistry[22, 23], energy management[24, 25], social management[26, 27], and many more. For a comprehensive survey of PROMETHEE methods see[28]. Among all the methods of this family, PROMETHEE II is the most acclaimed and frequently used one.

Among numerous other MCDM methods, TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) is a compensatory method that determines the solution which is at the shortest distance from the positive ideal solution and the farthest distance from the negative ideal solution of the MCDM problem. It has been often used across different application areas since its inception by Hwang and Yoon[29]. TOPSIS has been successfully applied in supply chain management and logistics[30, 31], health, safety and environment management[32, 33], human resources management[34, 35], energy management[36, 37], and many more. One can refer to[29, 38] for a detailed

description and working of the method and to the work of Behzadian[39] for a comprehensive survey on the use of TOPSIS method in MCDM. We have briefly mentioned this method here for it has been used in the latter part of this paper for performance analysis and comparison. Both methods require a priori knowledge of criteria weights for aggregating the information in data for penultimate ranking of alternatives. This assumption makes the methods sensitive to the choice of criteria weights.

3 Proposed methodology

An MCDM problem involves m alternatives A_i, $i = 1, \ldots, m$, to be evaluated on n criteria C_j, $j = 1, \ldots, n$. Let $I = \{1, \ldots, m\}$ and $J = \{1, \ldots, n\}$ be the indices sets of alternatives and criteria respectively. Let an alternative-criteria evaluation matrix be $R = [r_{ij}]_{m\times n}$, where r_{ij} represents the evaluation of ith alternative on jth criteria. The proposed method involves the following steps:

Step 1: The classical PROMETHEE II method characterizes six types of preference functions to compare the alternatives under various cost and beneficial criteria. The provision of distinct preference functions makes the normalization process a choice in classical PROMETHEE II. In this paper, the pairwise difference is adopted as the preference function that requires the uniformity of all criteria. Hence, the normalization process becomes a fundamental requirement for the method.

Normalize the matrix R using linear normalization given as $R_{norm} = [\widetilde{r}_{ij}]_{m\times n}$, where

$$\widetilde{r}_{ij} = \begin{cases} \dfrac{r_{ij} - \min\limits_{i\in I}\{r_{ij}\}}{\max\limits_{i\in I}\{r_{ij}\} - \min\limits_{i\in I}\{r_{ij}\}} & \text{if } j\text{th criterion is a beneficial criterion;} \\ \dfrac{\max\limits_{i\in I}\{r_{ij}\} - r_{ij}}{\max\limits_{i\in I}\{r_{ij}\} - \min\limits_{i\in I}\{r_{ij}\}} & \text{if } j\text{th criterion is a cost criterion.} \end{cases}$$

The max-min normalization is preferred as the above process converts all criteria into the beneficial criteria that makes nature of all criteria uniform and it is also an essential requirement for the upcoming step.

Step 2: For a given alternative A_i, find the maximum value over all criteria

$$M_i = \max_{j\in J}\{\widetilde{r}_{ij}\}.$$

Let M_i be achieved on k_i number of criteria ($k_i \leq n$), say, $C_{\gamma_i(1)}, \ldots, C_{\gamma_i(k_i)} \in \{C_1, \ldots, C_n\}$; set $J_i = \{\gamma_i(1), \ldots, \gamma_i(k_i)\}$.

Step 3: For $j \in J_i$, evaluate the preference index $P_j(A_i, A_{i'})$, $i, i' \in I$, $i \neq i'$, as follows:

$$P_j(A_i, A_{i'}) = \begin{cases} 0 & \text{if } \widetilde{r}_{ij} \leq \widetilde{r}_{i'j} \\ \widetilde{r}_{ij} - \widetilde{r}_{i'j} & \text{otherwise.} \end{cases}$$

Here $P_j(A_i, A_{i'})$ is the preference index (value) of the alternative A_i over alternative $A_{i'}$ under criterion C_j. So, for the ith alternative, the preference indices $P_j(A_i, A_i')$, $j \in J_i$, are computed only under those criteria which correspond to the maximum evaluation of the alternative A_i.

The conventional PROMETHEE II method requires $mn(m-1)$ preference indices to participate in the procedure to rank m alternatives on n criteria whereas the designed method reduces it to evaluate only $((m-1)\sum_{i=1}^{m} k_i)$ number of preference indices.

Step 4: Compute the cumulative preference index $\pi(A_i, A_i')$, for $i, i' \in I$, $i \neq i'$, as follows:

$$\pi(A_i, A_i') = \frac{1}{k_i} \sum_{j \in J_i} P_j(A_i, A_{i'}).$$

The conventional PROMETHEE II method assumes the criteria weight vector beforehand. In the proposed method, since the criterion(a) having maximum value represents the best feature(s) of an alternative, hence, for each $i \in I$, we propose to attach the uniform weight $1/k_i$ with criteria in J_i, and weight zero with all remaining criteria in $J \setminus J_i$.

The concern of the prior knowledge of weight vector may be dropped as the statistical results given in Sect. 5 validate the efficiency and competency of the proposed algorithm. The ranking obtained by proposed algorithm is highly correlated as well as more stable than the classical PROMETHEE II ranking. It also makes the approach independent of the decision-makers' perspective on the criteria importance/weights.

Step 5: Determine the outgoing and the incoming flows described respectively as follows:

$$\phi^+(A_i) = \frac{1}{m-1} \sum_{\substack{i'=1 \\ i \neq i'}}^{m} \pi(A_i, A_{i'}), \quad i \in I,$$

$$\phi^-(A_i) = \frac{1}{m-1} \sum_{\substack{i'=1 \\ i \neq i'}}^{m} \pi(A_{i'}, A_i), \quad i \in I,$$

and, compute the net flow

Table 1 Computational time complexity of PROMETHEE II and designed methodology

Principle activity	PROMETHEE II	Proposed method
1. Normalization	$O(mn)$	$O(mn)$
2. Preference indices and cumulative preference indices	$O(mn(m-1))$	$O(m(m-1)+mn)$
3. Net flows	$O(m(m-1))$	$O(m(m-1))$
4. Sorting the net flows values to rank	$O(m\log m)$	$O(m\log m)$
Computational time complexity	$O(m^2n)$	$O(mn)$

$$\phi(A_i) = \phi^+(A_i) - \phi^-(A_i), \qquad i \in I.$$

Step 6: The alternatives are ranked with the largest value of $\phi(A_i)$ considered as the best alternative.

Remark 3.1 For a given $m \times n$ alternatives-criteria evaluation matrix, the computational time complexity of suggested scheme is $O(mn)$ which is indeed less than the complexity $O(m^2n)$ of the conventional PROMETHEE-II method. Table 1 depicts the same.

The time complexity of proposed method has also decreased to a noteworthy extent. Hence, the designed methodology abridges the complexity of PROMETHEE II method.

The flowchart given in Fig. 1 summarizes the proposed method.

4 Illustrative example

Consider an MCDM problem with 7 alternatives and 9 criteria, out of which, criteria C_1, C_2, C_3, C_6, C_8, C_9 are the beneficial criteria and criteria C_4, C_5, C_7 are the cost criteria. The 7×9 alternative-criteria evaluation matrix is populated with a random data from $U(0, 1)$ distribution to illustrate the proposed method. Let this matrix be R.

$$\begin{bmatrix} 0.2956 & 0.9715 & 0.5639 & 0.2711 & 0.1345 & 0.3183 & 0.3315 & 0.5127 & 0.6675 \\ 0.5696 & 0.6743 & 0.3813 & 0.104 & 0.0997 & 0.2724 & 0.4562 & 0.2391 & 0.4185 \\ 0.5231 & 0.1166 & 0.0101 & 0.5562 & 0.4602 & 0.0472 & 0.7755 & 0.876 & 0.9642 \\ 0.6077 & 0.9339 & 0.9153 & 0.3164 & 0.717 & 0.2123 & 0.8049 & 0.8063 & 0.238 \\ 0.4758 & 0.5839 & 0.8552 & 0.9422 & 0.1279 & 0.3202 & 0.346 & 0.2821 & 0.6372 \\ 0.2278 & 0.4051 & 0.1107 & 0.924 & 0.7876 & 0.2157 & 0.0529 & 0.2402 & 0.1467 \\ 0.4636 & 0.2086 & 0.5459 & 0.6851 & 0.4435 & 0.6745 & 0.4678 & 0.6737 & 0.898 \end{bmatrix}.$$

The normalized matrix is evaluated using the normalization process defined for cost and beneficial criteria in step 1 of Sect. 3 and is given as R_{norm}.

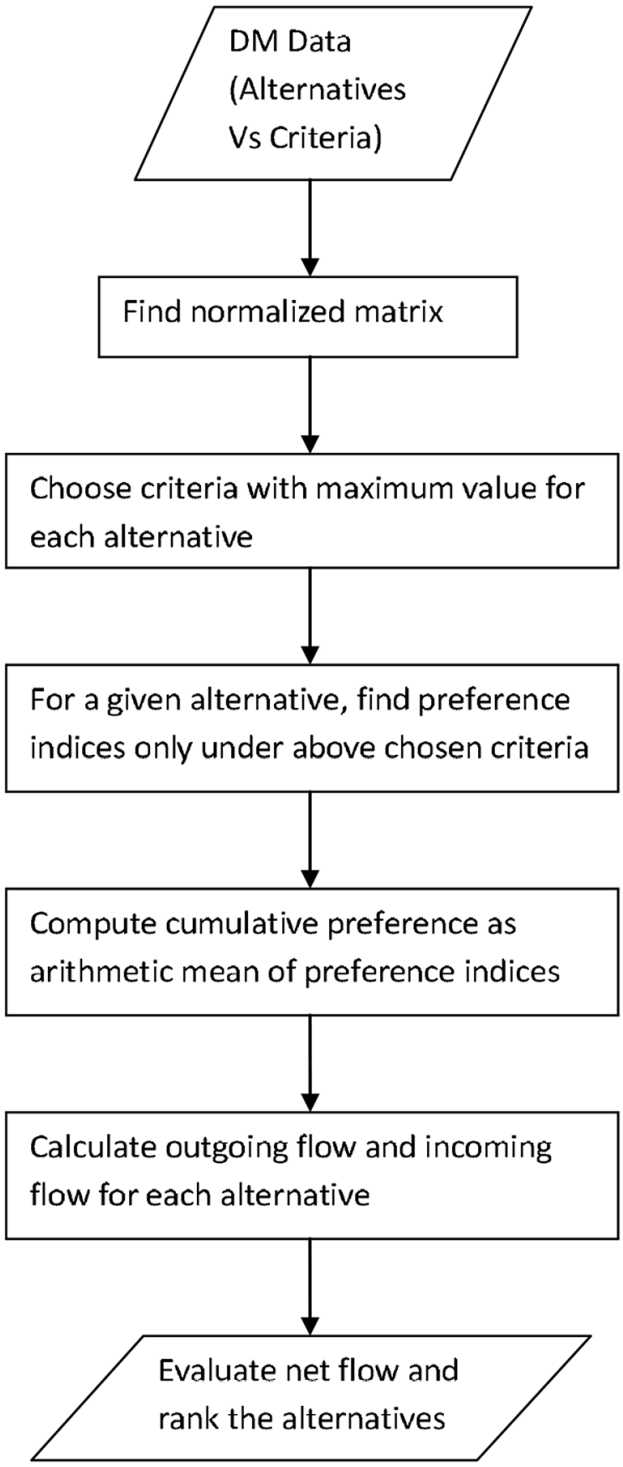


Fig. 1 Flow chart for the proposed methodology

$$R_{norm} = \begin{bmatrix} 0.1785 & 1.0000 & 0.6118 & 0.8006 & 0.9494 & 0.4322 & 0.6295 & 0.4296 & 0.6371 \\ 0.8997 & 0.6524 & 0.4101 & 1.0000 & 1.0000 & 0.3590 & 0.4637 & 0.0000 & 0.3325 \\ 0.7773 & 0.0000 & 0.0000 & 0.4605 & 0.4759 & 0.0000 & 0.0391 & 1.0000 & 1.0000 \\ 1.0000 & 0.9560 & 1.0000 & 0.7466 & 0.1026 & 0.2632 & 0.0000 & 0.8906 & 0.1117 \\ 0.6528 & 0.5466 & 0.9336 & 0.0000 & 0.9590 & 0.4352 & 0.6102 & 0.0675 & 0.6000 \\ 0.0000 & 0.3375 & 0.1111 & 0.0217 & 0.0000 & 0.2686 & 1.0000 & 0.0017 & 0.0000 \\ 0.6207 & 0.1076 & 0.5919 & 0.3067 & 0.5002 & 1.0000 & 0.4483 & 0.6824 & 0.9190 \end{bmatrix}$$

Table 2 Maximum value and the index set of criteria at which the maximum value is achieved

Alternative	Maximum value (M_i)	Index set of criteria J_i
A_1	1	{2}
A_2	1	{4, 5}
A_3	1	{8, 9}
A_4	1	{1, 3}
A_5	0.9590	{5}
A_6	1	{7}
A_7	1	{6}

The maximum value M_i and the index set J_i of each alternative are given in Table 2.

The preference indices are evaluated under the criteria that attain the maximum value for given alternative as mentioned in Table 3.

Note the number of preference indices to be evaluated is 60, whereas the conventional PROMETHEE II method would have taken 378 evaluations.

The cumulative preference indices are given in Table 4.

For each alternative, the outgoing flow, the incoming flow, and the net flow are depicted in Table 5.

By net flow values, A_1 is the best alternative, followed by A_7 and A_2.

To validate the proposed method, the ranking of alternatives are also obtained by conventional PROMETHEE II and the TOPSIS method. The weights of the criteria in the former approach are taken to be uniform for the nine criteria. The two variants of TOPSIS method described in the literature, please see[40], differ only in the normalization process. Here, we use TOPSIS with linear normalization (same as in Step 1 of the designed methodology) to make a conducive comparison with the ranking of the suggested method. In this work, the solution procedures for PROMETHEE II and TOPSIS methods are not provided, but one can see[9, 38] for the reference. The rankings of seven alternatives by the three methods are tabulated in Table 6.

The ranking achieved by the designed method is same as ranking from the two conventional methods, making the proposed method promising.

5 Empirical analysis and discussion

Although in certain situations, the proposed methodology, PROMETHEE II method, and TOPSIS (with linear normalisation) can result in similar rankings (as in above example), but eventually they can lead to different rankings. Therefore, compatibility and stability of rankings obtained by various decision-making methods must be analysed for the validity and efficiency of the new method.

Table 3 Preference indices $P_j(A_i, A_{i'}),\ j \in J_i,\ i, i' \in I,\ i \neq i'$

	C1	C2	C3	C4	C5	C6	C7	C8	C9
$P(A_1, A_2)$		0.3476							
$P(A_1, A_3)$		1							
$P(A_1, A_4)$		0.0440							
$P(A_1, A_5)$		0.4534							
$P(A_1, A_6)$		0.6625							
$P(A_1, A_7)$		0.8924							
$P(A_2, A_1)$				0.199356	0.050589				
$P(A_2, A_3)$				0.5395	0.5241				
$P(A_2, A_4)$				0.2534	0.8974				
$P(A_2, A_5)$				1	0.041				
$P(A_2, A_6)$				0.9783	1				
$P(A_2, A_7)$				0.6933	0.4998				
$P(A_3, A_1)$								0.5704	0.3629
$P(A_3, A_2)$								1	0.6675
$P(A_3, A_4)$								0.1094	0.8883
$P(A_3, A_5)$								0.9325	0.4
$P(A_3, A_6)$								0.9983	1
$P(A_3, A_7)$								0.3176	0.0810
$P(A_4, A_1)$	0.8215		0.3882						
$P(A_4, A_2)$	0.1003		0.5899						
$P(A_4, A_3)$	0.2227		1						
$P(A_4, A_5)$	0.3472		0.0664						
$P(A_4, A_6)$	1		0.8889						
$P(A_4, A_7)$	0.3793		0.4081						
$P(A_5, A_1)$					0.0096				
$P(A_5, A_2)$					0				
$P(A_5, A_3)$					0.4831				
$P(A_5, A_4)$					0.8564				
$P(A_5, A_6)$					0.9590				
$P(A_5, A_7)$					0.4588				
$P(A_6, A_1)$							0.3705		
$P(A_6, A_2)$							0.5363		
$P(A_6, A_3)$							0.9609		
$P(A_6, A_4)$							1		
$P(A_6, A_5)$							0.3898		
$P(A_6, A_7)$							0.5517		
$P(A_7, A_1)$						0.5678			
$P(A_7, A_2)$						0.6410			
$P(A_7, A_3)$						1			
$P(A_7, A_4)$						0.7368			
$P(A_7, A_5)$						0.5648			
$P(A_7, A_6)$						0.7314			

Table 4 Cumulative preference indices

Cumulative preference indices			
$\pi(A_1, A_2) = 0.3476$	$\pi(A_2, A_1) = 0.1250$	$\pi(A_3, A_1) = 0.4667$	$\pi(A_4, A_1) = 0.6049$
$\pi(A_1, A_3) = 1.0000$	$\pi(A_2, A_3) = 0.5318$	$\pi(A_3, A_2) = 0.8338$	$\pi(A_4, A_2) = 0.3451$
$\pi(A_1, A_4) = 0.0440$	$\pi(A_2, A_4) = 0.5754$	$\pi(A_3, A_4) = 0.4989$	$\pi(A_4, A_3) = 0.6113$
$\pi(A_1, A_5) = 0.4534$	$\pi(A_2, A_5) = 0.5205$	$\pi(A_3, A_5) = 0.6662$	$\pi(A_4, A_5) = 0.2068$
$\pi(A_1, A_6) = 0.6625$	$\pi(A_2, A_6) = 0.9891$	$\pi(A_3, A_6) = 0.9991$	$\pi(A_4, A_6) = 0.9444$
$\pi(A_1, A_7) = 0.8924$	$\pi(A_2, A_7) = 0.5965$	$\pi(A_3, A_7) = 0.1993$	$\pi(A_4, A_7) = 0.3937$
$\pi(A_5, A_1) = 0.0096$	$\pi(A_6, A_1) = 0.3705$	$\pi(A_7, A_1) = 0.5678$	
$\pi(A_5, A_2) = 0.0000$	$\pi(A_6, A_2) = 0.5363$	$\pi(A_7, A_2) = 0.6410$	
$\pi(A_5, A_3) = 0.4831$	$\pi(A_6, A_3) = 0.9609$	$\pi(A_7, A_3) = 1.0000$	
$\pi(A_5, A_4) = 0.8564$	$\pi(A_6, A_4) = 1.0000$	$\pi(A_7, A_4) = 0.7368$	
$\pi(A_5, A_6) = 0.9590$	$\pi(A_6, A_5) = 0.3898$	$\pi(A_7, A_5) = 0.5648$	
$\pi(A_5, A_7) = 0.4588$	$\pi(A_6, A_7) = 0.5517$	$\pi(A_7, A_6) = 0.7314$	

Table 5 Different flows for each alternative

Alternative	Outgoing flow	Incoming flow	Net flow
A_1	0.5667	0.3574	0.2093
A_2	0.5564	0.4506	0.1057
A_3	0.6107	0.7645	− 0.1539
A_4	0.5177	0.6186	− 0.1009
A_5	0.4611	0.4669	− 0.0058
A_6	0.6349	0.8809	− 0.2461
A_7	0.707	0.5154	0.1916

Table 6 Ranking of alternatives by three methods

	Method	Rank of alternatives
1.	Proposed method	$A_1 > A_7 > A_2 > A_5 > A_4 > A_3 > A_6$
2.	PROMETHEE II	$A_1 > A_7 > A_2 > A_4 > A_5 > A_3 > A_6$
3.	TOPSIS with linear normalization	$A_1 > A_7 > A_2 > A_4 > A_5 > A_3 > A_6$

Wu et al.[41] discussed the quantitative method for measuring the stability of rankings in group decision-making. The stability degree of an MCDM method is a quantitative measure, with values in [0, 1], and the highest degree gives the most stable ranking.

Conformity measure evaluates the performance of an MCDM method vis a vis other MCDM methods, and it is to be minimised to improve the performance of an MCDM method, see[42].

The rank correlation coefficient is used to analyse the similarity between the rankings given by each method. Commonly used rank correlation coefficients are

Spearman's rank-order correlation coefficient, Kendall's tau-b and Pearson's correlation coefficient. The Kendall's tau-b is a non-parametric rank correlation coefficient which computes the degree of similarity between two sets of ranks for the same set of objects. The Kendall's tau-b coefficient also takes into account ties in rankings, a feature desirable in the context of our present study.

For m (number of alternatives), n (number of criteria), and q (number of randomly generated MCDM matrices), the procedure of computing the aforementioned measures is discussed below.

Step 1: For $t = 1$ to q,
generate a random alternative-criteria evaluation matrix R_t of size $m \times n$.
Step 2: Pass each R_t, $t = 1, \ldots, q$, into three methods {Proposed Method, PROMETHEE II, TOPSIS} to obtain rankings for the m alternatives. The total number of such rankings is q for each of the three methods.
Step 3: Computing the stability degree:

3(a)(i) Determine the score matrices $(SM)_{Algo} = [N_{is}]_{m \times m}$, where N_{is} is the number of times an alternative A_i appears at the sth position in q ranks obtained in Step 2, and $Algo \in \{$ Proposed Method, PROMETHEE II, TOPSIS $\}$. For example, in matrix $(SM)_{ProposedMethod}$, the N_{21} depicts the number of times the alternative A_2 appears at 1st position while ranking the alternatives using the proposed methodology.

3(a)(ii) For each method, evaluate the weighted score vector as $S_{Algo} = (SM)_{Algo} W$, where the weight vector is taken as $W = (m-1, m-2, \ldots, 1, 0)^T$ from the voting rule of Borda count method, as discussed in[41], and $Algo \in \{$ Proposed Method, PROMETHEE II, TOPSIS $\}$. The alternatives are ranked with the first one having the highest score in S_{Algo} vector[1].
The combined (overall) ranking of alternatives and the score vector S_{Algo}, where $Algo \in \{$ Proposed Method,PROMETHEE II, TOPSIS $\}$, are used to calculate the stability degree of the rankings.

3(b)(i) Let $I^{(1)}_{Algo}$ and $I^{(2)}_{Algo}$ be the sets of indices in I corresponding to the alternatives having rank respectively, 1 and 2 in the combined ranking obtained by the $Algo$. Note that the sets $I^{(\ell)}_{Algo}$, $\ell = 1, 2$, can have more than one single index. Calculate the vector $U_{Algo} = (u^1_{Algo}, \ldots, u^m_{Algo})$, as follows:

[1] The vector S_{Algo} is formed by combining q alternative-criteria matrices to rank the alternatives; this scheme can easily be extended to the Group Decision MCDM problems.

$$u^k_{Algo} = \begin{cases} S^k_{Algo} & \text{if } \; k \notin I^{(1)}_{Algo} \cup I^{(2)}_{Algo} \\ \frac{S^k_{Algo}+S^b_{Algo}}{2} & k \in I^{(1)}_{Algo}, \\ & \text{and for any index } \; b \in I^{(2)}_{Algo} \\ \frac{S^k_{Algo}+S^b_{Algo}}{2} & k \in I^{(2)}_{Algo}, \\ & \text{and for any index } \; b \in I^{(1)}_{Algo}. \end{cases}$$

Here, S^k_{Algo}, S^b_{Algo} are the k-th and b-th components of the vector S_{Algo} respectively.

3(b)(ii) Find the stability degree ρ_{Algo} of $Algo \in$ { Proposed Method, PROMET-HEE II, TOPSIS }, by computing sine of the angle between the vectors S_{Algo} and U_{Algo} as follows:

$$\rho_{Algo} = \frac{||S_{Algo} \times U_{Algo}||}{||S_{Algo}||\,||U_{Algo}||}, \quad \times \text{ stands for the cross product .}$$

Step 4: Computing the conformity values:

4(i) For $t = 1$ to q, compute the conformity values

$$C^t_{Algo} = \sum_{i=1}^{m} \left(\text{ Net Flow}^t_{Algo,(i)} - \text{ Mean Flow}^t_{(i)} \right)^2,$$

$Algo \in$ { Proposed Method, PROMETHEE II, TOPSIS }. For TOPSIS, the net flow value of each alternative is replaced by its relative closeness value. And the mean flow for each alternative is evaluated by taking the arithmetic mean of the net flow values (or relative closeness values for TOPSIS) of m alternatives, for each of the three methods.

4(ii) Determine the mean and the variance of these conformity values.

$$\overline{C}_{Algo} = \frac{1}{q} \sum_{t=1}^{q} C^t_{Algo}$$

$$\sigma^2_{C_{Algo}} = \frac{1}{q} \sum_{t=1}^{q} (C^t_{Algo} - \overline{C}_{Algo})^2.$$

4(iii) Apply the hypothesis testing to test the significance of the difference between the three vectors $C_{Algo} = \{C^1_{Algo}, C^2_{Algo}, \ldots, C^q_{Algo}\}$, $Algo \in$ { Proposed Method, PROMETHEE II, TOPSIS }.

Step 5: Computing the rank correlation coefficient:

5(i) Go to step 3 and obtain the combined rank for each of the three methods using 3(a)(i) and 3(a)(ii).

5(ii) Evaluate Kendall's tau-b correlation coefficient to analyse the similarity between the rankings given by each method.

Table 7 Stability degrees from three methods

S. No.	No. of alternatives	No. of criteria	Proposed method	PROMETHEE	TOPSIS
1	3	5	0.00097	0.00026	0.00032
2	5	6	0.00106	9.3287×10^{-5}	0.00019
3	7	9	0.00033	5.3452×10^{-6}	1.4254×10^{-5}
4	10	10	0.00056	0.00015	4.9690×10^{-5}
5	10	15	0.00023	3.2299×10^{-5}	0.00013
6	12	15	0.00014	0.00011	3.9711×10^{-5}
7	15	20	0.00014	2.9994×10^{-5}	3.8341×10^{-5}
8	20	25	8.7711×10^{-5}	6.1415×10^{-5}	5.4092×10^{-5}
9	25	40	0.00016	4.0659×10^{-5}	6.6232×10^{-5}
10	30	50	0.00011	0.0001	0.0001

In the empirical analysis, we took different pairs (m, n) for alternatives-criteria and for each such pair we randomly simulated 1,00,000 independent and identically distributed matrices of size $m \times n$. The global stability degrees of the combined rankings of these 1,00,000 matrices with different numbers of alternatives and criteria are reported in Table 7.

It is evident from Table 7 that, for each pair of alternatives and criteria, the rankings delivered by the designed method are more stable than those given by the other two methods. This approves the advantage of the proposed method in solving MCDM problems.

Next, we compute the conformity vectors C_{TOPSIS}, $C_{\text{PROMETHEE}}$ and $C_{\text{Proposed Method}}$ for 1,00,000 matrices of size $m \times n$, as generated in stability analysis. To check if there is a significant difference between these three vectors, we set the two null hypotheses $H_0^{(1)}$ and $H_0^{(2)}$ and their corresponding alternate hypotheses $H_1^{(1)}$ and $H_1^{(2)}$ as follows:

$$
\begin{aligned}
H_0^{(1)} &: \overline{C}_{\text{TOPSIS}} - \overline{C}_{\text{Proposed Method}} = 0,\\
H_1^{(1)} &: \overline{C}_{\text{TOPSIS}} > \overline{C}_{\text{Proposed Method}},\\
H_0^{(2)} &: \overline{C}_{\text{Proposed Method}} - \overline{C}_{\text{PROMETHEE}} = 0, \quad \text{and}\\
H_1^{(2)} &: \overline{C}_{\text{Proposed Method}} > \overline{C}_{\text{PROMETHEE}}.
\end{aligned}
$$

The one-tailed Z-test is conducted at 99% level of significance. The results are reported in Table 8.

From Table 8, the proposed method performs better than the TOPSIS method. And the Z-statistics confirms the rejection of the null hypothesis $H_0^{(1)}$ at 99% significance level which concludes that the mean conformity of the proposed algorithm is less than that of the TOPSIS method to a noteworthy extent. It is also worth noticing that although the mean conformity of the PROMETHEE II

Table 8 First two moments of the conformity vectors and their Z-statistics

S. No.	No. of alternatives	No. of criteria	TOPSIS		Proposed method		PROMETHEE II		Z-statistics	
			Mean	S.d.	Mean	S.d.	Mean	S.d.	TOPSIS—proposed method	Proposed method—PROMETHEE
1	3	5	0.3661	0.0488	0.1303	0.0390	0.1119	0.0268	3.7742	0.3899
2	5	6	0.5882	0.0677	0.1972	0.0364	0.1543	0.0191	5.0877	1.0422
3	7	9	0.8038	0.0734	0.2547	0.0356	0.2091	0.0193	6.7296	1.1262
4	10	10	1.1420	0.0954	0.3561	0.0387	0.2951	0.0239	7.6376	1.3424
5	10	15	1.1319	0.0770	0.3354	0.0326	0.2919	0.0198	9.5262	1.1424
6	12	15	1.3578	0.0900	0.4032	0.0357	0.3496	0.0227	9.8578	1.2674
7	15	20	1.6901	0.0948	0.4878	0.0352	0.4335	0.0239	11.8830	1.2738
8	20	25	2.2478	0.1084	0.6382	0.0375	0.5751	0.0272	14.0383	1.3634
9	25	40	2.7988	0.1035	0.7667	0.0337	0.7120	0.0260	18.6740	1.2827
10	30	50	3.3546	0.1081	0.9082	0.0341	0.8517	0.0272	21.5876	1.2969

is lower than the designed method yet the null hypothesis $H_0^{(2)}$ is accepted given the Z-statistics at 99% confidence. It signifies that there is essentially no significant difference between $\overline{C}_{\text{PROMETHEE II}}$ and $\overline{C}_{\text{Proposed Method}}$. Hence, the efficiency of the proposed method and the PROMETHEE II method are analogous.

To analyse the performance of the suggested scheme, for a small sample size data, the above procedure is followed for sample size 10 and 20. The conformity vectors C_{TOPSIS}, $C_{\text{PROMETHEE}}$ and $C_{\text{Proposed Method}}$ are evaluated and hypothesis testing is carried out to check if there is a significant difference between these three vectors. The two null hypotheses $H_0^{(1)}$ and $H_0^{(2)}$ are same as before.

The one-tailed t-test is conducted at 99% level of significance. The results are reported in Table 9 for sample size 10 and in Table 10 for sample size 20. The p-value is mentioned in the bracket.

From Tables 9 and 10, we can conclude that the proposed algorithm performs better than the TOPSIS method with the same reason as for the large sample size. And also, the efficiency of the suggested method and PROMETHEE II are comparable. The analysis indicates the efficiency and adeptness of the designed method.

For the next analysis, various pairs (m, n) for alternatives-criteria are taken and for each such pair, 10 independent and identically distributed matrices of size $m \times n$ are randomly simulated. The combined rankings for each of the three methods are obtained, and Kendall's tau-b correlation coefficients are evaluated using IBM SPSS Statistics 21 software as given in Table 11.

It is evident from Table 11, that the ranking obtained by the proposed method is highly correlated with the rankings acquired by PROMETHEE II and TOPSIS with statistically significant p-values.

In summary, the proposed algorithm is an efficient and conducive approach. The method obtains the most stable ranking of the alternatives that is highly correlated with the rankings acquired by the existing methods. Also, the conformity values assert the better performance of the suggested scheme in comparison to TOPSIS while the efficacy of the suggested approach and PROMETHEE II are analogous.

Note: The experts and researchers working in this field may arise a concern that the weights of the criteria reflect the preferences of the DM and hence, the omission of pre-defined criteria weight vector is not justifiable. In reference to this issue, the authors agree with the fact that the individual's preference of the criteria is of significant importance, but it is not always possible for DM to quantify his preference for all available criteria. It may also be claimed that various well-known techniques to evaluate the exact criteria weights exist in the literature. The adoption of such techniques is the DM's choice, but these methodologies add to the complexity of the decision situation and emanate the decision-making as a time-consuming process. In such a scenario, the suggested method calculates the criteria weights itself to overcome the subjectivity of human thinking in the decision-making. The statistical results given in Sect. 5 confirm the legitimacy of the proposed method and hence, validate the efficacy and competency of the scheme.

Table 9 First two moments of the conformity vectors and their t-statistics for sample size 10

S. No.	No. of alternatives	No. of criteria	TOPSIS		Proposed method		PROMETHEE II		t-statistics	
			Mean	S.d.	Mean	S.d.	Mean	S.d.	TOPSIS—proposed method	Proposed method—PROMETHEE
1	3	5	0.3663	0.0546	0.1368	0.0307	0.1053	0.0183	3.6637	0.8846 (0.0126)
2	5	6	0.6024	0.0706	0.2039	0.0323	0.1573	0.0184	5.1342	1.2525 (0.0034)
3	7	9	0.7803	0.0782	0.2463	0.0380	0.2037	0.0169	6.1421	1.0239 (0.0020)
4	10	10	1.1383	0.0870	0.3301	0.0268	0.2925	0.0231	8.8836	1.0635 (0)
5	10	15	1.1328	0.0617	0.3242	0.0255	0.2896	0.0156	12.1055	1.1549 (0)
6	12	15	1.3608	0.0688	0.3768	0.0225	0.3477	0.0174	13.5967	1.0230 (0)
7	15	20	1.7167	0.0563	0.4746	0.0112	0.4384	0.0149	21.6561	1.9405 (0)
8	20	25	2.2268	0.1048	0.6038	0.0400	0.5644	0.0275	14.4756	0.8113 (0)
9	25	40	2.8097	0.0780	0.7372	0.0248	0.7081	0.0195	25.3284	0.9210 (0)
10	30	50	3.3425	0.0445	0.8688	0.0124	0.8415	0.0115	53.5804	1.6126 (0)

Table 10 First two moments of the conformity vectors and their t-statistics for sample size 20

S. No.	No. of alternatives	No. of criteria	TOPSIS		Proposed method		PROMETHEE II		t-statistics	
			Mean	S.d.	Mean	S.d.	Mean	S.d.	TOPSIS—proposed method	Proposed method—PROMETHEE
1	3	5	0.3629	0.0540	0.1271	0.0381	0.1129	0.0337	3.5698	0.2797 (0.0251)
2	5	6	0.5625	0.0562	0.2066	0.0335	0.1505	0.0169	5.4409	1.4944 (0)
3	7	9	0.7800	0.0947	0.2485	0.0344	0.2047	0.0243	5.2762	1.0387 (0)
4	10	10	1.1293	0.0996	0.3563	0.0374	0.2921	0.0249	7.2648	1.4282 (0)
5	10	15	1.1339	0.0742	0.3239	0.0201	0.2911	0.0178	10.5360	1.2197 (0)
6	12	15	1.3707	0.1020	0.4073	0.0394	0.3540	0.0253	8.8107	1.1364 (0)
7	15	20	1.6514	0.0710	0.4780	0.0358	0.4240	0.0210	14.7582	1.3006 (0)
8	20	25	2.2724	0.0855	0.6363	0.0280	0.5797	0.0206	18.1891	1.6268 (0)
9	25	40	2.8155	0.1350	0.7722	0.0414	0.7167	0.0344	14.4742	1.0295 (0)
10	30	50	3.3469	0.1274	0.9075	0.0384	0.8498	0.0310	18.3385	1.1691 (0)

Table 11 Kendall's tau-b correlation coefficients of proposed method ranking with PROMETHEE and TOPSIS rankings, the 2-tailed significance values are given in brackets

S. No.	No. of alternatives	No. of criteria	PROMETHEE	TOPSIS
1	3	5	1.000** (0)	1.000** (0)
2	5	6	0.894* (0.037)	0.894* (0.037)
3	7	9	0.683* (0.033)	0.586 (0.068)
4	10	10	0.395 (0.121)	0.524* (0.043)
5	15	20	0.546** (0.005)	0.476* (0.015)

*indicates significance of correlation at 1% and **indicates significance of correlation at 5%

6 Conclusions

PROMETHEE II is a highly researched and most applied outranking method. In this paper, inspired by the PROMETHEE II method, we proposed a scheme which overcomes two concerns with the PROMETHEE II method. One is the prior knowledge of the criteria weight vector, and second is the number of computations going in the performance indices, especially when the number of alternatives is large. The computational time complexity of the classical PROMETHEE II has also reduced from $O(m^2n)$ to $O(mn)$ that makes the algorithm efficient. The stability analysis confirms the suitability and usage of the proposed method by experts/decision-makers. The conformity analysis proves that the designed method is equally as good as the classical PROMETHEE II method. The correlation coefficient approves the ranking given by the proposed method makes the method more desirable.

We envision that the proposed method can be extended to the scenarios when information is expressed through linguistic variables, interval-valued linguistic variables, or fuzzy numbers.

Acknowledgements The authors wish to express their gratitude to the esteemed referees for their valuable observations and comments in improving the quality of this research work.

Authors' Contributions Anjali Singh performed the computations and drafted the manuscript. Anjana Gupta and Aparna Mehra conceived of the presented theory and verified the analytical methods. All authors provided critical feedback and helped shape the research, analysis and manuscript. All authors read and approved the final manuscript.

Funding Not applicable.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Availability of data and material Not applicable.

References

1. Lootsma, F.A.: Multi Criteria Desicion Analysis via Ratio and Difference Judgement. Kluwer Academic Publishers, Dordrecht (1999)
2. Beccali, M., Cellura, M., Mistretta, M.: Decision-making in energy planning: application of the ELECTRE method at regional level for the diffusion of renewable energy technology. Renew. Energy **28**, 2063–2087 (2003)
3. Hokkanen, J., Salminen, P.: ELECTRE III and IV decision aids in an environmental problem. J. MultiCriteria Decis. Anal. **6**, 215–226 (1997)
4. Keyser, W.D., Peeters, P.: Theory and methodology: a note on the use of PROMETHEE multicriteria methods. Eur. J. Oper. Res. **89**, 457–461 (1996)
5. Shiha, H.S., Shyurb, H.J., Lee, E.S.: An extension of TOPSIS for group decision-making. Math. Comput. Modell. **45**, 801–813 (2007)
6. Wiecek, M.M., Ehrgott, M., Fadel, G., Figueira, J.R.: Multiple criteria decision-making for engineering. Omega **36**, 337–339 (2008)
7. Brans, J.P.: Lingenierie de la decision. Elaboration dinstruments daide a la decision. In: Methode PROMETHEE, Nadeau, R., Landry, M., (eds.) Laide a la Decision: Nature, Instruments et Perspectives Davenir. Presses de Universite Laval, Quebec, Canada, pp. 183–214 (1982)
8. Vincke, Ph, Brans, J.P.: A preference ranking organization method: the PROMETHEE method for multiple criteria decision-making. Manag. Sci. **31**, 647–656 (1985)
9. Pohekar, S.D., Ramachandran, M.: Multi-criteria evaluation of cooking energy alternatives for promoting parabolic solar cooker in India. Renew. Energy **29**, 1449–1460 (2004)
10. Albadvi, A., Chaharsooghi, S.K., Esfahanipour, A.: Decision-making in stock trading: an application of PROMETHEE. Eur. J. Oper. Res. **177**, 673–683 (2007)
11. Queiruga, D., Walther, G., Gonza'lez-Benito, J., Spengler, T.: Evaluation of sites for the location of WEEE recycling plants in Spain. Waste Manag. **28**, 181–190 (2008)
12. Singh, A., Gupta, A., Mehra, A.: An AHP-PROMETHEE II method for 2-tuple linguistic multicriteria group decision-making. In: 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), IEEE Xplore, pp. 1–6 (2015). https://doi.org/10.1109/ICRITO.2015.7359374
13. Singh, A., Gupta, A., Mehra, A.: Energy planning problems with interval valued 2-tuple linguistic information. Oper. Res. Int. J. (2016). https://doi.org/10.1007/s12351-016-0245-x
14. Mareschal, B.: Weight stability intervals in multicriteria decision aid. Eur. J. Oper. Res. **33**, 54–64 (1988)
15. Brans, J.P., Mareschal, B.: The PROMETHEE VI procedure: how to differentiate hard from soft multicriteria problems. J. Decis. Syst. **4**, 213–223 (1995)
16. Calders, T., Assche, D.V.: PROMETHEE is not quadratic: an $O(qn \log(n))$ algorithm. Omega. (2017). https://doi.org/10.1016/j.omega.2017.04.003
17. Macharis, C., Brans, J.P., Mareschal, B.: The GDSS PROMETHEE procedure: a PROMETHEE-GAIA based procedure for group decision support. J. Decis. Syst. **7**, 283–307 (1998)
18. Mareschal, B., Brans, J.P.: Geometrical representations for MCDA: the GAIA module. Eur. J. Oper. Res. **34**, 69–77 (1988)
19. Brans, J.P., Mareschal, B.: The PROMETHEE GAIA decision support system for multicriteria investigations. Investig. Oper. **4**, 107–117 (1994)
20. Figueira, J., de Smet, Y., Brans, J.P.: MCDA methods for sorting and clustering problems: Promethee TRI and Promethee CLUSTER, Universite Libre de Bruxelles, Service deMathematiques de la Gestion, Working Paper 2004/02 (2004)
21. Halouani, N., Chabchoub, H., Martel, J.M.: PROMETHEE-MD-2T method for project selection. Eur. J. Oper. Res. **195**, 841–895 (2009)
22. Hendriks, M.M.W.B., De Boer, J.H., Smilde, A.K., Doornbos, D.A.: Multicriteria decision-making. Chemometr. Intell. Lab. Syst. **16**, 175–191 (1992)
23. Herngren, L., Goonetilleke, A., Ayoko, G.A.: Analysis of heavy metals in roaddeposited sediments. Anal. Chim. Acta **571**, 270–278 (2006)
24. Pohekar, S.D., Ramachandran, M.: Application of multi-criteria decision-making to sustainable energy planning—a review. Renew. Sustain. Energy Rev. **8**, 365–381 (2004)
25. Zhou, P., Ang, B.W., Poh, K.L.: Decision analysis in energy and environmental modeling: an update. Energy **31**, 2604–2622 (2006)

26. Johnson, M.P.: Spatial decision support for assisted housing mobility counseling. Decis. Support Syst. **41**, 296–312 (2005)
27. Brans, J.P., Macharis, C., Kunsch, P.L., Chevalier, A., Schwaninger, M.: Combining multicriteria decision aid and system dynamics for the control of socio-economic processes, an iterative real-time procedure. Eur. J. Oper. Res. **109**, 428–441 (1998)
28. Behzadian, M., Kazemzadeh, R.B., Albadvi, A., Aghdasi, M.: PROMETHEE: a comprehensive literature review on methodologies and applications. Eur. J. Oper. Res. **200**(1), 198–215 (2010)
29. Hwang, C.L., Yoon, K.P.: Multiple Attribute Decision-Making: Methods and Applications. Springer, New York (1981)
30. Chen, C.T., Lin, C.T., Huang, S.F.: A fuzzy approach for supplier evaluation and selection in supply chain management. Int. J. Prod. Econ. **102**, 289–301 (2006)
31. Zhang, G., Shang, J., Li, W.: An information -granulation entropy-based model for third-party logistics providers evaluation. Int. J. Prod. Res. **50**, 177–190 (2012)
32. Grassi, A., Gamberini, R., Mora, C., Rimini, B.: A fuzzy multi-attribute model for risk evaluation in workplaces. Saf. Sci. **47**, 707–716 (2009)
33. Aiello, G., Enea, M., Galante, G., La Scalia, G.: Clean agent selection approached by fuzzy TOPSIS decision-making method. Fire Technol. **45**, 405–418 (2009)
34. Boran, F.E., Genc, S., Kurt, M., Akay, D.: Personnel selection based on intuitionistic fuzzy sets. Hum. Factors Ergon. Manuf. Serv. Ind. **21**(5), 493–503 (2011)
35. Saremi, M., Mousavi, S.F., Sanayei, A.: TQM consultant selection in SMEs with TOPSIS under fuzzy environment. Expert Syst. Appl. **36**, 2742–2749 (2009)
36. Doukas, H., Karakosta, C., Psarras, J.: Computing with words to assess the sustainability of renewable energy options. Expert Syst. Appl. **37**, 5491–5497 (2010)
37. Chamodrakas, I., Martakos, D.: A utility-based fuzzy TOPSIS method for energy efficient network selection in heterogeneous wireless networks. Appl. Soft Comput. **11**, 3734–3743 (2011)
38. Wei, G.W.: Extension of TOPSIS method for 2-tuple linguistic multiple attribute group decision-making with incomplete weight information. Knowl. Inf. Syst. **25**, 623–634 (2010)
39. Behzadian, M., Otaghsara, S.K., Yazdani, M., Ignatius, J.: A state-of the-art survey of TOPSIS applications. Expert Syst. Appl. **39**(17), 13051–13069 (2012)
40. Opricovic, S., Tzeng, G.H.: Compromise solution by MCDM methods: a comparative analysis of VIKOR and TOPSIS. Eur. J. Oper. Res. **156**, 445–455 (2004)
41. Wu, F., Zhao, Y., Chen, Y.: Stability analysis of group decision-making under weighted scoring rules. Int. J. Syst. Sci. (2015). https://doi.org/10.1080/00207721.2015.1128577
42. Razaei, J.: Best-worst multi-criteria decision-making method. Omega **53**, 49–57 (2015)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal of Public Health | pp. 1–9 | doi:10.1093/pubmed/fdaa099

Breast cancer burden in Africa: evidence from GLOBOCAN 2018

Rajesh Sharma

University School of Management and Entrepreneurship, Delhi Technological University, Delhi 110095, India
Address correspondence to Rajesh Sharma, E-mail: rajesh.sharma@dtu.ac.in.

ABSTRACT

Background Breast cancer is the leading malignancy in African females. This study aims to examine the breast cancer burden in Africa using recently released GLOBOCAN 2018 estimates.

Methods The incidence and mortality estimates of age- and country-wise burden of breast cancer in 54 African countries were obtained from GLOBOCAN 2018.

Results In Africa, breast cancer caused 74 072 deaths, and 168 690 cases were estimated to have occurred in 2018. The age-standardized incidence rate stood at 37.9/100 000 in Africa, varying from 6.9/100 000 in the Gambia to 69.6/100 000 in Mauritius. The age-standardized mortality rate stood at 17.2/100 000 in 2018, ranging from 4/100 000 in the Gambia to 29.1/100 000 in Somalia in 2018. Nigeria was the leading country in terms of absolute burden with 26 310 cases and 11 564 deaths, followed by Egypt with 23 081 new cases and 9254 deaths. The mortality-to-incidence ratio for Africa stood at 0.44, varying from 0.24 in Libya to 0.68 in the Central African Republic.

Conclusion To tackle breast cancer burden in Africa, the main challenges are late-stage disease presentation, lack of screening and therapeutic infrastructure, lack of awareness and limited resources.

Keywords Africa, breast cancer, GLOBOCAN, incidence, mortality

Introduction

Breast cancer is the leading cancer group in African females and its burden has been rising steadily in the last three decades owing to factors such as longevity, westernization of lifestyle characterized by delayed marriages, a shorter period of breast-feeding and lower age at menarche.[1–5] The mortality rates are higher and survival rates are lower in Africa than developed countries due to poor oncologic care, ranging from prevention to palliation.[6] The different disease outcomes and survivorship in Africa were ascribed to genetic variation[7] and tumor stage at diagnosis.[8] Much of breast cancer (80–90%) arises sporadically due to a multitude of risk factors, whereas a small percentage is associated with inherited gene mutations (e.g. BRCA 1 or BRCA 2).[9, 10]

This study examines the burden of breast cancer in Africa using recently released GLOBOCAN 2018 estimates.[11] We examined the estimates of age- and country-wise, and age-standardized rates of incidence and mortality of breast cancer in 54 African countries in 2018 in the light of country-level human development index (HDI).[12] The HDI is a summary indicator of three indicators of development: health (life expectancy), education (mean years of schooling and expected years of schooling) and income (gross national income (GNI) per capita). It is a better indicator of human development than its components as it not only reflects healthcare resources (through income per capita) in a country but also encompasses general conditions of population health (through life expectancy) and general awareness levels (through years of schooling). Country-level survival rates of breast cancer were approximated using mortality-to-incidence ratio (MIR), which is a useful proxy of survival rates of different neoplasia.[2, 13, 14] The examination of leading malignancy in African females is expected to help realign public health priorities toward emerging public health concerns such as breast cancer and shape the public health policy toward cancer control and management to achieve

Rajesh Sharma, Assistant Professor

© The Author(s) 2020. Published by Oxford University Press on behalf of Faculty of Public Health. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Downloaded from https://academic.oup.com/jpubhealth/article-abstract/doi/10.1093/pubmed/fdaa099/5869901 by Delhi Technological University user on 20 July 2020

better disease outcomes and improve the quality of life of breast cancer patients.

Methods

Breast cancer estimates for 54 African countries were obtained from GLOBOCAN 2018.[11, 15, 16] The estimates were produced only for invasive carcinoma classified as C50 by the International Classification of Diseases, tenth version (ICD-10). GLOBOCAN 2018 estimates were produced for 18 nonoverlapping age-groups: under 5, 5 to 9, 10 to 14, 15 to 19, 20 to 24, 25 to 29, 30 to 34, 35 to 39, 40 to 44, 45 to 49, 50 to 54, 55 to 59, 60 to 64, 65 to 69, 70 to 74, 75 to 79, 80 to 84 and 85 plus. All-age and age-specific incidence and death counts were reported along with 95% uncertainty intervals (UIs), which reflects the uncertainties associated with cancer registry coverage and death registration, data quality and timeliness of data reporting.[11, 15]

All the GLOBOCAN 2018 estimates are available at global cancer observatory (gco.iarc.fr), and a brief account of Globocan methodology is presented in Appendix 1. The 5-year survival rate of breast cancer was approximated using the country-level MIR, which was calculated by dividing all-age death count with all-age incidence count.

Human development index

We also examined the association of breast cancer burden and country-level HDI.[12, 17] HDI is an indicator of country-level development, which is a composite index of three indices: income index, education index and health index. Each of these indices had values varying between 0 (lowest) and 1 (highest), and the final HDI was calculated by taking the geometric mean of three indicators.[12]

Odd os breast cancer incidence

We also examined the cumulative risk of developing and dying from breast cancer in African countries during their lifetime (below the age of 75 years) using the following formula.

$$\text{Cumulative Risk} = 1 - \exp^{-\text{CR}}$$

In the above equation, CR denotes cumulative rate and was calculated by $\text{CR} = \sum_i \text{Rate}_i * T_i$ and Rate_i is the rate of incidence or mortality in ith age-group and T_i is the number of years in the ith age-group. The odds of developing or dying from breast cancer were calculated as per the following.

$$\text{ODDS} = \frac{1}{\text{Cumulative Risk}}$$

Results

Geographical distribution of breast cancer

In Africa, breast cancer caused an estimated 74 072 deaths [95% UI, 67 345–81 472] with 168 690 [152 332–186 804] cases were estimated to have occurred in 2018. The age-standardized incidence rate (ASIR) stood at 37.9/100 000 in Africa, whereas the age-standardized mortality rate (ASMR) stood at 17.2/100 000 in 2018. Nigeria was the leading country in terms of absolute burden with 26 310 [23 610–29319] cases and 11 564 [10 302–12 981] deaths followed by Egypt with 23 081 [21 734–24 512] new cases and (9254 [8707–9836]) deaths in 2018 (Fig. 1a and b; Supplementary Table S1). Geographically, high all-age incidence counts were also reflected in high all-age death numbers, the age-standardized rates, however, were quite asymmetric between incidence and mortality in Africa (Fig. 1a–d).

The ASIR was highest in Mauritius (69.6/100 000) and lowest in the Gambia (6.9/100 000) in 2018 (Fig. 1c; Supplementary Table S1). Besides Mauritius in eastern Africa, there were pockets of high incidence rates in north Africa (e.g. Algeria: 55.6/100 000; Egypt: 52.4/100 000), central Africa (e.g. Republic of Congo: 37.8/100 000) and southern Africa (e.g. South Africa: 49/100 0000; Zimbabwe: 38.4/100 000) (Fig. 1c). The ASMR was the highest in Somalia (29.1/100 000), with the Gambia recording the lowest ASMR of 4/100 000 in 2018. The high ASMRs were mostly concentrated near central and northeast Africa toward the great horn of Africa (e.g. Central African Republic: 23.9/100 000; Somalia: 29.1/100 000; Eritrea: 24.8/100 000; Sierra Leone 25.4/100 000). Low ASMRs were largely concentrated toward the southeastern region of Africa (e.g. Botswana: 7.1/100 000; Madagascar: 7.5/100 000).

The MIR for Africa stood at 0.44 in 2018 with widespread variation across the region ranging from 0.24 in Libya to 0.68 in Central African Republic (Fig. 1e; Supplementary Table S1). High MIRs were mostly concentrated in central sub-Saharan Africa (SSA) (e.g. Central African Republic: 0.68; Equatorial Guinea: 0.64), eastern SSA (e.g. Somalia: 0.675; Eritrea: 0.58) and western SSA (e.g. Niger: 0.64; Chad:0.59). The MIRs were the lowest in north Africa (e.g. Algeria: 0.28; Libya: 0.24).

Age-wise, breast cancer incidence and deaths increased up to the 45–49 age-group and declined after that (Fig. 2). This pattern was preserved in nearly all African countries, with the incidence being the highest in middle age-groups (30–49) (Supplementary Figure S1a and b).

Association between breast cancer burden and HDI

Figure 3a–c displays the relationship between various metrics of breast cancer burden and country-level HDI. Both

Downloaded from https://academic.oup.com/jpubhealth/article-abstract/doi/10.1093/pubmed/fdaa099/5869901 by Delhi Technological University user on 20 July 2020

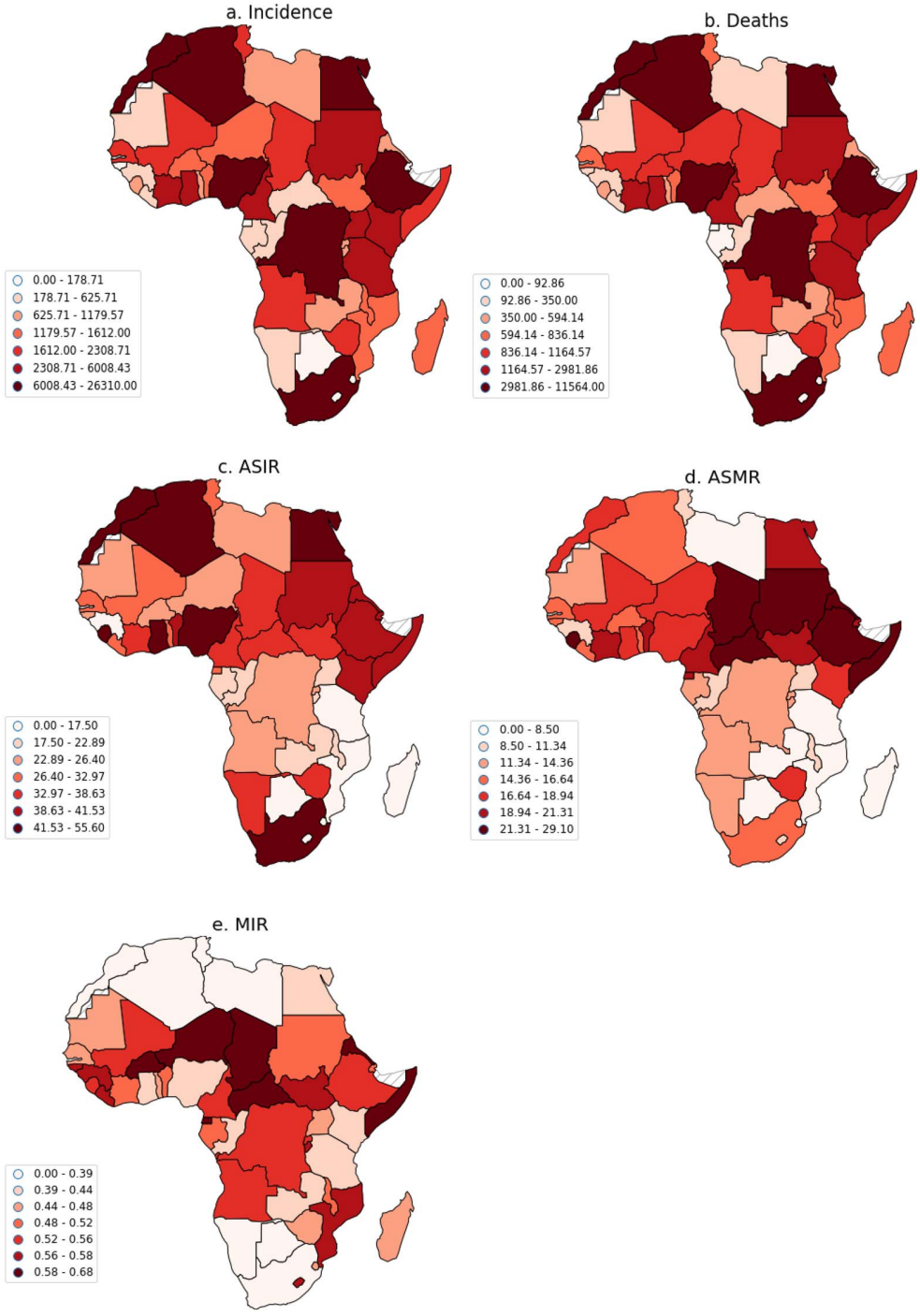


Fig. 1 Geographical distribution of breast cancer burden in Africa (a) Incidence (b) Deaths (c) ASIR (d) ASMR (e) MIR. Incidence: All-age new cases; Deaths: All-age death counts; ASIR: Age-standardized incidence rate; ASMR: Age-standardized mortality rate; MIR: Mortality-to-incidence ratio. Data Source: GLOBOCAN 2018 (IARC). Figure couldn't show the heatmap for the countries Cape Verde, Comoros, Mauritius and Sao Tome and Principe whose country-specific values can be obtained from Supplementary Table S1.

all-age incidence count and ASIRs were positively linked with HDI, albeit with few exceptions, with comparatively developed countries having high counts as well as incidence rates (Fig. 3a). The relationship between all-age incidence and HDI, although statistically insignificant (at 5% level of significance), was positive; the relationship between ASIR and HDI, however, was positive and statistically significant (Supplementary Table S2). Mortality, however,

Downloaded from https://academic.oup.com/jpubhealth/article-abstract/doi/10.1093/pubmed/fdaa099/5869901 by Delhi Technological University user on 20 July 2020

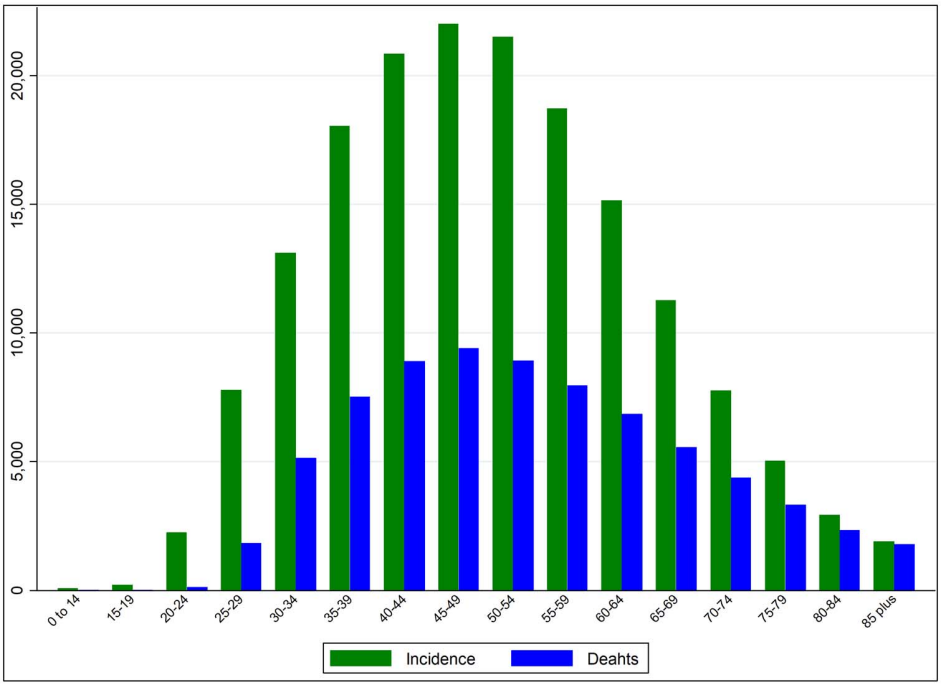


Fig. 2 Age-group wise distribution of breast cancer incidence and death counts in 2018. Incidence: Age-wise new cases; Deaths: Age-wise death counts. Data Source: GLOBOCAN 2018 (IARC).

showed a slightly different relationship with HDI; all-age death counts were positively linked (although statistically insignificant) with HDI, the ASMRs, however, were negatively associated (although statistically insignificant) with HDI (Fig. 3b; Supplementary Table S2). MIR, a proxy for 5-year survival rate had strong negative and statistically significant association (at 1% level of significance) with HDI (R-squared: 0.6384), with high HDI countries of Africa (e.g. Algeria, HDI:0.75 and MIR:0.28; Mauritius, HDI: 0.79 and MIR: 0.33) registering comparatively higher survival rates (i.e. low MIR; Fig. 3c and Supplementary Table S2).

Odds of breast cancer

In Africa, the odds of developing breast cancer were 1 in 25 females (probability: 3.94%) during their lifetime (before the age 75 years), whereas 1 in 57 females (probability: 1.77%) were expected to die of breast cancer during their lifetime (Fig. 4; Supplementary Table S3). The risk of developing breast cancer varied 10-fold in Africa; it was the highest in Mauritius (7.27%, odds: 1 in 14) and the lowest in the Gambia (0.71%, odds: 1 in 141). The probability of dying from breast cancer varied 7-fold in Africa, from 0.42% (odds: 1 in 240) in the Gambia to Somalia (2.97%, odds: 1 in 34). We noted that the odds of developing cancer were higher in comparatively developed countries of Africa, whereas the risk of breast cancer death displayed a negative association with HDI (Supplementary Figure S2a and b).

Discussion

Main finding of this study

Across Africa, the breast cancer burden rose from 92 600 cases and 50 000 deaths in 2008 (as per GLOBOCAN 2008[18]) to 168 690 cases and 74 072 deaths in 2018 (as per GLOBOCAN 2018). The age-standardized rates also displayed increases, with ASIR rising from 28/100 000 in 2008[18] to 37.9/100 000 in 2018, whereas the ASMR rose marginally from 16/100 000 in 2008[18] to 17.2/100 000 in 2018. While the incidence rates increased in developed countries, too, the rise in mortality rates is specific to Africa in line with previous research.[2] As per GLOBOCAN 2018, the ASIR was 84.8/100 000 in North America and 86.7/100 000 in Oceania, whereas the same stood at 37.2/100 000 in Africa in 2018. ASMR, in contrast, was lower in North America (12.6/100 000) and Oceania (14.9/100 000) in comparison with Africa (17.2/100 000). As per GLOBOCAN 2018, in terms of ASIR, no African country featured in top 10, whereas three African countries featured in top 10 in terms of ASMR (In fact, no African country featured in top 25 in terms of ASIR, but 12/25 countries belonged to Africa in terms of ASMR.). High incidence rates, along with low mortality rates in developed regions of the world, reflect diagnosis of breast cancer at an early stage and the fact that better treatment is available in these countries.[19]

Downloaded from https://academic.oup.com/jpubhealth/article-abstract/doi/10.1093/pubmed/fdaa099/5869901 by Delhi Technological University user on 20 July 2020

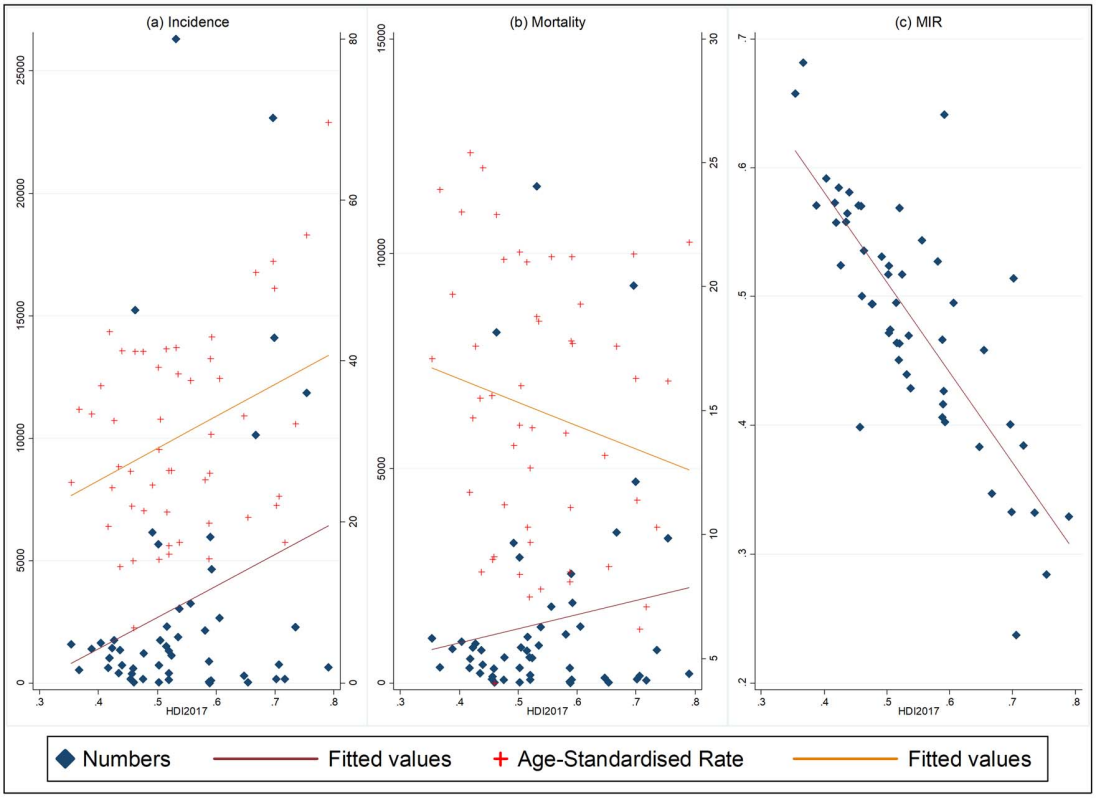


Fig. 3 Bivariate relationship between country-level breast cancer burden and HDI (a) Incidence versus HDI (b) Deaths versus HDI (c) MIR versus HDI. Incidence: All-age new cases in 2018; Deaths: All-age death counts; ASIR: Age-standardized incidence rate; ASMR: Age-standardized mortality rate; MIR: Mortality-to-incidence ratio; HDI: Human Development Index. The data of HDI for 2018 was procured from united nations development program (UNDP). Data Source: GLOBOCAN 2018 (IARC).

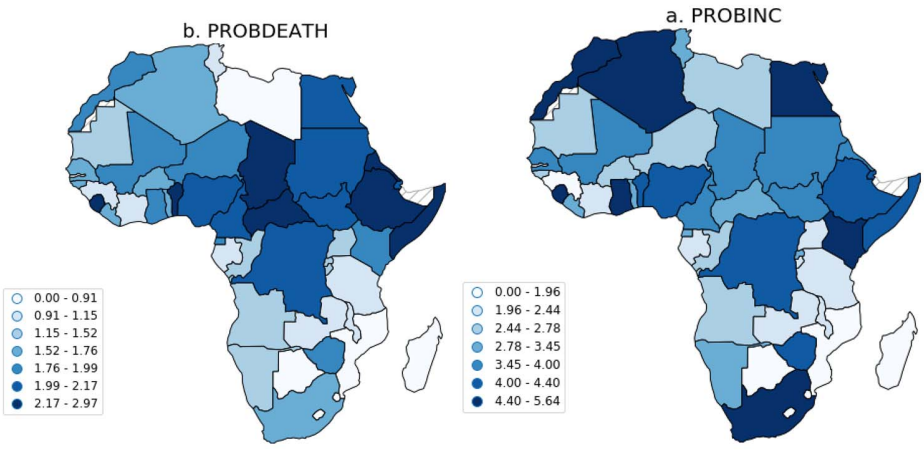


Fig. 4 Geographical distribution of risk of developing and dying from breast cancer in Africa (a) PROBINC (b) PROBDEATH PROBINC: Probability of developing breast cancer during lifetime (before the age 75); PROBDEATH: Probability of dying from breast cancer during lifetime (before the age 75). Figure couldn't show the heatmap for the countries Cape Verde, Comoros, Mauritius and Sao Tome and Principe whose country-specific values can be obtained from Supplementary Table S1.

Downloaded from https://academic.oup.com/jpubhealth/article-abstract/doi/10.1093/pubmed/fdaa099/5869901 by Delhi Technological University user on 20 July 2020

What is already known on this topic

International Agency for Research on Cancer (IARC) previously published GLOBOCAN estimates for the year 2008 and 2012.[20, 21] In the latest iteration, the cancer estimates were provided for the most recent year 2018. While few studies examined the global burden of breast cancer in multicountry context,[2, 5] no study has examined the most recent estimates of breast cancer burden in Africa.

What this study adds

This study examined the burden of breast cancer in 54 African countries measured in terms of age-wise and age-standardized rates of incidence and mortality. The MIR ratio was employed to approximate the 5-year survival rates and the country-wise metrics of breast cancer burden were comprehensively examined in light of the development status of a country measured by its HDI. With few exceptions, greater socioeconomic status measured by HDI was linked with high ASIRs and low ASMRs (Fig. 3a and b). High ASIRs in developing countries of Africa relates to the similar risk factors as in developed countries as well as better screening of breast cancer in a few of these countries. MIR, a proxy for survival rate, had a strong negative and statistically significant association with HDI. The poor survival rates in SSA countries have been documented before[6, 22] and are due to low awareness levels, lack of screening availability and cancer care infrastructure, among other factors.[23]

We found that 50% of patients were below the age 50 (74% below the age of 60) with higher HDI countries had more breast cancer patients in older age-groups as high HDI characterizes greater development and longer life spans (Supplementary Fig S3). Previous studies have highlighted the issue of younger age-profiles of breast cancer patients in Africa.[24] Although lower life expectancy in African countries compared to developed countries is considered to be the biggest reason, younger age of breast cancer patients among African−American women[25] also suggests there must be few factors other than life expectancy such as genetic factors or environmental factors that predispose females of African origin to breast cancer at earlier age.[24] These patterns are particularly concerning given that breast cancer among young women tends to be clinically and pathologically more aggressive with higher mortality rate than that in older women.[26]

Late-disease presentation is a hallmark of breast cancer burden in Africa.[27] A meta-analysis on breast cancer presentation in Africa by Jedy-Agba *et al*.[27] found late-stage presentation varying from 30% in South Africa to 98% in Nigeria with disease presentation also differing among black and nonblack women. As per Abdulrahman and Rahman,[28] 50–75% of African women present with advanced disease, whereas only 11% of breast cancer patients presented with advanced disease (stage III or IV) in the USA.[29] In a country-specific study, Islami *et al*.[30] reported 74% of breast cancer patients in Cote d'Ivoire and 81% in the Democratic Republic of Congo presented with advanced disease.

The advanced disease presentation in Africa results in poor disease outcomes and hence, lower survival rates.[6] The survival rates were found to be 12 and 46% in Gambia and Uganda, respectively.[22] In Africa, lack of awareness, fear of mastectomy and societal fears were a few of the patient-specific reasons responsible for the late-disease presentation and higher mortality rates.[2–4, 23] Moreover, many African women also show up to a traditional healer on first noticing symptoms, who promises treatment without any painful surgery or medicine, which further delays proper diagnosis and poor prognosis ensues.[31–33] In addition to the above patient-specific factors, late-disease presentation also reflects poor oncologic infrastructure, week referral systems and lack of screening services in Africa[23].

While many developed countries can reduce case fatality rates for breast cancer by promoting early detection using mammography screening, many African countries lack either access to screening equipment and services or the means to promote them due to misconception and lack of population awareness. As far as screening is concerned, the developed countries' approach of screening through mammograms cannot be emulated in African countries due to a lack of resources and awareness.[34] As per breast health global initiative, there must be phased approaches toward breast cancer detection as per country's resource availability; in limited-resource settings, the focus must be on raising awareness and capacity building and in highest-resource settings, the focus must be on screening of asymptomatic women.[35]

As mammographic screening is not cost effective in low-resource settings in Africa, clinical breast examination (CBE) may prevent the delay in diagnosis and can facilitate earlier disease presentation. Preliminary results from randomized control trials (RCTs) on the efficacy of CBE in low- and medium-income countries have revealed success in downstaging and early detection of breast cancer.[36–38] Although CBE has not been recommended as primary screening device due to nonconclusive evidence of its mortality-reduction effects, its cost-effectiveness and relative success in downstaging of cancer in RCTs makes it suitable to be recommended for those women who do not receive regular mammograms either because it is not recommended (women aged < 40 and > 69) or because breast mammograms are costly. It can also present an opportunity for healthcare providers to educate women on breast cancer, risk factors, role of early detection in progression and lastly, self-breast examination.[39, 40]

Downloaded from https://academic.oup.com/jpubhealth/article-abstract/doi/10.1093/pubmed/fdaa099/5869901 by Delhi Technological University user on 20 July 2020

Breast cancer risk factors pertain to the duration of breast-feeding, age at menarche and menopause, hormone status, life expectancy, dietary risk factors, exposure to ionizing radiation and genetic predisposition.[41, 42] Few of these risk factors were not prevalent in Africa earlier; in recent times, however, life expectancy is increasing, infant mortality is falling along with declining fertility rates, marital behaviors, and reproductive patterns are changing in Africa.[43, 44] Cumulatively, these behavioral changes have been termed westernization of lifestyle, which has resulted in higher breast cancer risk in African women.[3]

Cancer care infrastructure ranging from preventive medicine to palliative care is often scarce in low- and medium-income countries, including Africa.[24, 45] For instance, even though surgery is the primary treatment option of breast cancer and sometimes the only treatment option, a small percent of breast cancer patients in low- and medium-income countries such as in Africa have access to safe, affordable and timely surgery.[46] Only 25 of the 43 responding African countries had CBE available at public primary healthcare level, whereas 15 of 43 African countries had mammograms available at the public primary health level (Supplementary Table S4) (Only 16 of 43 African countries had national level cancer control plan or strategy by 2014.).

Limitations of the Study

The quality and reliability of estimates depend crucially upon the degree of representativeness and the quality of source information[11, 15]. As per Bray *et al*.[47], across SSA, there were only four population-based registries (Kampala, Uganda; Harare, Zimbabwe; Blantyre, Malawi and PROMEC, South Africa). Moreover, as population-based cancer registries are not available in many of the African countries and estimates were obtained using data from neighboring countries, which may be underreported in few of these countries. Therefore, the breast cancer estimates reported and examined in this paper are expected to be biased. Moreover, the modeling effort may result in too wide UIs, which may render policy deductions lesser useful in few countries.

Conclusion

With ~ 170 000 cases and ~ 75 000 deaths, breast cancer has become one of the most significant public health threats in Africa with incidence rates fast approaching that of developed countries and mortality rates higher than that of developed countries. Breast cancer is a cause of concern in Africa owing to increasing incidence, high mortality, low survival rates, late-stage disease presentation and lack of cancer infrastructure. Increasing life expectancy and westernizing lifestyles will result in greater breast cancer burden in Africa in the near future, and the public health systems in Africa have to gear up to tackle this enormous public health threat. In addition to boosting oncologic infrastructure, there must be effective educational campaigns to educate women about the importance of early detection of malignancy and access to care[48]; for this purpose, low-cost interventions such as CBE can be employed to educate women on all issues relating to breast cancer. Lastly, there must be an increased focus on establishing cancer registries and expanding the coverage of already existing ones in Africa.

Supplementary data

Supplementary data mentioned in the text are available to subscribers in *PUBMED* online.

Ethical approval and Consent to Participate

The research was conducted using data available in the public domain and does not include any human participants and/or animals. Therefore, no ethical approvals were required.

Consent to Publish

The authors of the study have the consent and responsibility of submission to this journal.

Conflict of Interest

The authors declare that they have no conflict of interest.

Funding

The authors received no specific funding for this research.

References

1 Fitzmaurice C, Akinyemiju TF, Al Lami FH *et al.* Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2016: a systematic analysis for the global burden of disease study. *JAMA oncol* 2018;**4**:1553–68.

2 Sharma R. Breast cancer incidence, mortality and mortality-to-incidence ratio (MIR) are associated with human development, 1990–2016: evidence from global burden of disease study 2016. *Breast Cancer* 2019a;**26**:428–45.

Downloaded from https://academic.oup.com/jpubhealth/article-abstract/doi/10.1093/pubmed/fdaa099/5869901 by Delhi Technological University user on 20 July 2020

3 Porter P. "Westernizing" women's risks? Breast cancer in lower income countries. *N Engl J Med* 2008;**358**:213–6.

4 Arnold M, Pandeya N, Byrnes G *et al.* Global burden of cancer attributable to high body-mass index in 2012: a population-based study. *Lancet Oncol* 2015;**16**:36–46.

5 Li N, Deng Y, Zhou L *et al.* Global burden of breast cancer and attributable risk factors in 195 countries and territories, from 1990 to 2017: results from the global burden of disease study 2017. *J Hematol Oncol* 2019;**12**:140.

6 Allemani C, Matsuda T, Di Carlo V *et al.* Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet* 2018;**391**:1023–75.

7 Fregene A, Newman LA. Breast cancer in sub-Saharan Africa: how does it relate to breast cancer in African-American women? *Cancer* 2005;**103**:1540–50.

8 Awadelkarim KD, Arizzi C, Elamin EOM *et al.* Pathological, clinical and prognostic characteristics of breast cancer in Central Sudan versus northern Italy: implications for breast cancer in Africa. *Histopathology* 2008;**52**:445–56.

9 Bradbury AR, Olopade OI. Genetic susceptibility to breast cancer. *Rev Endocr Metab Disord* 2007;**8**:255–67.

10 Kim H, Choi DH. Distribution of BRCA1 and BRCA2 mutations in Asian patients with breast cancer. *J Breast Cancer* 2013;**16**:357–65.

11 Bray F, Ferlay J, Soerjomataram I *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;**68**:394–424.

12 UNDP. *Human Development Indices and Indicators 2018 Statistical Update*. http://hdr.undp.org/sites/default/files/2018_human_development_statistical_update.pdf (3 May 2019, date last accessed)

13 Vostakolaei F, Karim-Kos HE, Janssen-Heijnen ML *et al.* The validity of the mortality to incidence ratio as a proxy for site-specific cancer survival. *Eur J Public Health* 2010;**21**:573–7.

14 Sharma R. The burden of prostate cancer is associated with human development index: evidence from 87 countries, 1990–2016. *EPMA J* 2019b;**10**:137–52.

15 Ferlay J, Colombet M, Soerjomataram I *et al.* Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer* 2019;**144**:1941–53.

16 Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F. *Global Cancer Observatory*. https://gco.iarc.fr/ (1 May 2019, date last accessed)

17 Human Development Database. http://hdr.undp.org/en/data# (3 May 2019, date last accessed)

18 Jemal A, Bray F, Forman D *et al.* Cancer burden in Africa and opportunities for prevention. *Cancer* 2012;**118**:4372–84.

19 Youlden DR, Cramb SM, Dunn NA *et al.* The descriptive epidemiology of female breast cancer: an international comparison of screening, incidence, survival and mortality. *Cancer Epidemiol* 2012;**36**:237–48.

20 Ferlay J, Shin HR, Bray F *et al.* Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010;**127**: 2893–917.

21 Ferlay J, Soerjomataram I, Dikshit R *et al.* Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* 2015;**136**:E359–86.

22 Sankaranarayanan R, Swaminathan R, Brenner H *et al.* Cancer survival in Africa, Asia, and central America: a population-based study. *Lancet Oncol* 2010;**11**:165–73.

23 Brinton LA, Figueroa JD, Awuah B *et al.* Breast cancer in sub-Saharan Africa: opportunities for prevention. *Breast Cancer Res Treat* 2014;**144**:467–78.

24 Forouzanfar MH, Foreman KJ, Delossantos AM *et al.* Breast and cervical cancer in 187 countries between 1980 and 2010: a systematic analysis. *Lancet* 2011;**378**:1461–84.

25 Brinton LA, Sherman ME, Carreon JD, Anderson WF. Recent trends in breast cancer among younger women in the United States. *J Natl Cancer Inst* 2008;**100**:1643–8.

26 Anders CK, Johnson R, Litton J *et al.* Breast cancer before age 40 years. *Semin Oncol* 2009;**36**:237–49.

27 Jedy-Agba E, McCormack V, Adebamowo C, dos-Santos-Silva I. Stage at diagnosis of breast cancer in sub-Saharan Africa: a systematic review and meta-analysis. *Lancet Glob Health* 2016;**4**:e923–35.

28 Abdulrahman GO, Rahman GA. Epidemiology of breast cancer in Europe and Africa. *J Cancer Epidemiol* 2012;**2012**:915610.

29 Sariego J. Patterns of breast cancer presentation in the United States: does geography matter? *Am Surg* 2009;**75**:545–50.

30 Islami F, Lortet-Tieulent J, Okello C *et al.* Tumor size and stage of breast cancer in cote d'Ivoire and republic of Congo—results from population-based cancer registries. *Breast* 2015;**24**:713–7.

31 Clegg-Lamptey JN, Dakubo JC, Attobra YN. Psychosocial aspects of breast cancer treatment in Accra. *Ghana. East Afr Med J.* 2009;**86**(7):348–53.

32 Ezeome ER. Delays in presentation and treatment of breast cancer in Enugu. *Nigeria. Niger J Clin Pract* 2010;**13**(3):311–6.

33 De Ver DT, Bogale S, Hobden C *et al.* A mixed-method assessment of beliefs and practice around breast cancer in Ethiopia: implications for public health programming and cancer control. *Glob Public Health* 2011;**6**:719–31.

34 Harford JB. Breast-cancer early detection in low-income and middle-income countries: do what you can versus one size fits all. *Lancet Oncol* 2011;**12**:306–12.

35 Yip CH, Smith RA, Anderson BO *et al.* Breast health global initiative early detection panel. Guideline implementation for breast healthcare in low-and middle-income countries: early detection resource allocation. *Cancer* 2008;**113**:2244–56.

36 Pisani P, Parkin DM, Ngelangel C *et al.* Outcome of screening by clinical examination of the breast in a trial in the Philippines. *Int J Cancer* 2006;**118**:149–54.

37 Sankaranarayanan R, Ramadas K, Thara S *et al.* Clinical breast examination: preliminary results from a cluster randomized controlled trial in India. *J Natl Cancer Inst* 2011;**103**:1476–80.

38 Mittra I, Mishra GA, Singh S *et al.* A cluster randomized, controlled trial of breast and cervix cancer screening in Mumbai, India: methodology and interim results after three rounds of screening. *Int J Cancer* 2010;**126**:976–84.

Downloaded from https://academic.oup.com/jpubhealth/article-abstract/doi/10.1093/pubmed/fdaa099/5869901 by Delhi Technological University user on 20 July 2020

39 Smith RA, Saslow D, Sawyer KA *et al.* American Cancer Society guidelines for breast cancer screening: update. *CA Cancer J Clin* 2003;**53**:141–69.

40 Saslow D, Hannan J, Osuch J *et al.* Clinical breast examination: practical recommendations for optimizing performance and reporting. *CA: a cancer journal for clinicians* 2004;**54**:327–44.

41 Beral V, Bull D, Pirie K *et al.* Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. *Lancet Oncol* 2012;**13**:1141–51.

42 Colditz GA, Bohlke K. Priorities for the primary prevention of breast cancer. *CA Cancer J Clin* 2014;**64**:186–94.

43 Dicker D, Nguyen G, Abate D *et al.* Global, regional, and national age-sex-specific mortality and life expectancy, 1950–2017: a systematic analysis for the global burden of disease study 2017. *Lancet* 2018;**392**:1684–735.

44 Murray CJ, Callender CS, Kulikoff XR *et al.* Population and fertility by age and sex for 195 countries and territories, 1950–2017: a systematic analysis for the global burden of disease study 2017. *Lancet* 2018;**392**:1995–2051.

45 Kingham TP, Alatise OI, Vanderpuye V *et al.* Treatment of cancer in sub-Saharan Africa. *Lancet Oncol* 2013;**14**:e158–67.

46 Sullivan R, Alatise OI, Anderson BO *et al.* Global cancer surgery: delivering safe, affordable, and timely cancer surgery. *Lancet Oncol* 2015;**16**:1193–224.

47 Bray F, Ferlay J, Laversanne M *et al.* Cancer incidence in five continents: inclusion criteria, highlights from volume X and the global status of cancer registration. *Int J Cancer* 2015;**137**: 2060–71.

48 Zelle SG, Nyarko KM, Bosu WK *et al.* Costs, effects and cost-effectiveness of breast cancer control in Ghana. *Trop Med Int Health* 2012;**17**:1031–43.

Downloaded from https://academic.oup.com/jpubhealth/article-abstract/doi/10.1093/pubmed/fdaa099/5869901 by Delhi Technological University user on 20 July 2020

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-020-09199-5

Capsmf: a novel product recommender system using deep learning based text analysis model

Rahul Katarya[1] · Yamini Arora[2]

Received: 29 November 2019 / Revised: 19 April 2020 / Accepted: 8 June 2020
Published online: 29 June 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Researchers and data scientists have developed different Recommender System Algorithms such as Content-Based and Collaborative-Based in order to filter a large amount of information available on the internet and hence, recommend only the relevant and essential content based on the personalized interests of users. Information acquired explicitly by collecting users' ratings for an item lead to the problem of data sparsity. Many researchers have been currently working towards the improvement of rating prediction accuracy by integrating the auxiliary information along with the ratings provided by the users. This paper proposes a novel product recommender system called as "CapsMF", it applies the advanced neural network architecture Capsule Networks (Caps) for document representation, and MF represents Matrix factorization. In the proposed approach, we have enhanced a deep neural network text analysis model by adding a newly discovered neural network architecture; Capsule Networks stacked on bi-directional Recurrent Neural Network (Bi-RNN) for the robust representation of textual descriptions of items and users. The Deep Neural Network text analysis model is integrated with the Probabilistic Matrix Factorization to generate improved recommendations. The experiment has been performed on two real amazon datasets resulting in the enhancement of rating prediction accuracy, the recall, and the precision of top-n recommendations, in comparison to the basic and hybrid Recommendation System Algorithms. Also, text analysis model involving Capsule Networks stacked with Recurrent Neural Networks (RNNs) have outperformed the baseline models that have single Convolutional Neural Networks (CNN) or CNN combined with Bi-RNN in text analysis.

Keywords Recommender system · Matrix factorization · Capsule networks · Collaborative filtering

✉ Rahul Katarya
rahulkatarya@dtu.ac.in

Yamini Arora
arorayamini08@gmail.com

Extended author information available on the last page of the article

1 Introduction

In this digital era, people use e-commerce websites for shopping and social media sites for news and related information. The problem arises when there are varieties of options available while searching for a thing. It is tough for the end users to find the critical content they are looking for according to their interest. The solution for this new digital era issue is Recommender Systems (RS). These are being deployed on e-commerce websites and on the social media platforms that extract the personal information related to the users and recommend them the relevant content and information that they want depending on their interest, activity, and behavior. For example, movies, books, clothes, tweets are being recommended to the people visiting different web sites. Recommender System [22] is an application of Web Mining. This process discovers interests and behavior from large amounts of data without human interference and hence, ranks the most relevant content or items using the rating prediction of the users visiting a website. The researchers and data scientists that filter the information according to the personalized interests of the users have developed different types of Recommendation System Algorithms and techniques. A recommender system can extract information related to a user in two different ways: explicitly or implicitly.

One can get the information related to the users explicitly by collecting the ratings for an item given by a user, and it can be acquired implicitly by monitoring the behavior of users on the different websites related to the product [2]. Their behavior includes songs heard, applications downloaded, websites visited, movies watched, books read. Social information can be compiled by scraping the users' accounts on different social websites, for example, what kind of posts they like, their followers, followees and different activities on social media can tell a lot about the preference of the users, and hence, personalized recommendations can be more accurate and precise [15, 18].

Based on the information about different items and users, two categories of recommendation algorithms developed are Content-Based, Collaborative-based. Out of these, Matrix Factorization technique based on Collaborative Filtering (CF) gives the best recommendations [21]. The limitation observed in the current system is that information acquired explicitly by collecting users' ratings for an item lead to the problem of data sparsity because not every user rate every other movie. The accuracy of rating prediction for an item reduces, thereby not satisfying the needs of the customer. Many researchers inclined towards the improvement of rating prediction accuracy by integrating the auxiliary information along with the ratings provided by the users. It has been observed in related works that integrating the textual data along with rating data has brought an improvement in the rating prediction accuracy. However, document modeling approaches are different in different research papers.

Various topic modeling approaches and deep neural network models [36] are being used to represent the textual description of items and users for improving the rating prediction accuracy and recall and precision of top-n recommendations. The recent models integrate matrix factorization with textual data generate a textual representation of documents using deep neural network architectures; they are more effective as they extract semantic information from the text.

Capsule Networks bring an improvement in the old neural network architecture, and it works better than the Convolutional Neural Networks (CNN). There are certain disadvantages of using the convolutional neural networks, few areas where CNN lacks, Capsule Networks have overcome those limitations, and now it is a better neural network architecture for

developing models to solve the problems in different domains. Capsule Networks are primarily used for Classification of Images in different categories [8].

CN can also be applied in the areas of NLP and RS to utilize textual information more efficiently. Neural Networks can be trained to learn numeric representations of various words and phrases using models such as Word2Vec and Glove and can be applied to classify the data in different categories.

This research paper has successfully executed the following points:

We have addressed the limitations of CNN based document modeling approach as well as CNN and RNN based text representation approach.

We have proposed a novel and powerful Deep Neural Network text analysis model combining Capsule Networks and Bi-directional RNN to accomplish the textual representation of items' and users' reviews to improve recommendation performance.

Experiments results prove that exploiting ratings, items' and users' reviews have improved rating prediction accuracy, precision, and recall of top-n recommendations.

The rating Prediction accuracy on proposed model "CapsMF" has shown improvement over PMF, DRMF-Item, DRMF-User, and DRMF by 26%, 4.7%, 3.5%, and 1.7% in terms of MAE.

Different hybrid Recommendation Models using items' and users' reviews as auxiliary information have been described in the Related Work. The theory related to Capsule Networks and Bi-RNN and Recommendation algorithm, Probabilistic Matrix Factorization is explained in the Methodologies section. The proposed deep neural network text analysis model and the integration of PMF with a novel text analysis model are explained in the Proposed Approach section of the Paper. The last section of the paper deals with Experiment and Results along with Conclusion and Future Work associated with this approach.

2 Related work

Here, we have discussed different models that combine textual data with Recommendation System algorithms providing recommendations to the users. In [11], a brief about text classification using Capsule Networks(CN) has been explained. The different hybrid recommendation systems that have been proposed by the researchers that combine textual data and recommendation algorithms are explained further.

CTR, Collaborative Topic Regression [30], incorporates the advantages of collaborative-based, MF algorithm and probabilistic topic modeling approach for recommending research based articles to the scientists online. Content analysis is done using the Topic modeling approach (LDA) combined with latent factor models using MF for recommending the unseen articles [28]. Their approach worked well as it recommended completely unrated articles to the users that were useful and hence, predictions were made in the right manner. Various efficient algorithms have been proposed in further years that have worked upon these models and have improved them. In HFT [19] model, the authors have developed a statistical model that combines hidden dimensions in the rating matrix with the topics extracted from the text of the review given by users. This methodology helped in predicting users' rating on unseen movies and further recommends them. The models developed by them helped in the discovery of genre and identified the most informative reviews. Their model "HFT", Hidden Factor Topics has addressed the most occurred problem in recommendation tasks, Cold-Start. Their model performed better than the latent factor-based recommender system improving accuracy

by 5–10% [32]. The model RMR [17], proposed by the authors, have combined the content-based and collaborative based filtering techniques to develop a novel integrated recommender system. Their model "RMR", Rating meets reviews combined the rating model with a topic model to generate recommendations and thereby solving the "cold-start" problem, that is recommending the new products to the users that have not been rating using the description of the item. Their model also works on the side of the item. For modeling the ratings, they have used a mixture of Gaussian, unlike above researchers that have used MF method. They have compared their model with the above described, HFT and CTR and has shown that their model has performed better in terms of rating prediction accuracy. CDL [32], Collaborative Deep Learning brought a change in the topic modeling approach by learning the numeric representation of textual data by applying deep learning techniques on the item description document. In this paper, they have proposed the hierarchal Bayesian Model, which applies deep learning on the textual data and Collaborative filtering approach on rating data in a joint manner. They have used SDAE, Stacked Denoising Autoencoders, neural network approach combined with the Collaborative Filtering method in order to integrate the rating and content information, thereby advancing the quality of recommender systems. Convolutional Matrix Factorization (ConvMF) [14], this model addresses the rating data sparsity by learning the text document using single CNN. The CNN considers the surroundings of the word order that has improved the representation of latent features in the description of the items' documents. Their model integrates PMF and CNN to upgrade the rating prediction task and hence, refine the quality of recommendation of items or movies to the end-users. LZhang et al. [35] proposed combining CF algorithm with an artificial neural network (ANN) to enhance the scalability and remove the sparsity of recommendation systems. Quadric Polynomial Regression model extracts the latent features, and these features then becomes the input to the deep neural network for predicting the rating scores. The model used to extract the hidden features improves the traditional matrix factorization method [35]. Dual-Regularized Matrix Factorization (DRMF) [34], this is the recent research work in the field of integrating textual data description with PMF. They have exploited the textual description of items and users both and then predicted the rating score. They have also adopted a unique multi-layered neural network model that combines CNN and bi-directional GRU that has given a better latent representation of the text. The content representation regularized the latent models for the items and users in MF.

Different expert recommendation systems have also been designed to recommend the experts in different domains for solving problems using the information retrieval process. It helps in the detection of knowledgeable people expert in his or her area. The combination of search engines and NLP can be used to retrieve the experts in this kind of recommendation systems [12]. Different tags and their syntactic patterns associated with Web 2.0 such as audio, video, movies can be exploited for recommendation to solve the cold start problem [1, 23].

This research paper is being inspired by the above-related work explained. Our proposed approach brings an improvement in the text analysis model so that the textual representation gets better, thereby improving the quality of recommendations.

3 Why capsule networks over CNN?

Capsule Networks were newly discovered Neural Network Architecture by Geoffrey Hinton in 2017. They have been applied in various domains such as Image Classification, Natural Language Processing (NLP), text classification, etc. [31]. This is a breakthrough in the domain

of Deep Learning. It has overcome the various drawbacks in the CNN and shown effective results in the classification problems. The max pooling function used in CNN to extract the highest features is not efficient and results in reduced computational complexity. It has been replaced by Dynamic Routing Algorithm that is very efficient in recognizing the spatial relationships and hence improving the classification accuracy. "Routing-by-agreement" proposed by Geoffrey Hinton, helps in retaining the relevant information and features of the content and sends the features from lower-level to higher level only if the content is being matched [6].

CNN does not consider the spatial information, whereas CN does, and that increases its power. Replacing the CNN by CN in the deep neural network text analysis model has increased the rating prediction accuracy and hence reducing the root mean squared error (RMSE). The experiment and results are presented in the further sections.

4 Why deep neural networks for text analysis?

In recent times, the research in deep learning has enhanced. It is being used in different domains such as computer vision, image classification, NLP, Recommender Systems, etc. They have improved the accuracy in different classification and regression problems. They are more complex to understand and take more time in training, but the results generated are more effective and precise. They perform better in extracting semantics from documents and hence, generating an effective text representation of documents [9]. As we saw Capsule Networks perform better than CNN, as they have a dynamic routing algorithm that simultaneously takes feedback from simple entities.

Similarly, Bi-directional GRU is good at temporal modelling, that is, it stores information related to past and present. These neural networks are good in feature extraction and dimensionality reduction. Hence, we have proposed a DNN model for text analysis and generating the representation of items' reviews and users' reviews.

5 Methodologies

This section discusses the different methodologies adopted in order to implement a product recommender system.

6 Probabilistic matrix factorization

PMF [25] is collaborative filtering, model-based technique. The task of the algorithm is to infer the preferences of the users based on the ratings given by users. And what other similar users have liked? In this algorithm, the rating matrix of items and users are given, but it is a sparse matrix because not every user rate every other movie. The main aim is to fill the missing entries to predict the rating that a user would give to an unseen movie and then recommend those to the user based on predicted rating score.

PMF factorizes the rating matrix into a product of two smaller (low-rank) matrices. Given a rating matrix of M products and N users, and integer rating 1 to 5, the matrix is factorized into two latent feature matrices of users and items.

The two users and items latent feature-based matrices is learned from the observed rating matrix. PMF models ratings as drawn from a normal distribution. User and item features are determined using mean values. It assumes that ratings are normally distributed.

$$\widehat{r}_{ui} \sim N\left(p_u q_i^T, \sigma^2\right) \tag{1}$$

Where r_{ui} represents predicted rating score, N represents normal distribution, p_u represents user-feature vector and q_i represents item-feature vector.

σ^2 represents variance used to model noise of rating matrix.

User Feature Vector can be represented as:

$$p_{uf} \sim N\left(0, \sigma_p^2\right) \tag{2}$$

Here, p_{uf} represents user feature vector, N represents the normal distribution, σ_p^2 represents variance used to model noise of user latent feature vector matrix.

Item Feature Vector can be represented as:

$$q_{if} \sim N\left(0, \sigma_p^2\right) \tag{3}$$

Here, q_{if} represents item feature vector, N represents the normal distribution, σ_p^2 represents variance used to model noise of item latent feature vector matrix.

In this research, we have compared our model with the basic Probabilistic Matrix Factorization approach. In this proposed integrated model, parameters can be learned using the MAP approach.

7 Deep neural network for text representation

The Deep Neural Network (DNN) Text Analysis Model is being used for generating the text representation of both items and users. A unique and powerful DNN model has been implemented to learn the hidden features of textual reviews. The DNN Text Analysis Model first goes through three Capsule Network (CN) Layers and learns the text representation with these CNs. After learning the semantics and orientation of the text by concatenating three CN Layers, the text is encoded into numbers using Bi-directional Gated Recurrent Unit (Bi-GRU), a variation of RNN. The Framework of DNN Text Analysis Model consists of the Embedding Layer first, and then three CN Layers stacked on a Bi-directional GRU Layer, and finally the Output Layer that projects the hidden features of the reviews.

8 Embedding layer

The input data needs to be encoded in integer; the algorithm does not take strings as input and work on it. Here, the Embedding Layer in Keras is essential. It helps in providing a dense representation of words and learns the relationship between the words. In this research, we have used a pre-trained embedding (Glove) [20] to get the dense representation of the reviews

of items and users and learn the relationship between the words in the reviews. It helps in extracting the semantic relationship between the words. It encodes the meaning of words into vectors.

9 Capsule networks

Capsule Networks [24] are newly developed Deep Neural Network architecture introduced by Geoffrey Hinton in 2017. These networks brought an improvement in the Convolutional Neural Networks (CNN). Capsule Networks have overcome the limitations observed in CNN's. The main advantages of CNs that makes it different from CNN and more efficient are dynamic routing and equi-variance. Dynamic routing Algorithm [7] is the central concept behind capsule networks. Analogous to backpropagation in Artificial Neural Network, Dynamic Routing between capsules help the capsules in communicating with each other and creating the text representations [27].

10 Routing by agreement

A capsule network is built up by combining several layers. The Capsules present in the lower level of capsule network architecture correspond to simple entities that further combine to form an image. For example, in the image of a house (complex entity), the two smaller entities are rectangle and triangle. According to the concept of Routing by Agreement as specified by Geoffrey Hinton, the capsules at lower level bet on the presence of a complex entity at a higher level. Moreover, these bets are combined to get the output of capsules at a higher level [30]. For example, the presence of two simple entities, rectangle, and triangle, works with each other to bet on the presence of a complex entity at a higher level, i.e. house. A coupling effect is observed in this concept. Here, capsules at lower level try to predict the output of the capsule at the higher level and if the predictions made by capsules at lower level conform to the actual output of capsule at a higher level, the coupling coefficient between the two capsules increases.

11 Mathematics behind capsule networks

Consider two layers t and t + 1 having m and n capsules respectively. The main aim is to compute the activations of the capsules at layer t + 1 given the activations at lower layer t. Let u denotes the activations of capsules at layer t. The task is to calculate v, activations of capsules at layer t + 1 [7].

For a capsule j at layer t + 1,

12 Calculation of prediction vectors

The first step involves the calculation of the prediction vectors of the upper layer using the capsules at lower layer. The prediction vector is calculated by a capsule i (of layer t) for the capsule j (of layer t + 1):

$$\overline{U_{j\backslash i}} = W_{ij}U_i \tag{4}$$

Where Wij is the weight matrix

13 Calculation of the output vector

The output vector s_j is calculated using the weighted sum of all the prediction vectors given by the capsules of layer t for the capsule j [19].

$$s_j = \sum_{i=1}^{m} c_{ij}\overline{u_{j\backslash i}} \tag{5}$$

The scalar cij is the coupling coefficient between capsule i (of layer t) and j (of layer t + 1). The iterative dynamic routing algorithm determines the coupling coefficient.

14 Apply the squashing function

After calculating the output vector, the squashing function is applied to the output vector v_j of the capsule j:

$$\mathrm{v_j} = \mathrm{squash}\left(\mathrm{s_j}\right) \tag{6}$$

Squash Function performs the non-linear transformation and works as an activation function for capsule networks like ReLU function used in convolution networks.

15 Bi-directional gated recurrent unit

We have used Bi-Directional Gated Recurrent Unit stacked on Capsule Networks to enhance the text representation. Gated Recurrent Unit (GRU) solves the different problems observed in RNNs [29] such as long-term dependency and vanishing/ exploding gradient problem. These are the most effective neural networks when the context of the input is required. GRU uses two gates, update gate and reset gate in order to maintain the information passed long ago.

Mathematically, Update gate is represented using z_t for the time step t using the following equation,

$$z_t = \sigma\left(W^{(z)}x_t + U^{(z)}h_{t-1}\right) \tag{7}$$

Where x_t represents input unit, W (z) represents the weight, h_t-1 represents the information from previous input units and U (z) is the weight associated with the previous input units. They are both added, and the output is obtained between zero and one due to the sigmoid function applied.

The update gate reduces the vanishing gradient problem as it helps in determining the amount of past information required to be forwarded in the time ahead [5].

The Reset gate also plays an important role to forget the not so required past information. The equation for the reset gate is:

$$r_t = \sigma\left(W^{(r)}x_t + U^{(r)}h_{t-1}\right) \tag{8}$$

Where x_t represents input unit, W(r) represents weights associated with the input unit, h_t-1 represents information from previous input units and U(r) represents weights associated with the previous input units. The sigmoid function is applied after summing the results of the input unit and previous input units.

Now, the following formula shows the final output that would be generated using the reset gate that restores the past information and removes the irrelevant information from the past [25]. The new memory content generated using the Reset gate is given as:

$$h_t^{'} = \tanh(W_{x_t} + r_t \odot U h_{t-1}) \tag{9}$$

The h_t' represents the ongoing memory content. The present memory content is generated by multiplying the corresponding weights to the input x_t and multiplying the corresponding weight of previous input units. After that, the element-wise product is calculated between reset gates and weighted previous input units. Then, the non-linear activation function tanh is applied to generate the results. The closing Memory at the present time step t is calculated using the following equation:

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot h_t^{'} \tag{10}$$

In this equation, h_t represents the closing memory content vector generated, z_t represents the update gate and h'_t represents present memory content.

16 Output layer

After going through the embedding layer, Capsule Network Layer and bi-GRU Layers, the features extracted by the last layer are outlined on a k-latent dimensional space by using hyperbolic tanh activation function.

$$\alpha = \tanh\left(G^{*}h + b\right) \tag{11}$$

Here, G represents the projection matrix obtained in the output Layer, h represents the output obtained from the bi-directional GRU layer, and b is the bias. Dropout [26] is also used in the model to prevent the over-fitting in the projection matrix. The main aim behind Dropout is dropping hidden or visible neurons in a neural network model to prevent the over-fitting. Dropout is a regularization technique that helps in reducing the interdependence between the nodes in the neural network.

The above-mentioned text analysis neural network layers described converting a raw document into a densely distributed representation, as shown in Fig. 1. Given X, the document of items' reviews and Y, the documents of users' reviews, the DNN text analysis model transforms X to beta and Y to alpha. Alpha and beta are the densely distributed representation of the raw documents of users and items, respectively.

Keras functional API [4] is being used to define the neural network model for text analysis and getting the projection matrix for the further process. The model parameters are learnt by applying the back-propagation algorithm. The input sequence, output, and the corresponding loss function are defined for the model. In the Proposed Work section, the algorithm is being explained used for implementing the DNN text analysis model based on Keras.

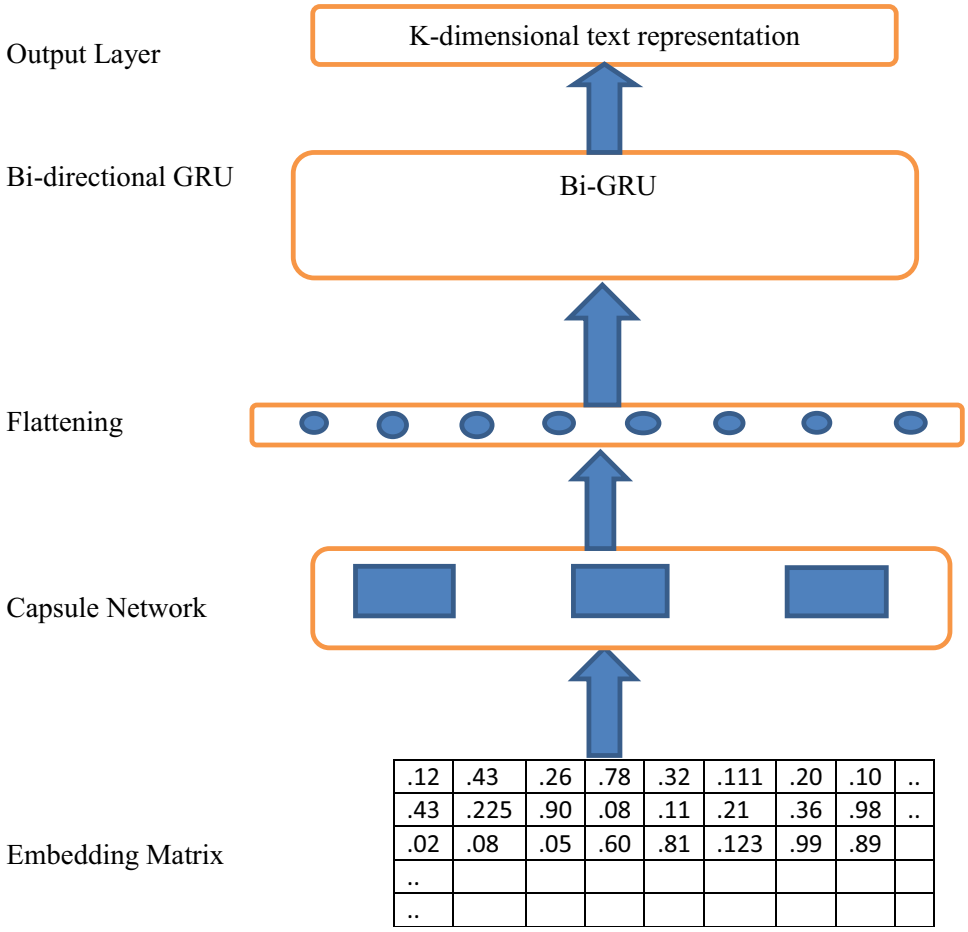


Fig. 1 Deep Neural Network text analysis model generating the representation of users' and items' reviews

17 Proposed work

In this segment, we have described the algorithm implemented for text analysis of reviews of items and users of different products. The matrix obtained from the text analysis is then integrated with the collaborative filtering based PMF method to generate more accurate and precise recommendations.

18 The model

The Algorithm gives the details about the implementation of Text Analysis Neural Network Model that involves different layers starting from Embedding Layer, Capsule Network Layers and Bi-directional GRU layers that extract features from the description documents and project them into k-dimensional space. It outputs a latent document k-dimension matrix that helps in improving the product recommendations.

The earlier similar models have used convolutional neural networks (CNN) [37] for latent document matrix that had certain limitations. These limitations have been overcome using capsule network layers.

The Algorithm is documented below, showing all the different layers that have been applied and implemented to get the k-dimensional matrix of the user and item reviews document. The document is first converted into a numerical matrix using the embedding layer representing the

text as a sequence of word embedding. Then, the document representation is learned through capsule networks, and the semantics of the reviews are encoded by going through the bi-directional GRU layers.

Algorithm: Deep neural network (dnn) text analysis model

```
Input: vocabulary size, e, d, s, gru_dim, k, init_W, n_capsule, n_routings, caps_dim;
Output: DNN text analysis Model
{Embedding Layers}
input_sequence = Input (shape= (d * s,))
seq_emb = Embedding (vocabulary size, e, weights= [init_W / 20], trainable=False) (input_sequence)
{Capsule Layers}
caps_1 = Capsule (num_capsule=n_capsule, dim_capsule = caps_dim,
routings = n_routings, share_weights=True) (seq_emb)
x = Flatten () (caps_1)
caps_2 = Capsule (num_capsule=n_capsule, dim_capsule=caps_dim,
    routings = n_routings, share_weights=True) (x_3)
x = Flatten () (caps_2)
caps_3 = Capsule (num_capsule=n_capsule, dim_capsule=caps_dim,
    routings=n_routings, share_weights=True) (caps_3)
x = Flatten () (caps_3)
x = concatenate ([caps_1, caps_2, caps_3])
{Bi-directional GRU Layers}
x_gru = Bidirectional (GRU (gru_dim, return_sequences=True)) (x)
x = Flatten () (x_gru)
{Output Layer}
seq_dropout = Dropout (dropout_rate) (x)
out_seq = Dense (output_dimesion, activation='tanh') (seq_dropout)
dnn_text_analysis = Model (input_sequence, out_seq)
dnn_text_analysis (optimizer='rmsprop', loss='mse')
return dnn_text_analysis;
```

19 CapsMF: integration of deep neural network text analysis model and PMF

Here, we describe the integration of PMF with the text analysis model and the equations involved in generating user latent feature vector and item latent feature vector. The main aim of the MF technique is to generate user and item hidden feature vectors [16] that particularly describe the choices of the users, and they are obtained from both the ratings and the reviews of items and users.

20 User latent vector

User i as a latent vector represented as:

$$p_i = €_i + \alpha_i \tag{12}$$

Where p_i represents user latent vector, $€ \sim N(0, \sigma_p^2 I)$, α_i represents user document latent vector.

Integrating the above we obtain the user latent model as:

$p_i \sim N(\alpha_i, \sigma_p^2)$, here I represent the identity matrix.

21 Item latent vector

Item j as a latent vector represented as:

$$q_j = €_j + \beta_j \tag{13}$$

Where q_j represents item latent vector, $€ \sim N(0, \sigma_q^2 I)$, β_j represents item document latent vector.

Integrating the above, we obtain item latent model as:

$q_j \sim N(\beta_j, \sigma_p^2 I)$, here I is the identity matrix.

22 Prediction

Suppose, M represents the number of users, and N represents the number of items, R represents the given rating matrix. The missing rating in the rating matrix is predicted using the formula "P^TQ", where P represents the user latent matrix of dimension k X M and Q represents the item latent matrix of dimension k X N.

$$\overline{R} = N\left(P_i^T Q_j, \sigma^2\right) \tag{14}$$

Next is the Flow Chart representing the Learning of different parameters in this model.

Figure 2 above describes the learning of different parameters used in the model "CapsMF" using Maximum a Posteriori (MAP) approach. First, rating matrix, item document, and user document are given as input to the model. Then, Item Content and User Content are pre-processed by removing the stop words and applying lemmatization and convert into embedding input using a pre-trained embedding (Glove) [20] to get the dense representation of the

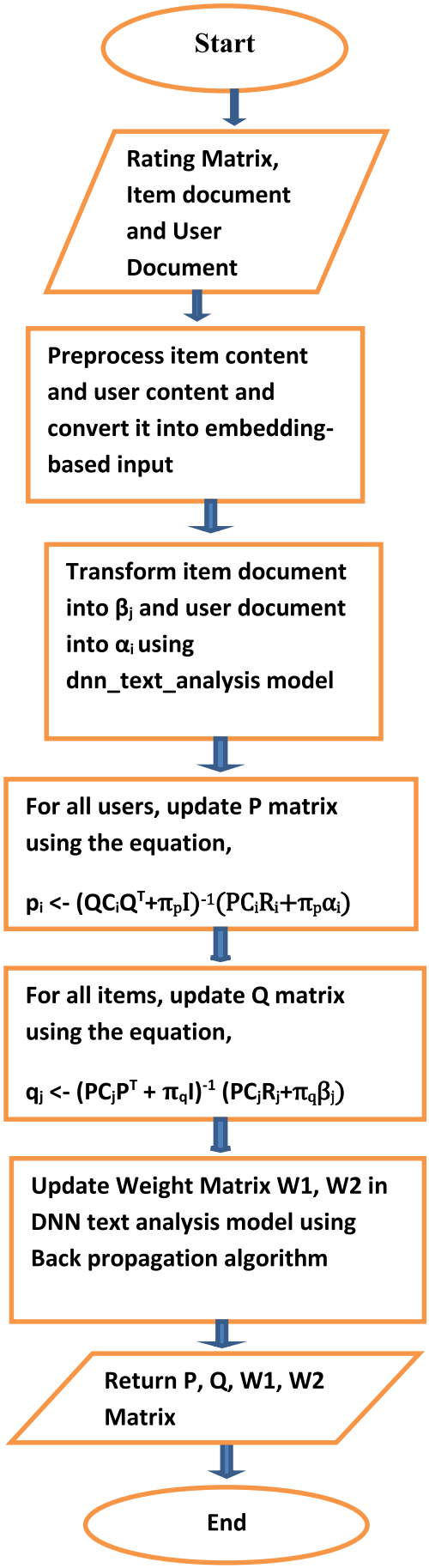


Fig. 2 Learning parameters in the proposed model CapsMF

Table 1 Description of the datasets

Datasets	Number of Users	Number of Items	Total Reviews/Ratings	Rating Scale
AIV (Amazon Instant Video)	5130	1685	37,126	1–5
AA (Apps for Android)	87,721	13,209	752,937	1–5

reviews of items and users. The document matrices generated for users and items and the text representation is learned through DNN text analysis model. Item document is transformed into β_j, and user document is transformed into α_i using the algorithm described in Section 6.1. After transforming the item reviews document and user reviews document, the user latent matrix and item latent matrix is updated using the equations given in the flow chart. The Weight Matrices W1 and W2 applied in DNN text analysis model are updated using Back Propagation Algorithm. When the number of iterations is over, return the User Latent Matrix, Item Latent Matrix, and Weight Matrices.

23 Experiment and results

We have conducted the experiments on the two amazon datasets, one is "Amazon Instant Video (AIV)", and the other is "Apps for Android (AA)". Both the datasets are taken from Amazon Product Reviews site (http://jmcauley.ucsd.edu/data/amazon/). This site contains reviews of different kind of products as well as metadata regarding the product. We have considered only the reviews and ratings of the products. We have taken 5-core datasets, in these 5-core datasets, the users and items have five reviews each. Table 1 describes the datasets that have been used for implementation. The Rating is given from one to five to an item by a user.

We have conducted the experiments on Google Colaboratory Notebook using GPU since the dataset is too large and deep learning models take time in training. We have used different evaluation metrics to compare the proposed model with the recommendation models already published. Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) have been extensively used for comparing the performance of rating predictions.

Table 2 Parameter settings of π_p and π_q for rating prediction

	Amazon Instant Video		Apps for Android	
Models	Π_p	Π_q	Π_p	Π_q
PMF	0.1	0.01	0.1	0.01
DRMF-Item	1	10	1	100
DRMF-User	10	0.1	100	0.1
DRMF	100	10	10	100
CapsMF	100	10	10	100

Where $\Pi p = \sigma^2 / \sigma_p^2$ and $\Pi q = \sigma^2 / \sigma_q^2$.

σ^2 represents variance used to model noise of rating matrix, σ_P^2 represents variance used to model noise of user latent feature vector matrix and, σ_P^2 represents variance used to model noise of user latent feature vector matrix.

Also, parameters related to Capsule Layer are set accordingly for "CapsMF":

$n_c = 50$; $n_r = 5$; $n_d = 100$

Here, n_c = number of capsules, n_r = number of routings and n_d = Capsule Dimension.

Table 3 Parameter settings of Π_p and Π_q for top-N recommendation

	Amazon Instant Video		Apps for Android	
Models	Π_p	Π_q	Π_p	Π_q
PMF	1	5	5	10
DRMF-Item	0.5	5	10	1
DRMF-User	5	0.5	10	1
DRMF	5	0.5	10	1
CapsMF	5	0.5	10	1

Where $\Pi p = \sigma^2 / \sigma_p^2$ and $\Pi q = \sigma^2 / \sigma_q^2$.

σ^2 represents variance used to model noise of rating matrix, σ_P^2 represents variance used to model noise of user latent feature vector matrix and, σ_P^2 represents variance used to model noise of user latent feature vector matrix.

Also, parameters related to Capsule Layer are set accordingly for "CapsMF":

$n_c = 50$; $n_r = 5$; $n_d = 100$

Here, n_c = number of capsules, n_r = number of routings and n_d = Capsule Dimension.

$N = 300$ for top-N recommendations.

We have also calculated the precision and recall for assessing the accuracy of top- n recommendations.

24 Different evaluation metrics

The formulas used for different evaluation metrics is described here:

$$\mathrm{MAE} = \sum_{i,j} \left| r_{ij} - \overline{r_{ij}} \right| / N \tag{15}$$

Where r_{ij} represents the actual rating, r'_{ij} represents the predicted rating and N represents the count of samples.

$$\mathrm{RMSE} = \text{square root} \left(\sum_{i,j} \left(r_{ij} - r'_{ij} \right)^2 \right) / N \Big) \tag{16}$$

Precision and Recall are calculated to assess the accuracy of top-N recommendations, we have calculated these metrics for top-300 recommendations given to the users.

$$\text{Precision@N} = (\#\text{of recommended items@N that are relevant}\} / (\#\text{of recommended items@N}) \tag{17}$$

Table 4 Rating prediction performance of different models on AIV dataset

Models	MAE	RMSE
PMF	0.9627	1.2088
DRMF-Item	0.7419	1.0101
DRMF-User	0.7325	0.9866
DRMF	0.7190	0.9661
CapsMF	0.7064	0.9593

Table 5 Rating prediction performance of different models on AA dataset

Models	MAE	RMSE
PMF	1.1829	1.4773
DRMF-Item	1.0664	1.3763
DRMF-User	0.9638	1.2617
DRMF	0.9071	1.1930
CapsMF	**0.8878**	**1.157**

The best values are shown in bold.

According to the accuracy comparisons shown in the tables above, it can be concluded that CapsMF, the model described has shown improvement in terms of MAE and RMSE on the two datasets: AIV and AA.

$$\text{Recall@N} = (\#\text{of recommended items@N that are relevant})/(\text{total}\#\text{of relevant items}) \quad (18)$$

25 Parameter settings

Different values that have been set for π_p and π_q are given in Tables 2 and 3.

26 Results

The different values of evaluation metrics obtained on implementing the different models are shown in this section. The baseline models are PMF, DRMF-Item, DRMF-User, and DRMF, as explained earlier in the related work. For the right results, we have averaged the results of five-folds in order to estimate the prediction and recommendation performance accurately.

Below Table 4 shows the results of rating prediction performance of different models on AIV dataset in term of MAE and RMSE.

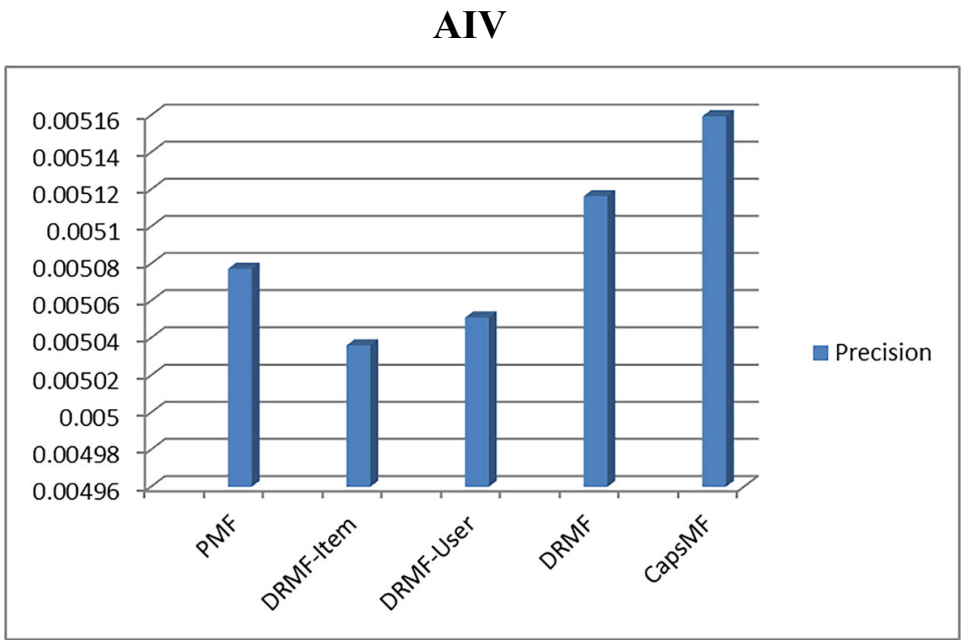


Fig. 3 Top-300 ranking performance in terms of precision

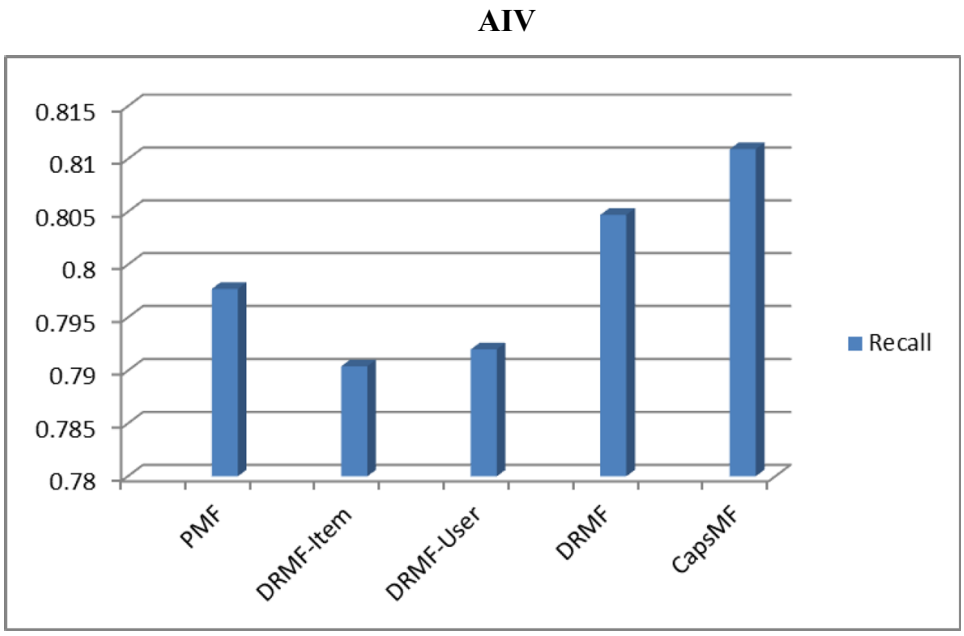


Fig. 4 Top-300 ranking performance in terms of recall

Table 5 shows the results obtained when rating prediction is implemented using different Models on AA dataset. Rating Prediction performance of different models is shown based on MAE and RMSE, respectively.

27 Graphical analysis of results

Here, we have shown the ranking performance of different Models on the two datasets AIV and AA graphically in term of Precision and Recall.

Below Figs. 3 and 4 graphically represent the top-300 recommendation performance of different Models on AIV dataset in term of precision and recall respectively.

Figures 5 and 6 show the results obtained while implementing the task of Top-N recommendations on AIV dataset in term of precision and recall, respectively. Here, the value of N varies from 50 and ends at 300. It is noteworthy that there is not much gain in top-n recommendation performance.

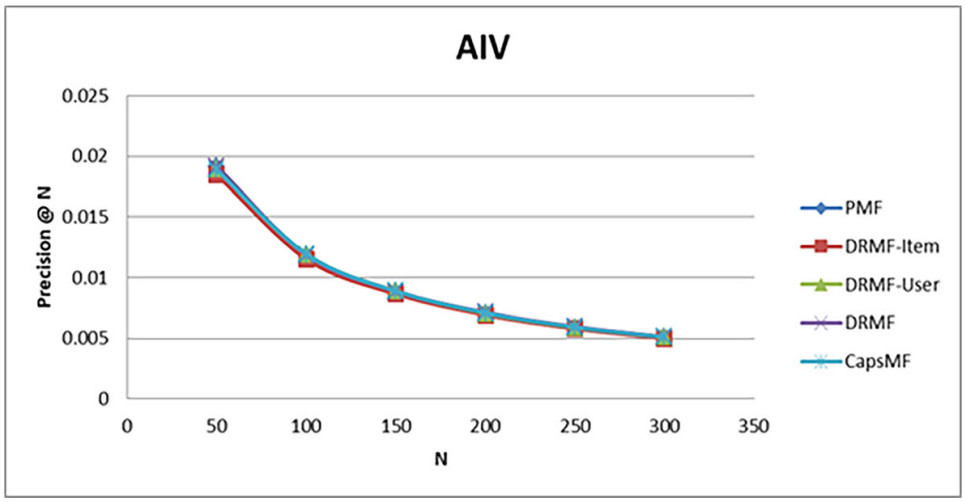


Fig. 5 Top – N recommendation in term of precision

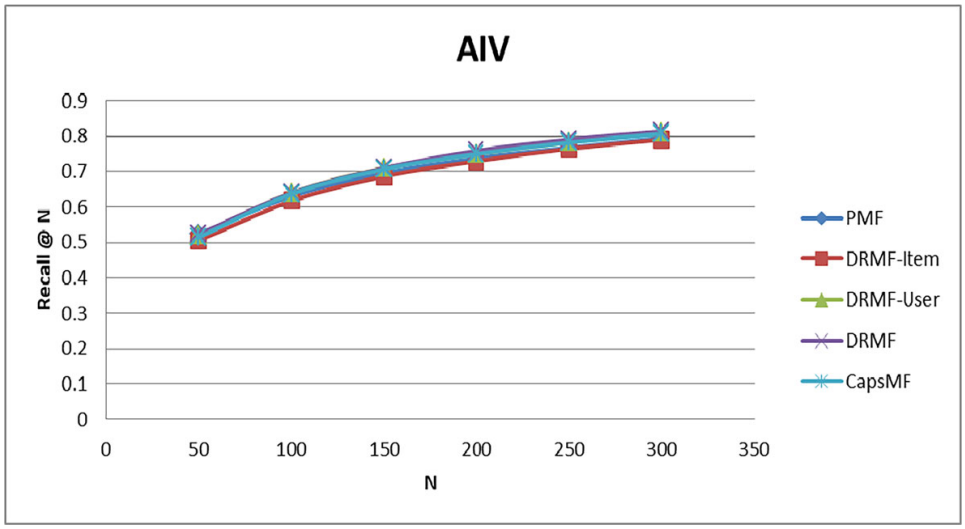


Fig. 6 Top-N ranking performance in term of recall

We can observe in the charts above that the proposed approach "CapsMF" has shown better recommendation performance on AIV dataset. Both RMSE and MAE have smaller values than the baseline Models implemented. All the models have been implemented on Google Colaboratory Notebook in Python using GPU for getting the results faster. Higher values of Precision and Recall prove that recommendation performance has improved over the baseline Models described in the related work.

Following are the graphical representations of the results obtained on "Apps for Android (AA)" dataset. Figure 7 shows the ranking performance of top-300 recommendations in term of Precision graphically on the AA dataset.

Figures 8 and 9 graphically represents the ranking performance of items in term of Precision and Recall on AA dataset. The value of N in Top-N recommendation performance varies from 50 to 300.

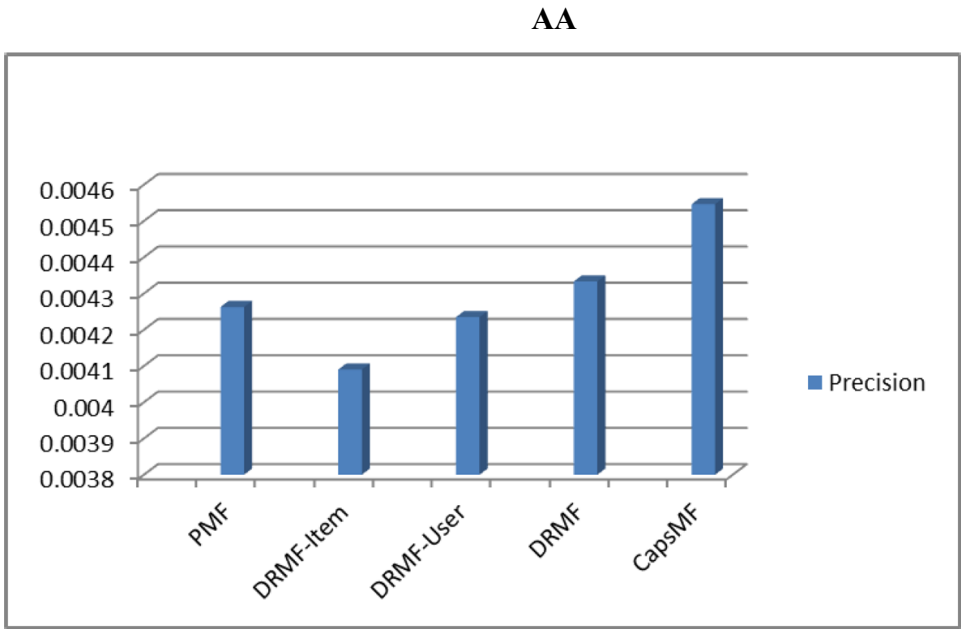


Fig. 7 Top-300 recommendation performance in terms of precision

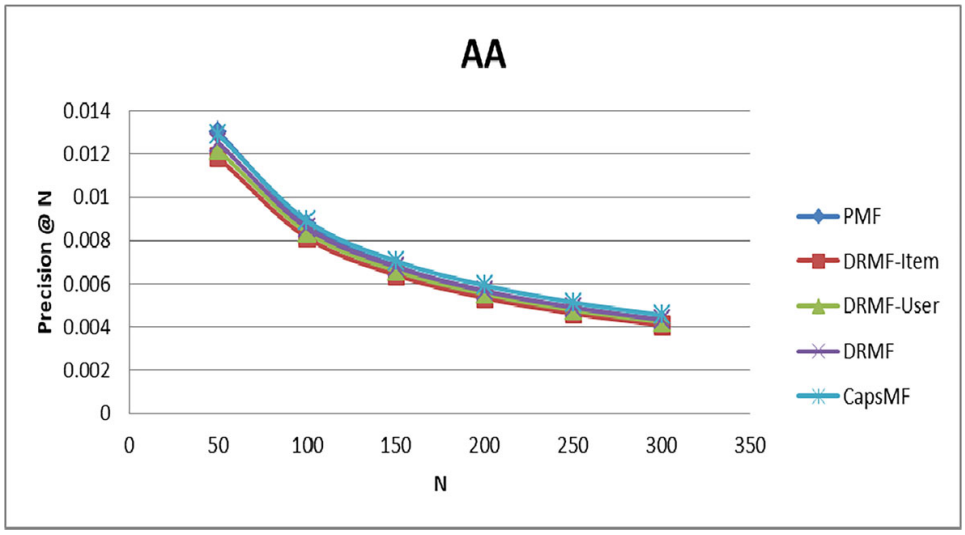


Fig. 8 Top-N recommendation performance in term of precision

28 Performance comparisons

Above we have shown the results in the tables and analysis of the different models on rating prediction and top-n recommendations using column charts. As we can observe that CapsMF, the proposed deep neural network model for text analysis has shown better performance when integrated with Matrix factorization (MF) Technique. For the task of rating prediction on AIV dataset, CapsMF has shown improvement over PMF, DRMF-Item, DRMF-Use, and DRMF by 26%, 4.7%, 3.5%, and 1.7% respectively in term of MAE. In term of RMSE, CapsMF has improved over PMF, DRMF-Item, DRMF-User, and DRMF by 20.6%, 5.02%, 2.76%, and 0.7% respectively.

For the task of top-n recommendation on AIV dataset, CapsMF has shown significant improvement when compared with the other models of similar type in terms of both precision and recall. We have chosen $n = 300$. It has improved by 1.6% when compared with PMF and 2.3%, 2%, and 0.8% improvement w.r.t DRMF-Item, DRMF-User and DRMF in term of Precision. Similar significant improvements have been observed in terms of Recall @ 300 on AIV Dataset.

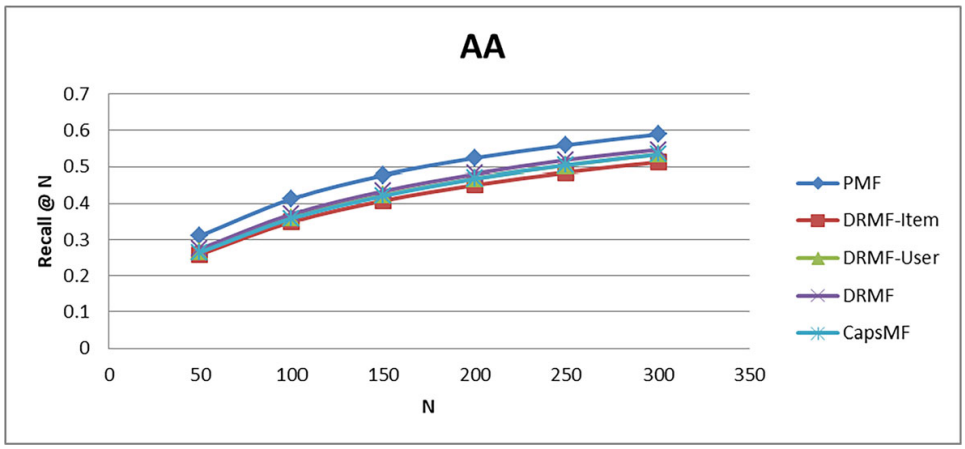


Fig. 9 Top-N ranking in term of recall similarly, when the proposed approach "CapsMF" is implemented on Apps for Android (AA) dataset has shown improvement over the baseline models on two different tasks, one of rating prediction and another, the ranking performance of items that are being recommended to the users

When the task of rating prediction and top-n recommendation are implemented on "Apps for Android (AA)" dataset, significant improvements have been recorded. CapsMF has shown better rating prediction accuracy over the baseline models. It has outperformed the baseline models PMF, DRMF-Item, DRMF-User, and DRMF by 25%, 16.7%, 7.8%, and 2.12% respectively in term of MAE. Similarly, RMSE values have also been reduced when calculated on the models by 21.6% when compared with PMF, 15.9%, 8.29%, 3.01% when compared with DRMF-Item, DRMF-Use, and DRMF respectively.

We can observe that the proposed model CapsMF has shown better results on a larger dataset, "Apps for Android (AA)".

29 Conclusion and future work

In this research paper, we have addressed the limitations of CNN based document modeling approach as well as CNN and RNN based text representation approach. We have proposed a novel and powerful DNN text analysis model stacking Capsule Networks and Bi-directional RNN to generate the textual representation of users' and items' reviews integrated with PMF to improve recommendation performance. Experiments results have proved that exploiting ratings, items' and users' reviews have improved rating prediction accuracy and precision and recall of top-n recommendations. It can also be concluded that the proposed model, "CapsMF", performs better on more massive datasets as compared to smaller datasets. Both the users and items reviews have been considered to build the separate deep neural network for text analysis. Hence, it has resulted in better recommendations to the users and also solves the cold-start problem in recommender systems.

There are still certain limitations that come with Capsule Networks; one that has been observed while performing the experiments is that capsule networks take times in training. The advantage is that it takes fewer data points in training as compared to CNN.

In future works, we would refine the deep neural network for text analysis with new deep learning technologies getting discovered for analysis of text for getting a better representation of the text. We would apply the proposed model in different recommendation datasets and scenarios such as social recommendations [3], joint-recommendation, group recommendation [33], etc. Other contextual information related to users and items such as images, tweets [10, 13], metadata about products can be integrated with collaborative based filtering technique to enhance the personalized recommendations.

Compliance with ethical standards

Conflict of interest There is "NO any" direct or indirectly related conflict of interest for this manuscript and with authors.

References

1. Belém FM, Heringer AG, Almeida JM, Gonçalves MA (2019) Exploiting syntactic and neighborhood attributes to address cold start in tag recommendation. Inf Process Manag 56(3):771–790
2. Bobadilla J, Ortega F, Hernando A, Gutiérrez A (2013) Knowledge-based systems recommender systems survey. Knowledge-Based Syst 46:109–132

3. Chen K, Chen T, Zheng G, Jin O, Yao E, Yu Y (2012) Collaborative personalized tweet recommendation. Proc. 35th Int. ACM SIGIR Conf Res Dev Inf Retr - SIGIR '12, no 800, p 661
4. Chollet F et al. Keras: deep learning library for theano and tensorflow. https://keras.io/
5. Chung J (2014) Gated recurrent neural networks on sequence modeling. arXiv : 1412. 3555v1 [cs . NE], pp 1–9
6. Hinton GE, Krizhevsky A, Wang SD (2011) Transforming auto-encoders. In: Honkela T, Duch W, Girolami M, Kaski S (eds) Artificial neural networks and machine learning – ICANN 2011. Lecture notes in computer science, vol 6791. Springer, Berlin, Heidelberg
7. Hinton G, Sabour S, Frosst N (2018) Matrix capsules with EM routing. International Conference on Learning Representations (ICLR), pp 1–15
8. Jaiswal A, AbdAlmageed W, Natarajan P (2018) CapsuleGAN: generative adversarial capsule network. Cite arxiv:1802.06167 pp 1–10
9. Johnson R, Zhang T Effective use of word order for text categorization with convolutional neural networks. In Proceedings of the 2015 Conference of the North {A}merican Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 103–112. https://doi.org/10.3115/v1/N15-1011
10. Katarya R, Arora Y (2018) A survey of recommendation systems in twitter. 2018 4th international conference on Computational Intelligence & Communication Technology (CICT), Ghaziabad, pp 1–5. https://doi.org/10.1109/CIACT.2018.8480264
11. Katarya R, Arora Y (2019) Study on text classification using capsule networks. Accepted in International Conference on Advanced Computing & Communication Systems (ICACCS), IEEE
12. Khasmakhi NN, Balafar MA, Derakhshi MRF (2019) Engineering Applications of Artificial Intelligence The state-of-the-art in expert recommendation systems ☆. *Eng. Appl. Artif. Intell.* 82(August 2018):126–147
13. Khater S, Elmongui HG, Gra D (2014) Tweets you like: personalized tweets recommendation based on dynamic users interests. IEEE, pp 1–10
14. Kim D, Park C, Oh J, Lee S, Yu H (2016) Convolutional matrix factorization for document context-aware recommendation. Proceedings of the 10th ACM Conference on Recommender Systems, page 233–240. New York, NY, USA, ACM
15. Kong X, Member S (2018) A survey of collaborative filtering-based recommender systems : From traditional models to hybrid models based on social networks. IEEE Access 6:64301–64320
16. Koren Y, Bell R, Volinsky C (2009) Matrix Factorization Techniques for Recommender Systems. Computer 42(8):30–37. https://doi.org/10.1109/MC.2009.263
17. Ling G, Lyu MR, King I (2014) Ratings meet reviews, a combined approach to recommend. In Proceedings of the 8th ACM Conference on Recommender systems (RecSys '14). ACM, New York, NY, USA, 105–112. https://doi.org/10.1145/2645710.2645728
18. Liphoto M, Du C, Ngwira S (2016) A survey on recommender systems. 2016 International Conference on Advances in Computing and Communication Engineering (ICACCE), Durban, pp 276–280. https://doi.org/10.1109/ICACCE.2016.807376
19. Mcauley J (2013) Hidden factors and hidden topics : understanding rating dimensions with review text. Proceedings of the 7th ACM Conference on Recommender Systems, RecSys '13, pages 165–172, ACM
20. Pennington JJ, Socher R, Manning CD (2014) GloVe : global vectors for word representation. Empirical Models in Natural Language Processing (EMNLP), pp 1532–1543
21. Ramlatchan A, Yang M, Liu Q, Li M, Wang J, Li Y (2018) A survey of matrix completion models for recommendation systems. In Big Data Mining and Analytics, vol 1, no 4, pp 308–323(IEEE), December. https://doi.org/10.26599/BDMA.2018.9020008
22. Razghandi M, Golpaygani SAH (2017) A context-aware and user behavior-based recommender system with regarding social network analysis. 2017 IEEE 14th Int Conf E-bus Eng pp 208–213
23. Rojo D, Valdiviezo-díaz P (2018) Hybrid collaborative filtering based on users rating behavior. IEEE Access vol 6
24. Sabour S, Nov CV, Hinton GE (2017) Dynamic routing between capsules. arXiv:1710.09829, Nips
25. Salakhutdinov R, Mnih A (2007) Probabilistic matrix factorization. In Proceedings of the 20th International Conference on Neural Information Processing Systems (NIPS'07), Curran Associates Inc., USA, 1257–1264
26. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: A simple way to prevent neural networks from overfitting. J Mach Learn Res 15(1):1929–1958
27. Srivastava S, Khurana P, Tewari V (2018) Identifying aggression and toxicity in comments using capsule network. Proceedings of the First Workshop on Trolling, Aggression, and Cyberbullying (TRAC-2018), pp 98–105

28. Steyvers M, Smyth P, Rosen-Zvi M, Griffiths T (2004) Probabilistic author-topic models for information discovery. Proc. 2004 ACM SIGKDD Int Conf Knowl Discov Data Min. - KDD '04, no. 1990, p 306
29. Van Merri B, Fellow CS (2014) Learning phrase representations using RNN encoder – decoder for statistical machine translation. In Proceedings of the 2014 Conference on empirical models in natural language processing (EMNLP), Association for Computational Linguistics, pp 1724–1734, October
30. Wang C, Blei DM (2011) Collaborative topic modeling for recommending scientific articles. Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, page 448–456, New York, NY, USA, ACM
31. Wang Y, Sun A, Han J (2018) Sentiment analysis by capsules. WWW '18 Proceedings of the 2018 World Wide Web Conference, Pages 1165–1174 vol 2
32. Wang H, Wang N (2015) Collaborative deep learning for recommender systems. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 1235–1244. https://doi.org/10.1145/2783258.2783273
33. Wang W, Zhang G, Lu J (2016) Member contribution-based group recommender system. Decis Support Syst 87:80–93
34. Wu H, Zhang Z, Yue K, Zhang B, He J, Sun L (2018) Knowledge-based systems dual-regularized matrix factorization with deep neural networks for recommender systems R. Knowledge-Based Syst. 145:46–58
35. Zhang L, Luo T, Zhang F, Wu Y (2018) A recommendation model based on deep neural network. IEEE Access 6:9454–9463. https://doi.org/10.1109/ACCESS.2018.2789866
36. Zhang S, Yao L, Sun A, Tay Y (2019) Deep Learning Based Recommender System: A Survey and New Perspectives. ACM Comput. Surv. 52:5:1–5:38. https://doi.org/10.1145/3285029
37. Zheng L, Noroozi V, Yu PS (2017) Joint deep modeling of users and items using reviews for recommendation. In proceedings of the tenth ACM international conference on web search and data mining (WSDM '17). ACM, New York, NY, USA, 425–434. https://doi.org/10.1145/3018661.3018665.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Rahul Katarya[1] · **Yamini Arora**[2]

[1] Department of Computer Science and Engineering, Delhi Technological University, Delhi, India

[2] Department of Information Technology, Delhi Technological, Delhi, India

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Citation-Based Scientific Paper Summarization Using Game Theory

Anil Singh Parihar[1], Anjali Jain[2], Archit Gupta[3]
Department of Computer Science and Engineering
Delhi Technological University
New Delhi, India
[1]parihar.anil@gmail.com, [2]anjali7230@gmail.com, [3]archit.g1997@gmail.com

***Abstract*—Given the vast amount of research-oriented literature, any new research work asks for a comprehensive literature review. This process is typically very time-consuming. Citation information of a paper is a reflection of its contributions to the researcher's group. We use the citations of the paper to generate a comprehensible and informative summary, which will reduce the time taken to review the paper. In this paper, we propose a novel evolutionary game-theoretic approach to summarize scientific papers. A set of two-player games is played to choose the most relevant sentences in the paper that should be a part of the summary. Our method doesn't require an annotated dataset with gold summaries, thus levitating the need for a large, manually created dataset, which would be a primary requirement for any supervised approach. We have evaluated our results against human-expert generated summaries and state-of-the-art methods.**

Keywords—Extraction based text summarization, game theory

I. INTRODUCTION

Scientific articles are being published at a rate faster than ever. Whenever someone tries to enter a new field of research, they have to go through a large amount of literature to become familiar with the current progress achieved in that area. Reducing the complexity of this task is the main motive behind developing automatic summarizers for scientific papers. Scientific paper summarization is divergent from summarization of general articles [1] as these are divided into different sections, involve complex technical terms, and are much longer as compared to general articles. As of now, Scientific paper summarization follows two strategies. One of them uses the paper's abstract. Now, while the authors' perspective of the contents of a paper is concisely reflected in its abstract, it is also essential to include the views of the scientific community and the impact the paper has had on it. As a remedy to this problem, the second strategy, i.e., citation-based summaries, was proposed [2]. The citation summaries are formed by using a set of citations of an article.

We propose an approach that combines both these strategies. Also, instead of using citation sentences directly (as they may not contain enough relevant information or are even inaccurate sometimes), we follow the methodology of [3] , by making use of cited text spans, i.e., A collection of sentences in the cited paper that is referred to, by the citing sentences from other papers, for generating summaries. Our model takes a reference paper to be summarised and constructs its summary by aggregating and integrating contents from the abstract, and the cited text spans using game theoretical techniques. Hence it includes both the author's vision and the scientific communities' perspective.

We use the SciSummNet corpus as described in [4] as our dataset. It provides over 1,000 papers in the ACL anthology network with their citation networks (e.g., citation sentences, citation counts) and their comprehensive, manual summaries. For evaluation, we use the benchmark gold summaries provided in the CL-SciSumm shared task [5] . The evaluations show that our method outperforms the existing state-of-the-art methods and provides summaries which inculcate the outlook of the author and the research community.

II. RELATED WORK

Early works in scientific article summarization [6, 7] , produced the summaries of target papers based only on their contents. [1] formulated these summaries by identifying suitable content from the papers using a naïve Bayes classifier. But, the discourse of the paper didn't provide any information about the impact of the paper on the scientific community and their views on it. As a solution to this problem, citation-based summarization was proposed. A summary of an article created by using a set of sentences in other articles that cite that article is called its citation summary. According to [8] , citation summaries, while overlapping with the abstract of an article, also focus on other aspects and provide additional information that is not contained in the abstract of an article.

Further research deemed the direct use of citation sentences for summarization as inaccurate as they may not provide enough information and sometimes may even be imprecise. So, [3, 9] proposed identifying the relevant context (usually a set of sentences) in the reference paper (original paper) for each citation and using that for summarization. Such a set of sentences is usually called a 'text

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:00:35 UTC from IEEE Xplore. Restrictions apply.

span,' and in this case, it is known as cited-text span. The summaries formed using these cited-text spans would include the insights of the research community while being composed of discourse from a stable source (only a single paper) and, thus, be more comprehensible and coherent. [10] performed experiments that show that summaries formed by using cited text spans outperformed those formed by using the citation sentences directly. CL-SciSumm shared task [5] also uses the cited text-spans based approach.

III. PRELIMINARIES

Skip thoughts, [11] is a model consisting of an encoder based on a recurrent neural network that is used to encode input sentences to a given size output feature vector. This output is used as an input to two decoders. One of the two generates the previous sentence from the context of this vector, and the other one generates the next sentence of the document given the context of the input vector. Outputs from the two decoders are used to train the encoder. Skip thoughts network is trained on the dataset of reference papers whose summaries need to be generated to convert a given sentence of the paper into a fixed-size feature vector.

In classic game theory, we define a set of players, player ={1,2…..n}and every player has a corresponding set of strategies, $strategy_i$ = {1,2…..m}. The payoff is the reward a player earns by playing a strategy in a particular game setting. Utility function relates strategies chosen by all players to the payoff, mathematically defined as, $U=strategy_i \times \ldots \times strategy_j$. To play the game, each player selects a strategy, and the utility function depends on the combination of strategies played simultaneously by all the players involved rather than one player. A crucial assumption being that all the players are rational, and their essential goal is to boost the estimation of U. In non-cooperative games, players first consider what the other people can play, and then choose an ideal strategy.

Nash equilibria can be defined as the stable state of a system where changing the strategy of only one player while keeping the strategies of other players the same gives incentives to other players. In other words, it is the state of a system when each player gets the highest payoff at the given strategy spaces of the complete game. Furthermore, most games involve mixed strategies for players, which are probability distributions over pure strategies. Consider a set of pure strategies, $strategy_i$ = {1,2….m}. In a mixed strategy setting, each pure strategy is assigned a probability, $X = \{x_1, x_2 \ldots .. x_m\}$ where each component x_i is the probability of a player selecting its ith strategy.

In a game between two players, the payoff of the first player from a strategy i whose probability in mixed strategy setting is p concerning the strategy j of the other player whose probability in mixed strategy setting is q will be $u_{ij} = p.A_1.q$ where A_1 is the payoff matrix of the first player.

Evolutionary game hypothesis veers off from the old-style idea as in a player can play the game with all the co-players winning a halfway result after each game and afterward acquires the complete total of every single fractional result as the last result. In this hypothesis, the players over and over take part in games and update their methodology space at every emphasis, considering the adjustments granted during past games. The payoff of a pure strategy h for player i in this setting is given as,

$$u_i\left(e_i^h\right)=\Sigma_{j=1}^{n}\left(A_{ij}x_j\right)_h \quad (1)$$

where x_k denotes the mixed strategy space of player k and A_{ij} is the payoff matrix between player i and j. The average payoff for this player is evaluated as,

$$u_i(x)=\Sigma_{j=1}^{n} x_i^T A_{ij} x_j \quad (2)$$

The probabilities in the mixed strategy space of each player are refreshed by the replicator dynamic condition [12] to accomplish Nash equilibria.

$$\dot{x}=\left[u\left(e^h,x\right)-u(x,x)\right].x^h \ \forall h \in x \quad (3)$$

We have utilized the discrete rendition of this condition, in our work [13] in our work. At each time step, the strategy space is updated following this equation until the system attains Nash equilibrium.

$$x^h(t+1)=x^h\frac{u\left(e^h,x\right)}{u(x,x)} \quad (4)$$

IV. PROPOSED WORK

We took the concept of evolutionary game theory and applied it to the task of text summarization. We take abstract sentences and the citation sentences of the reference paper as the players of the game. Each player plays a game against all the other players, similar to an evolutionary game setting. Weightage of a game in updating the strategy space of a player depends on the similarity between the two players (or sentences). We build a weight matrix W as,

$$w_{ij}=0.7*\text{similarity}\left(\text{player}_i, \text{player}_j\right)+0.3*\text{auth_score}_{(i,j)} \quad (5)$$

where $player_i$ & $player_j$ denote players i & j respectively and $auth_score_{(i,j)}$ is relative authority score of the paper which contains $player_i$ to the authority score of the paper which contains $player_j$ and similarity is cosine similarity between vectors of the respective players. The relative weightage of these two scores for which this method works best has been calculated through experimentation. This set of weights gives the best results. We then convolve the W matrix with a Gaussian filter to smooth it. The set of strategies of each player is a set of sentences similar to the player sentence in the reference paper. The size of the strategy space of each player is kept the same. We initialize the probabilities equal in the mixed strategy setting. We define the payoff matrix A_{ij}

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:00:35 UTC from IEEE Xplore. Restrictions apply.

between players i and j where similarity is the cosine distance between the vectors of $strategy_{ik}$ and $strategy_{jl}$ as,

$$a_{ij}^{kl} = similarity\left(strategy_{ik}, strategy_{jl}\right) \quad (5)$$

We follow the aforementioned discrete dynamic equation (4) for updating probability to attain Nash equilibria in the above system where the payoff of the h^{th} pure strategy of $player_i$ is defined as,

$$u_i\left(e_i^h\right) = \Sigma_{j=1}^{n}\left(w_{ij} A_{ij} x_j\right)_h \quad (6)$$

and the average payoff of $player_i$ is defined as,

$$u_i(x) = \Sigma_{j=1}^{n} x_i^T (w_{ij} A_{ij} x_j) \quad (7)$$

We can terminate updating the probabilities once Nash equilibrium is attained that is the state of the game becomes stable or after a fixed number of iterations.

Our method finds the most suitable representation of abstract and citation sentences in the paper and scores them in order of their relevance in a summary.

For each player, the best strategy that is the one having maximum probability in the player's mixed strategy is chosen as the best representative of that player. Then all these representative strategies are compared together and sorted in the decreasing order of the average payoff of their respective player. The sentences are chosen to be a part of the summary in the priority of their sorted order and taking care that similar sentences are not being added to the summary. The sentences chosen for the summary by the above method are then arranged in the order in which they appear in the reference paper, and then the summary can be reduced to 200 words by omitting last sentences as later parts of a paper are very specific for a summary compared to sentences at the beginning of the paper.

The intuition behind this being that for a player, the strategy which is coherent with all other strategies of all other players, will have a better pure strategy payoff than the rest of the other strategies of that player, which will lead to a gain in its probability, increasing its probability from the same initial value in the mixed strategy of that player. Thus, the strategy with maximum probability can be taken as the representative of that player. It is the most suitable sentence in the reference paper, which is similar to the player and also coherent with the other sentences of the summary. The importance of the representative strategy, in summary, will be indicated by the average payoff of the respective player as more will be the average payoff of a player more relevant will be its strategy to the summary.

V. Experimentation and Results

We have used human-annotated, Scisumm dataset [2], ScisummNet which constitutes of 1000 papers in the domain of ACL anthology network along with their human expert generated summaries to test our model.

We trained the skip thoughts network with the reference papers of the dataset and used the cosine distance between skip thought vectors as a similarity measure to create strategy space for each player. For building the weight matrix and payoff matrix, we used a weighted score of cosine distance between skip thoughts vectors and tf-idf vectors of the sentences as an altered cosine similarity measure. We used a weight of 0.8 for skip thought vectors similarity and 0.2 for tf-idf vectors similarity, the values of these weights were determined experimentally. We then convolved the weight matrix with a Gaussian filter to smooth it. We followed the system dynamics for the above-mentioned evolutionary game theory model to generate summaries.

An overall framework of our model is shown in Fig. 1. Abstract and citation sentences of the reference paper are players of the game. Tf-idf and skip-thoughts vector encodings are used to create the strategy space of each player and for building payoff and weight matrices. Evolutionary game dynamics are followed to get a suitable representative of every player and to rank those representatives in order of their importance in the summary. Dissimilar sentences are chosen in the order they have been ranked in. This set of sentences is arranged in the order in which they appear in paper to generate the summary.

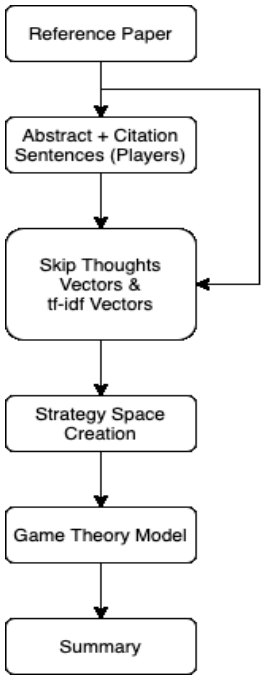


Fig. 1. Proposed Framework.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:00:35 UTC from IEEE Xplore. Restrictions apply.

In Fig. 2, we present an abstract of a reference paper from the above dataset on the topic "Unsupervised Semantic parsing". The summary generated by our method is shown in Fig. 4.

We took two players from the same reference paper whose abstract is cited in Fig. 3, one being a relevant sentence("We evaluate our approach by using it to extract a knowledge base from biomedical abstracts and answer questions.") to the summary and the other being a sentence("The sentence can be rephrased as "Utah is next to Idaho," "Utah shares a border with Idaho," etc.") that is too specific for a summary and doesn't add much information to it and show the contrast in the curve of average payoff between a coherent, informative player that is integral to a summary and a player that is not important concerning a summary. The obtained graph (Fig. 3) is in concurrence with our intuition.

"We present the first unsupervised approach to the problem of learning a semantic parser, using Markov logic. Our USP system transforms dependency trees into quasi-logical forms, recursively induces lambda forms from these, and clusters them to abstract away syntactic variations of the same meaning. The MAP semantic parse of a sentence is obtained by recursively assigning its parts to lambda-form clusters and composing them. We evaluate our approach by using it to extract a knowledge base from biomedical abstracts and answer questions. USP substantially outperforms Text Runner, DIRT and an informed baseline on both precision and recall on this task."

Fig. 2. Abstract of reference paper on the topic "Unsupervised Semantic parsing".

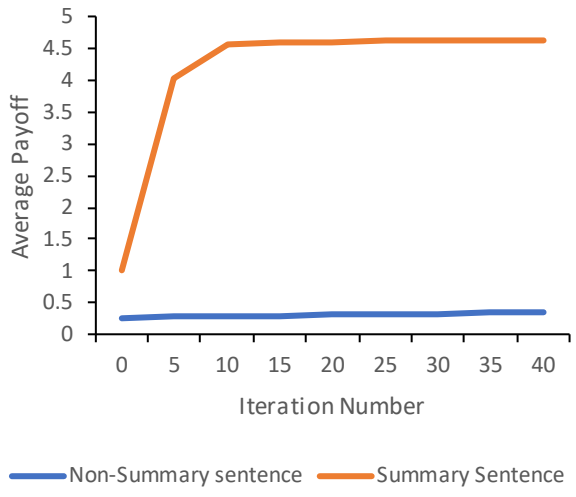


Fig. 3. Average payoff vs. iteration number graph between a summary and a non-summary sentence.

"Our USP system transforms dependency trees into quasi-logical forms, recursively induces lambda forms from these, and clusters them to abstract away syntactic variations of the same meaning. We evaluate our approach by using it to extract a knowledge base from biomedical abstracts and answer questions. In this paper we develop the first unsupervised approach to semantic parsing, using Markov logic. The semantic parse of a sentence is derived by starting with logical forms in the lexical entries and recursively composing the meaning of larger fragments from their parts. Markov logic makes it possible to compactly specify probability distributions over complex relational domains, and has been successfully applied to unsupervised coreference resolution and other tasks.To address this problem, we impose local normalization constraints on specific groups of formulas that are mutually exclusive and exhaustive, i.e., in each group, we require that Eki=1 ewi = 1, where wi are the weights of formulas in the group.6Regularizations, e.g., Gaussian priors on weights, alleviate this problem by penalizing large weights, but it remains true that weights within a short range are roughly equivalent. and update agenda and candidate operations until agenda is empty return the MLN with learned weights and the semantic parses Another major challenge in USP learning is the summation in the likelihood, which is over all possible semantic parses for a given dependency tree."

Fig. 4. Summary generated by our method.

We have evaluated the results, both qualitatively and quantitatively.

A. Quantitative Measure

We have used ROGUE metrics (Recall-Oriented Understudy for Gisting Evaluation) [14] for quantitative evaluation. ROGUE-2 estimates recall and F1 Measure for the bigram overlap between gold and generated summary. We gather from Table 1 that our approach performs as good as the GCN model of the ScisummNet [2] generated summaries, which is the benchmark for this dataset.

TABLE I. ROGUE-2 SCORES

Summarizer	Recall	F1 Measure
Abstract	28.52	29.40
Our Approach	34.61	32.73

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:00:35 UTC from IEEE Xplore. Restrictions apply.

Baseline	33.88	31.54

B. *Qualitative Measure*

We gathered a group of 50 people and asked them to score summaries generated by our method on a scale of 0-10 on the parameters of coherence and covering the paper's citation information. Our summaries got an average score of 8.2, which symbolizes that the method generates a comprehensible summary that represents citation information adequately.

VI. CONCLUSION

Citation information of a paper is an outlook of its influence on the researcher's community; thus, appending citation information to a paper's summary is crucial to a faster and comprehensive reviewing of the paper and deciding its relevance in one's work. Our method proposes an unsupervised game-theoretic approach to the task of citation-based text summarization, which levitates the need of a large human-annotated dataset. Our experiments show that our method generates better and comprehensive summaries than the abstract of the paper and the previous citation-based text summarization work. Additional features can be added to bias the weight and payoff matrices. Generic sentence embedding approaches with the semantic representation of the input will lead to better results. We hope that our game-theoretic approach to this task contributes to more novel work in this domain.

REFERENCES

[1] S. Teufel and M. Moens, "Summarizing Scientific Articles: Experiments with Relevance and Rhetorical Status," Computational Linguistics, vol. 28, no. 4, pp. 409–445, 2002.

[2] V. Qazvinian and D. Radev, "Scientific Paper Summarization Using Citation Summary Networks", in Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008), pp. 689–696, ACL, 2008

[3] A. Cohan and N. Goharian, "Scientific document summarization via citation contextualization and scientific discourse", International Journal on Digital Libraries, vol. 19, no. 2-3, pp. 287–303, Sep. 2017.

[4] M. Yasunaga, J. Kasai, R. Zhang, A. R. Fabbri, I. Li, D. Friedman, and D. R. Radev, "ScisummNet: A Large Annotated Corpus and Content-Impact Models for Scientific Paper Summarization with Citation Networks," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 7386–7393, 2019.

[5] K. Jaidka, M.K. Chandrasekaran, S. Rustagi, M. Kan, "Overview of the CL-SciSumm 2016 Shared Task", in Proceedings of the Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL), pp. 93-102, 2016.

[6] H. P. Luhn, "The Automatic Creation of Literature Abstracts," IBM Journal of Research and Development, vol. 2, no. 2, pp. 159–165, 1958.

[7] C. Paice and P. Jones, "The identification of important concepts in highly structured technical papers", in Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '93), pp. 69-78, ACM, 1993.

[8] A. Elkiss, S. Shen, A. Fader, G. Erkan, D. States and D. Radev, "Blind men and elephants: What do citation summaries tell us about a research article?", Journal of the American Society for Information Science and Technology, vol. 59, no. 1, pp. 51-62, 2008.

[9] A. Cohan and N. Goharian, "Contextualizing Citations for Scientific Summarization using Word Embeddings and Domain Knowledge", in Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17), pp. 1133-1136, ACM, 2017.

[10] Q. Mei and C. Zhai, "Generating impact-based summaries for scientific literature", in ACL-08: HLT - 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, pp. 816-824, ACM, 2008.

[11] R. Kiros, Y. Zhu, R. Salakhudtinov, R. Zemel, A. Torralba, R. Urtasun and S. Fidler, "Skip-thought vectors", in Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'15), pp. 3294-3302, MIT Press, 2015.

[12] P. D. Taylor and L. B. Jonker, "Evolutionary stable strategies and game dynamics," Mathematical Biosciences, vol. 40, no. 1-2, pp. 145–156, 1978.

[13] A. Erdem and M. Pelillo, "Graph Transduction as a Non-cooperative Game," Graph-Based Representations in Pattern Recognition Lecture Notes in Computer Science, pp. 195–204, 2011.

[14] C. Lin, "ROUGE: A Package for Automatic Evaluation of Summaries", in Text Summarization Branches Out: Proceedings of the ACL-04 Workshop, pp. 74-81, ACL,2004.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:00:35 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Comparative Analysis of PWM Techniques and SVPWM operated V/f control of Induction Motor having different power ratings

Divyam
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
fastrack2030@gmail.com

Ayushi Saxena
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
ayushi.saxena1407@gmail.com

Bharat Singh
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
bharatsinghuae@gmail.com

J. N. Rai
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
jnrai@dce.ac.in

Corresponding Author: Bharat Singh, bharatsinghuae@gmail.com

***Abstract*— The paper makes a comparative study among the different Pulse Width Modulation (PWM) techniques including – Space Vector Pulse Width Modulation (SVPWM), Sinusoidal PWM (SPWM) and Multiple PWM (MPWM). These techniques are defined, examined and compared to generate gate pulses for a three phase Voltage Source Inverter (VSI). The modulation index is varied for achieving the comparative analysis of these techniques. Pulse Width Modulation techniques are preferred because of low harmonic content obtained in the output voltage and current, fixed switching frequency and higher utilization of the DC bus. Also, SVPWM is then used for achieving V/f control of squirrel cage Induction Motor having different power ratings.**

Keywords—pulse width modulation, space vector pulse width modulation, sinusoidal pulse width modulation, multiple pulse width modulation, induction machine, V/f control

Acronyms

- ***PWM (Pulse Width Modulation)***
- ***SPWM (Sinusoidal Pulse Width Modulation)***
- ***SVPWM (Space Vector Pulse Width Modulation)***
- ***MPWM (Multiple Pulse Width Modulation)***
- ***VSI (Voltage Source Inverter)***
- ***IM (Induction Machine)***

I. INTRODUCTION

An inverter is a power electronic converter used for obtaining ac power from dc power at the required output voltage and given frequency. They can be mainly categorized as Current Source Inverters (CSI) and Voltage Source Inverters (VSI). Voltage Source Inverters have a stiff dc voltage source at the input whereas Current Source Inverters have a stiff current source at the input. The output voltage waveform of the inverter can be sinusoidal, quasi-square wave or square wave with some distortions and can be controlled with the help of switches. The harmonic distortions in the output voltage and current can be considerably reduced by using proper switching schemes. Pulse Width Modulation is one such technique which can be implemented to achieve an efficient switching scheme.

Pulse Width Modulation is a method which uses a pulse wave whose pulse width is modulated to achieve a variation of the mean value of the output waveform. It is a way to control the switching sequence and time period for the power electronic switches used in an inverter. If the switch is ON for longer durations in a cycle, it will result in higher average output of the inverter. [1] The proportion of the ON duration of the switch to the total time period is known as the duty cycle. Also, the switching power loss is very low as negligible current flows in the OFF state of the device and the voltage drop is also negligible. Fig. 1 represents the block diagram for the PWM technique.

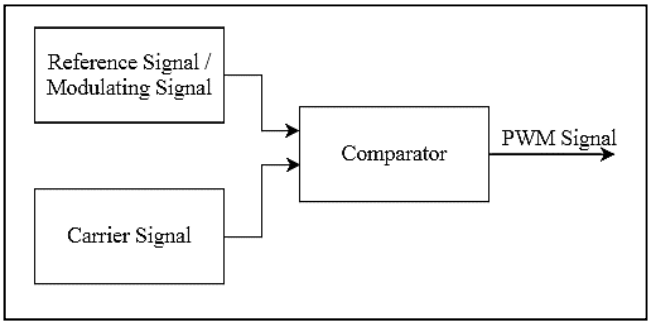


Fig. 1. Block Diagram depicting PWM Technique

The distortion in the output waveform is due to the presence of harmonics. Harmonics are the frequency component of the voltage present at integral multiples of the frequency of the main signal. This can be measured by the Total Harmonic Distortion (THD):

$$THD = \frac{\sqrt{(\sum_{n=2}^{\infty} V_{n_rms}^2)}}{V_{fund_freq}} \quad (1)$$

The THD is affected by the Modulation Index (M_A), which can be defined as the ratio of the peak value of the reference signal to the maximum value of the carrier signal.

IM are the most commonly used machines, especially as motors, because of excellent running conditions and low maintenance. Induction motors aggregate to 70-75% of total electrical load. Though the speed control techniques of IM are not much effective, V/f control is considered as the most popular method. In this paper, V/f control of Induction Motor having different power ratings is achieved by using SVPWM.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:53:43 UTC from IEEE Xplore. Restrictions apply.

II. TYPES OF PWM TECHNIQUES

A. Multiple PWM

In this method, multiple pulses are given to the switches in each half cycle. The reference signal is measured against a carrier wave in order to obtain the gate pulses of equal width for the switches of the VSI. It is also known as Uniform Pulse Width Modulation.

B. Sinusoidal PWM

In this method, the sinusoidal reference or modulating wave is measured against a large frequency carrier wave in order to generate gate pulses to be fed to the VSI.

C. Space Vector PWM

This is a technique which uses an algorithm for modifying the pulse width as per the given reference vector. In this method, vectors are used to represent all possible switching in a space having two dimensional voltage. This is obtained by the transformation to three phase independent vectors from the three phase dependent vectors. The assumption includes a constant reference vector for the corresponding sampling time. The sampling frequency is dependent on the ON and OFF times of the device used, which should be ideally infinite. Thus, for minimum total harmonic distortion and minimum switching losses, the maximum possible sampling frequency is selected. [3]

The line side frequency and magnitude of the fundamental component are regulated using the magnitude and frequency of the reference voltage vector. [1] In SVPWM, there is a unique correspondence between a three-phase system and a space vector in complex plane, which is represented in a mathematical manner.

III. METHODOLOGY

A. Implementation of SVPWM

SVPWM Technique involves transformation of three phase sinusoidal signal into revolving voltage vector in complex plane (abc to dq axis transformation) with constant magnitude revolving with constant angular frequency.

The six switches in a three phase inverter correspond to 8 switching states (all combinations of ON and OFF of the six switches) which are called voltage vectors in the three phase VSI structure. The switching sequences of the six switches should abide by these two restrictions - out of the six switches, three must always be in the ON state and the remaining switches must be in the OFF state. Also, both the switches of the same leg of the inverter should always be driven by two complementary pulse signals.

Fig. 2 shows the vectors in complex plane in different sectors. There is a 60° angle amidst any two neighboring non-zero vectors. In the meantime, zero voltage to the load is applied by V0 and V7 (two zero vectors) at the origin as there is no transfer of power in these switching states due to short circuit. The eight vectors- V0 to V7 are called basic space vectors. The required voltage reference vector V_{ref} can be obtained by applying the same transformation to desired output voltage in the dq plane. [5]

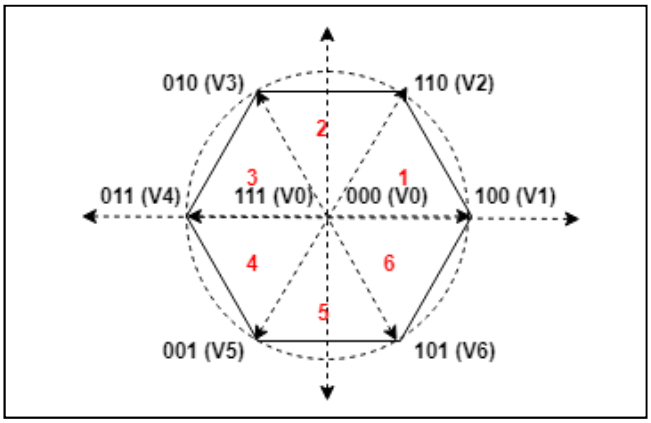


Fig. 2. Possible switching sequence for the VSI

Main function of SVPWM is to estimate voltage reference (V_{ref}) vector with the help of inverters eight switching pattern and this approximation is done by generating output voltage of inverter in small sampling period which is equivalent to V_{ref} in the same period. This V_{ref} voltage vector has discrete rotation in complex plane between sectors portioned at 60° hence forms hexagon trajectory. Decomposition of this vector on real and imaginary axis indicates that it coincides with the generating switching vector. Switching time at any sector by using V_{ref} as a combination of adjacent vectors can be determined by using Fig. 4 as:

$$T_1 = \frac{\sqrt{3}*T*|\overrightarrow{V_{ref}}|}{V_{dc}} sin(\frac{n\pi}{3} - \alpha), \quad (2)$$

$$T_2 = \frac{\sqrt{3}*T*|\overrightarrow{V_{ref}}|}{V_{dc}} sin(\alpha - \frac{(n-1)\pi}{3}), \quad (3)$$

$$T_0 = T - T_1 - T_2. \quad (4)$$

where T is the time period and n denotes number of sector ranging from 1 to 6. Also, α should lie between 0° and 60°. [5]

The switching time for the different sectors using (2), (3) and (4) for the VSI can be summarized as in Table I.

TABLE I. SWITCHING TIMES FOR VSI

Sector	Upper Switches (S_1, S_3, S_5)	Lower Switches (S_4, S_6, S_2)
1	$S_1=T_1+T_2+T_0/2$ $S_3=T_2+T_0/2$ $S_5=T_0/2$	$S_4= T_0/2$ $S_6= T_1+T_0/2$ $S_2= T_1+T_2+T_0/2$
2	$S_1=T_1+T_0/2$ $S_3= T_1+T_2+T_0/2$ $S_5= T_0/2$	$S_4= T_2+T_0/2$ $S_6= T_0/2$ $S_2= T_1+T_2+T_0/2$
3	$S_1= T_0/2$ $S_3= T_1+T_2+T_0/2$ $S_5= T_2+T_0/2$	$S_4= T_1+T_2+T_0/2$ $S_6= T_0/2$ $S_2= T_1+T_0/2$
4	$S_1= T_0/2$ $S_3= T_1+T_0/2$ $S_5= T_1+T_2+T_0/2$	$S_4= T_1+T_2+T_0/2$ $S_6= T_2+T_0/2$ $S_2= T_0/2$
5	$S_1= T_2+T_0/2$ $S_3= T_0/2$ $S_5= T_1+T_2+T_0/2$	$S_4= T_1+T_0/2$ $S_6= T_1+T_2+T_0/2$ $S_2= T_0/2$
6	$S_1= T_1+T_2+T_0/2$ $S_3= T_0/2$ $S_5= T_1+T_0/2$	$S_4= T_0/2$ $S_6= T_1+T_2+T_0/2$ $S_2= T_2+T_0/2$

The switching times as per Table I have been used in the MATLAB Simulink simulation for the Space Vector Pulse Width Modulation for VSI. Fig. 3, Fig. 4, Fig. 5 and Fig. 6 represent the implementation of this technique using MATLAB Simulink.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:53:43 UTC from IEEE Xplore. Restrictions apply.

B. *Implementation of Sinusoidal PWM*

In SPWM inverter the pole-voltage pulse widths for a time period varies in a sinusoidal manner. The sinusoidal pulse width modulation technique revolves around comparison of a triangular carrier signal having high frequency with a sinusoidal modulating signal which gives the fundamental frequency component of the output current and voltage waveform. The maximum value of the modulating signal should not be more than the maximum value of the carrier signal for linear operation. The lower and upper switches of the VSI are controlled using this comparator output. Fig. 7, Fig. 8, Fig. 9 and Fig. 10 represent the implementation of this technique using MATLAB Simulink.

C. *Implementation of Multiple PWM*

In MPWM inverter the pole-voltage pulse widths for a time period varies in a square wave manner. The multiple pulse width modulation technique revolves around comparison of a triangular carrier signal having high frequency with a square modulating signal which gives the fundamental frequency component of the output current and voltage waveform. The maximum value of the modulating signal should not be more than the maximum value of the carrier signal for linear operation. The lower and upper switches of the VSI are controlled using this comparator output. Fig. 11, Fig. 12, Fig. 13 and Fig. 14 represent the implementation of this technique using MATLAB Simulink.

D. *Implementation of V/f control of IM using SVPWM*

For the speed control of IM, though there is no best or effective way to control the speeds, the V/f control is still the most popular method. Since,

$$V = \sqrt{2}\pi f \emptyset_m N \quad (5)$$

$$B_m \propto V/f \quad (6)$$

We operate at the knee-point of the B-H curve, we do not want to vary the maximum magnetic field density (B_m). Hence, voltage and frequency are varied in such a manner that the V/f ratio is always constant. Since the maximum torque (τ_m)

$$\tau_m \propto (V/f)^2 \quad (7)$$

Hence, in V/f control, τ_m remains same and thus the power increases linearly with the rotor speed. Here, the V/f ratio is taken to be 8. This has been plotted for various induction motor drives having different power ratings using SVPWM and unity modulation factor. [9] Fig. 15 represents the implementation of this technique using MATLAB Simulink.

IV. SIMULATION RESULTS

The paper discusses about the software implementation of various PWM techniques namely SVPWM, SPWM and MPWM through MATLAB Simulink. The ultimate objective of either of these techniques is to obtain alternating output voltage with minimum harmonics and maximum fundamental component. By analyzing and comparing their THD Analysis with varying Modulation Indices, the observations can be summarized as follows:

A. *SVPWM*

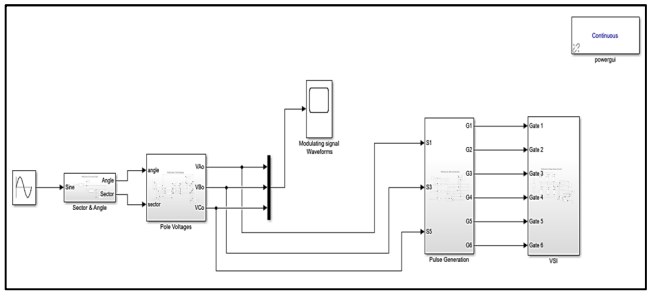


Fig. 3. MATLAB Simulink model for SVPWM with parallel RL load of 50Ω and 5 mH

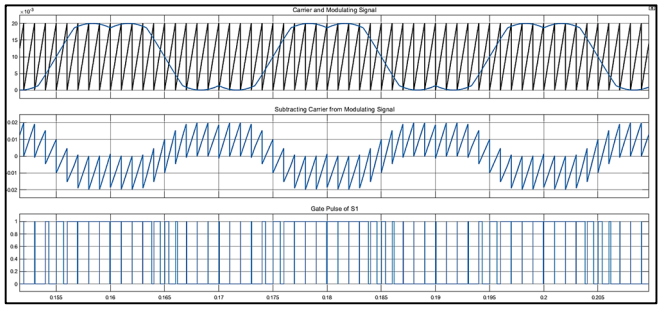


Fig. 4. Waveforms for carrier signal, modulating signal, comparator output and gate pulse for switch S1 at $M_A = 1$

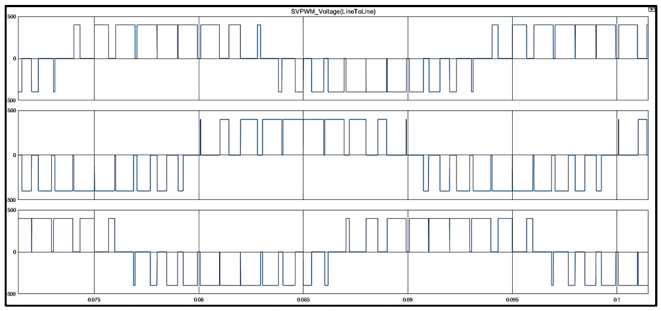


Fig. 5. Waveforms for output line voltages for SVPWM

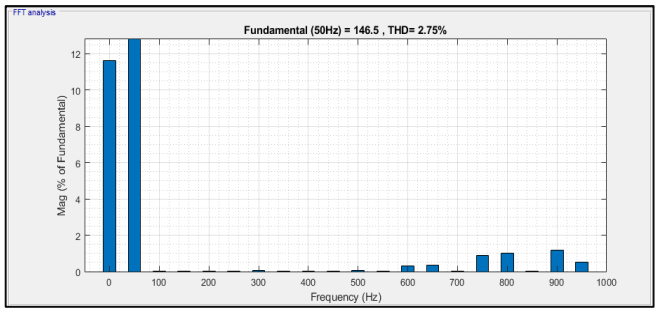


Fig. 6. THD for output current of VSI for SVPWM

B. *SPWM*

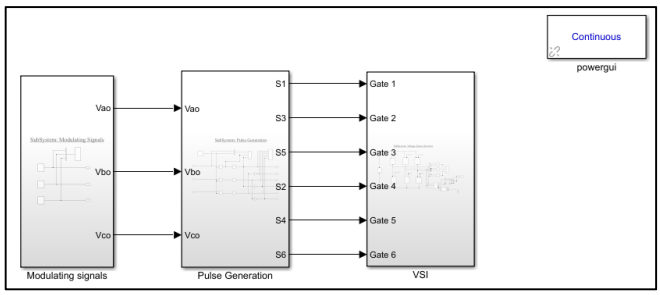


Fig. 7. MATLAB Simulation Simulink model for SPWM with parallel RL load of 50Ω and 5 mH

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:53:43 UTC from IEEE Xplore. Restrictions apply.

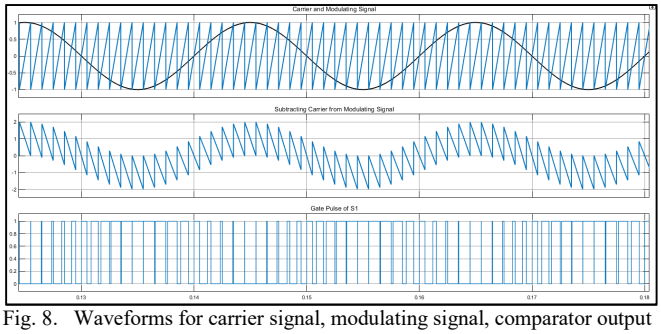


Fig. 8. Waveforms for carrier signal, modulating signal, comparator output and gate pulse for switch S1 at $M_A = 1$

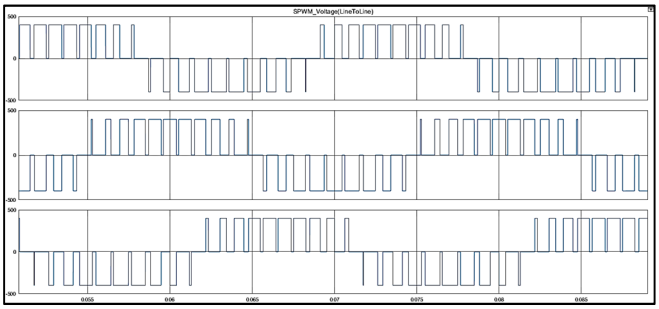


Fig. 9. Waveforms for output line voltages for SPWM

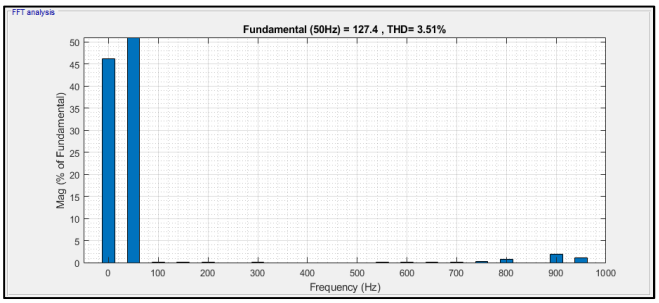


Fig. 10. THD for output current of VSI for SPWM

C. MPWM

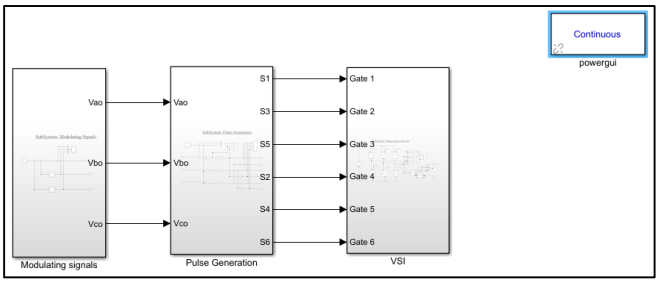


Fig. 11. MATLAB Simulink model for MPWM with parallel RL load of 50Ω and 5 mH

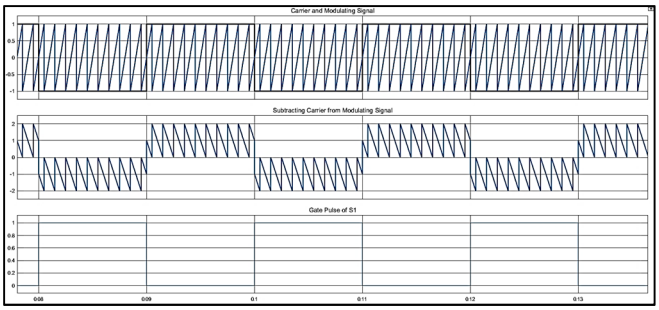


Fig. 12. Waveforms for carrier signal, modulating signal, comparator output and gate pulse for switch S1 at $M_A = 1$

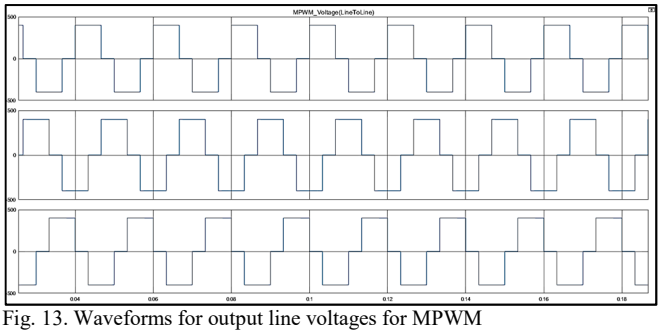


Fig. 13. Waveforms for output line voltages for MPWM

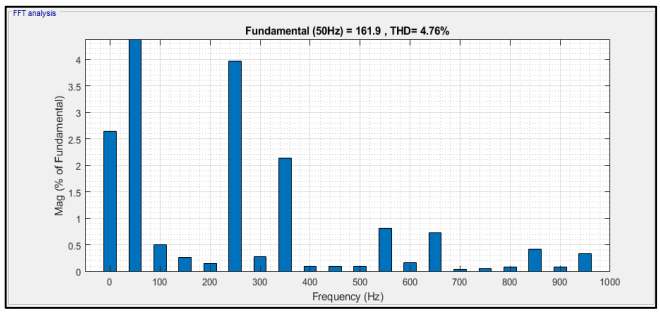


Fig. 14. THD for output current of VSI for MPWM

D. V/f control of IM using SVPWM

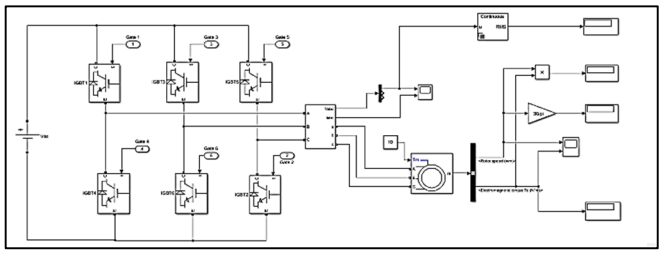


Fig. 15. MATLAB Simulink Subsystem for V/f control of IM using SVPWM

V. OBSERVATIONS

The three PWM techniques namely- MPWM, SPWM and SVPWM were compared for different modulation indices (M_A) ranging from 0 to 6 and the ratio of corresponding value of peak fundamental output line voltage (V_{01}) to constant dc supply voltage (V_{dc}=400V) was noted for all the three techniques. The observations have been summarized in Table II.

TABLE II. COMPARISON OF MPWM, SPWM AND SVPWM

MA	V_{01}/V_{dc}		
	MPWM	SPWM	SVPWM
0	0	0	0
0.4	0.44	0.3475	0.4005
0.6	0.6565	0.50725	0.598
0.8	0.8645	0.67025	0.79575
1.0	1.09575	0.848	1.0005
1.1	1.10625	0.9245	1.048
1.2	1.10625	0.95475	1.05775
1.3	1.10625	0.9645	1.058
1.4	1.10625	1.009	1.0795
1.5	1.10625	1.01025	1.08
1.6	1.10625	1.0235	1.08625
1.8	1.10625	1.04475	1.09225
2.0	1.10625	1.0525	1.092
2.2	1.10625	1.07225	1.094
2.4	1.10625	1.08325	1.10275
2.6	1.10625	1.0725	1.11175
2.8	1.10625	1.1	1.117
3.0	1.10625	1.0805	1.118

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:53:43 UTC from IEEE Xplore. Restrictions apply.

3.2	1.10625	1.094	1.115
3.5	1.10625	1.0805	1.11425
4.0	1.10625	1.10225	1.1115
4.5	1.10625	1.07475	1.11325
5	1.10625	1.08	1.1105
5.5	1.10625	1.08325	1.113
6	1.10625	1.077	1.11025

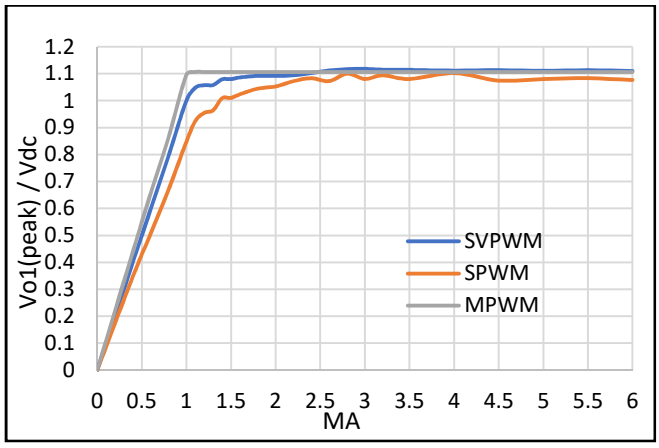


Fig. 16. Plot for V_{01}/V_{dc} Vs M_A for MPWM, SPWM and SVPWM

From Fig. 16 it is observed that the fundamental output voltage for SVPWM is 17.98% higher than that for SPWM at unity modulation index. Also, for a range of modulation index from 0.4 to 1.6, the mean fundamental output voltage for SVPWM is 12.37% higher than that for SPWM.

It is also observed from Fig. 6, Fig. 10 and Fig. 14 that the THD in output current is least for SVPWM (THD=2.75%), followed by that for SPWM (THD=3.51%) and the highest for MPWM (THD=4.76%).

Also, the peak line to line output VSI voltage is highest for MPWM as the pulse widths are almost same but the current produced by MPWM has highest distortion and thus is not preferred for practical use.

SVPWM technique, at unity modulation index and constant load torque of 10 Nm, was then used for achieving the V/f control of IM of different power ratings, the simulation results of which have been summarized in Table III.

TABLE III. COMPARISON OF V/F CONTROL OF DIFFERENT POWER RATING OF IM USING SVPWM

IM WITH POWER RATING OF 4 KW		
SPEED (RPM)	TORQUE (NM)	POWER (W)
1452	10.2	1561
1263	10.6	1409
1075	10.11	1139
886.8	9.95	923.1
698.9	9.94	727.3
513.2	9.83	527.9
337	10.8	387.2
77.55	10.19	82.74
IM WITH POWER RATING OF 7.5 KW		
SPEED (RPM)	**TORQUE (NM)**	**POWER (W)**
1477	9.69	1500
1289	10.08	1361
1101	9.71	1119
913	9.69	927
724.1	9.63	729.7
505.1	11.89	628.7
419.3	19.07	837.4
125.2	4.58	60.1

IM WITH POWER RATING OF 15 KW		
SPEED (RPM)	TORQUE (NM)	POWER (W)
1494	9.95	1557
1305	11.17	1527
1118	9.97	1167
929.6	10.21	994
740.6	12.31	954.7
530	53	3300

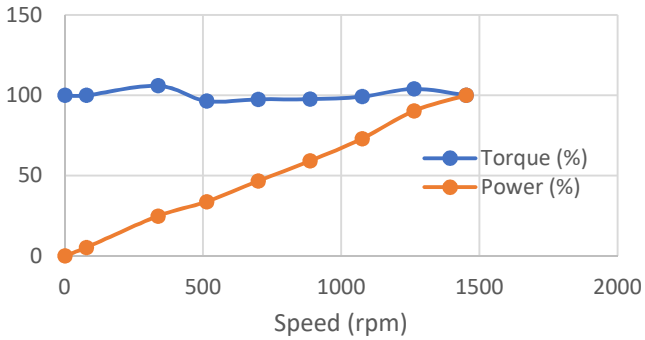


Fig. 17. 4 kW IM characteristics for V/f control

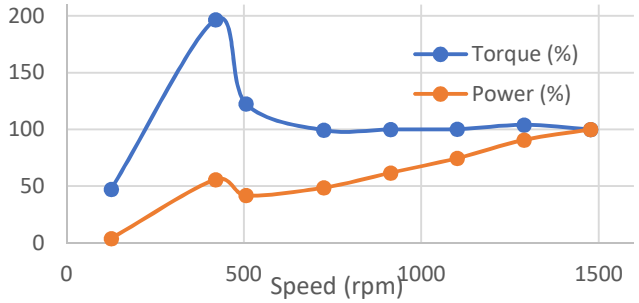


Fig. 18. 7.5 kW IM characteristics for V/f control

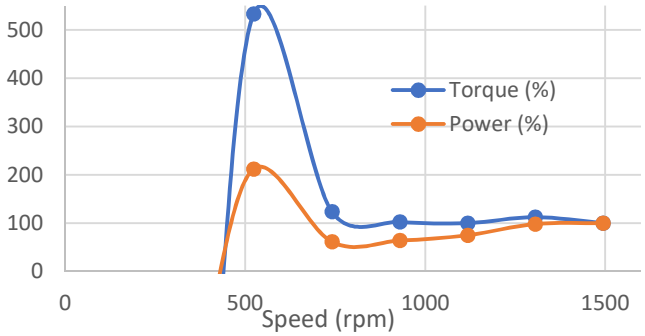


Fig. 19. 15 kW IM characteristics for V/f control

From Fig. 16, we observe that the torque remains almost constant throughout the V/f operation and the power varied linearly with the rotor speed.

From Fig. 17, Fig. 18 and Fig. 19, we observe that in the low frequency region, since the inductance becomes very low, the circuit becomes resistive in nature and hence the impedance does not vary with frequency, so with decreasing voltage and frequency, the current in the machine decreases and hence torque decreases.[12] Therefore, in low frequency region (< 37.5% of the rated frequency), the performance of IM deteriorates. Also, as the machine power rating increases, there is a decrease in the impedance of the machine and hence it becomes constant with respect to frequency even in the relatively higher frequency range (in the low frequency region) and accordingly we observe a greater dip in the torque with the increase in the power rating of the IM.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:53:43 UTC from IEEE Xplore. Restrictions apply.

VI. CONCLUSION

Here, SVPWM Technique was used to reduce harmonics in output current of the VSI. Also, it was compared with other PWM techniques, namely – SPWM and MPWM, it can be concluded that SVPWM provides the least THD in the output line current when compared to SPWM and MPWM and MPWM provides the highest THD. Also, the maximum line output VSI voltage is higher for SVPWM as compared to SPWM. Although the maximum line output VSI voltage is high for MPWM, it is not preferred due to presence of high harmonic distortions in current. For MPWM, V_{01} saturates for M_A greater than unity, i.e. over modulation region. However, in case of SPWM and SVPWM, V_{01} can be increased even in the over modulation region, though non-linearly, and does not saturate as in MPWM. Thus, based on above conclusions, SVPWM is a more sophisticated and effective technique for inverter operation and hence is used for the V/f control.

We also conclude that V/f control should not be preferred in the lower frequency range and the performance of IM deteriorates in this region.

Further, in future, the concept of Random PWM (RPWM) implementing random switching frequency method can be employed along with the above implementation of the pulse width modulation techniques for achieving a harmonic content spread over the entire harmonic spectrum and not particularly on one harmonic frequency.

ACKNOWLEDGMENT

We would like to thank Prof J.N. Rai, for his constant motivation and guidance during the project. Without his timely help and support, the completion of this paper wouldn't have been possible. We would also like to thank the publisher of the research papers from where we have taken reference.

REFERENCES

1. Avinash Mishra, Swaraj Save, Rohit Sen ," Space Vector Pulse Width Modulation" , International Journal of Scientific and Engineering Research, Feb 2014.
2. Wan Noraishah, Mohd. Firdaus Ismail, Ahmed Farid Abidin, "Multi-Phase Inverter Space Vector Modulation", 2013 IEEE 7th International Power Engineering and Optimization Conference (PEOCO2013),Malaysia,June 2013.
3. B.Arulvani, V.Padmathilagam, "Space Vector Pulse Width Modulation Controlled Multilevel CSI", International Research Journal of Computer Science (IRJCS), May 2015.
4. Lingling Fan "Realizing Space Vector Modulation in MATLAB and PSCAD", IEEE, 2013.
5. Renu Sharma, Deepak Kumar Goyal , Harpreet Singh, "Analysis of Space Vector PWM for 3 Phase Inverter and comparison with SPWM",International Journal of Advanced Research in Electrical , Electronics and Instrumentation Engineering (IJAREEIE), Jan 2016.
6. Ayman Y. Yusuf, "Space Vector Pulse Width Modulation Technique" International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE) ISSN: 0976-1353 Volume 15 Issue 1 –MAY 2015.
7. Ahmed Muhsin Abdul Majeed, Enas Hamid Ibrahim, Nadia Qassim, "Investigation of SPWM and SVPWM VSI for Induction Motor Drive" ,The Fourth Post Graduate Engineering Conference.
8. Umashankar S, Arun Shankar V.K., Geet Jain, Mohan L Kolhe, "Comparative Evaluation of Pulse Width Modulation Techniques on Effective DC Link Voltage Utilization of Grid Connected Inverter", International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) – 2016.
9. Ranju Bharti, Madhav Kumar, B.M Prasad, "V/F Control of Three Phase Induction Motor", International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN), 2019.
10. Di Wu, CS Chang, Y Fang, H Bai, "Performance improvement of V/f induction -motor control in the low-frequency range", 8th International Conference on Advances in Power System Control, Operation and Management, 2009.
11. Bose, B. K., Modern Power Electronics and AC drive, Prentice Hall, 2001.
12. Zhe Zhang, Yiqi Liu, Ali M. Bazzi, "An Improved High-performance Open-loop V/f Control Method for Induction Machines", IEEE Applied Power Electronics Conference and Exposition (APEC), 2017.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:53:43 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Comparative Analysis of Route Planning Algorithms on Road Networks

R. K. Yadav
Department of Computer Science and Engineering,
Delhi Technological University,
New Delhi, India
rkyadav@dtu.ac.in

Giriraj Kishor
Department of Computer Science and Engineering,
Delhi Technological University,
New Delhi, India
girirajrpvv@gmail.com

Himanshu
Department of Computer Science and Engineering,
Delhi Technological University,
New Delhi, India
himanshu21498@gmail.com

Kishan Kashyap
Department of Computer Science and Engineering,
Delhi Technological University,
New Delhi, India
kishankashyap5march@gmail.com

***Abstract* — It is merely impossible to know the history of Route Planning Algorithms. It can be imagined that from the very primitive era, animals have dedicated their time to find the shortest path to reach their food. This kind of algorithms is one of the most practised ones and forms the basis for solving many real-world problems. But like most of the graph algorithms, they were proposed in the mid-twentieth century. The applications of these algorithms are many, the most important one is its usage in Routing and Navigation based applications like Google Maps, Ola and, Uber. Here, the road network is resolved as a large graph with nodes as coordinates holding some meaning and edges as ways/paths connecting adjacent coordinates, then route planning algorithms are applied for finding the shortest path between two nodes in the graph created. A noteworthy application is in Routing data packets on the internet; the Open Shortest Path First (OSPF) routing protocol is based on the Dijkstra algorithm and is used in autonomous systems like Local Area Network (LAN). It is also being used in routing especially long-distance calls. Another interesting usage can be seen in the Currency Arbitrage strategy where currencies are bought and sold instantaneously from one market to another to make a profit from the exchange rate divergences. Due to applications in so many important fields, these algorithms are studied rigorously yet to come up with a more optimized version targeting certain scenarios or in general. It is in this understanding that this paper delves into the study and implementation of various route planning algorithms and portrays their performances based on efficiency, in comparison with each other when running on a graph obtained from real road network data. This paper will also deduce a set of parameters that affects the traffic flow and the evaluation metrics that should be considered when applying those algorithms.**

Keywords — Graph Theory, Route Planning Algorithms, Shortest Path, Dijkstra, A* search, Landmark Heuristic, Contraction Hierarchy, Transit Node Routing, Fitness Function, Genetic Algorithm.

I. INTRODUCTION

One of the first papers on the shortest path algorithms was given by Dantzig in the year 1958 to sort the list of edges based on the distance from the source node and to ignore the arcs of the list involving nodes which have been reached earlier. The complexity of the proposed algorithm was $O(n^2 logn)$, where n is the number of nodes in the graph. In the following year, Dijkstra proposed an improved method where a list of size n was maintained and for each vertex, the best path found so far was stored. The complexity was $O(n^2)$ and later it was realized that an efficient data structure like heaps can even reduce the running time to $O(m + nlogn)$. In 1964, Neel Nilson invented a heuristic-based approach to speed up the Dijkstra algorithm and named it A1; then in 1968 with the help of Bertram Raphael, Peter E. Hart gave the concept of consistent heuristic and named versions of the algorithms using it as A* [1]. In the case of road-networks, the size of the graph is huge and the Dijkstra algorithm is not time-efficient for the shortest path search queries due to enlarged search space [10]. Therefore, many improvements to the shortest path algorithm were proposed in 2007-2010 by the professors of KIT, Germany including D. Schultes, P. Sanders, and others. These algorithms include ALT, Contraction Hierarchies, and Transit Node Routing to speed up Dijkstra and are exhaustively being used by many routing based applications. The research is still going on in this field.

Estimating the optimal path to reach from one location to another on roadway networks is one of the most concerning tasks in the conveyance, distribution, and logistics applications. The routing algorithms should be highly scalable to accommodate various distributions amongst road networks that highly depend on the geography and demography of a given place.

This research paper aims to address the widely known algorithms for route planning that is being used on the motorway systems. Based on the implementation results of the route planning algorithms, a comparison based on the efficiency of these algorithms applied on a graph obtained from a real instance of a road network.

Afterwards, various parameters are inferred that create an impact over deciding the cost of a path. Various Evaluation metrics that should be taken into consideration for estimating the efficacy of a routing algorithm are also mentioned.

II. LITERATURE OVERVIEW

Route planning algorithms

There are several route planning algorithms applied in road networks, these are:

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:58:23 UTC from IEEE Xplore. Restrictions apply.

1) Dijkstra algorithm: This algorithm calculates the shortest paths from a given node (the start node) to all the other nodes in the graph [6]. One of the popular implementations of this algorithm is by maintaining a priority queue where priority is the degree of closeness with the start node. Hence, it comes under the famous Single Source Shortest Pathfinding algorithms. The nodes are visited in a sequence based on proximity with the start node and the algorithm terminates when either the target node is popped from the queue (i.e. settled) or all the nodes are settled depending upon the terminating condition being used. The correctness of the algorithm fails for negative edge weights and negative edge cycles. Bellman-Ford algorithm is a dynamic programming-based approach that works for such graphs and overcomes the limitation of the Dijkstra algorithm [3].

2) A* algorithm: A* is a successor of the Dijkstra algorithm and uses a heuristic function that approximates the distance to the target node which helps in leaning towards the target with an idea of the direction to expand the nodes rather than expanding them based on the closeness with source node. To perform an efficiently heuristic function is expected to be consistent or monotone and hence admissible [4]. A heuristic function is *consistent* if, for every node *n* and its neighbor *s*, its value for *n* is never greater than the sum of step cost of getting to *s* and the heuristic value for *s*. Hence:

$$h(n) \leq h(s) + dist(n, s) \quad (1)$$

3) A*, Landmarks, and Triangle Inequality algorithm: ALT (A* Landmark Technique) is a preprocessing-based, goal-directed and more efficient technique than the Dijkstra algorithm. The heuristic used by this algorithm uses one basic principle called reverse triangle inequality which states that the length of a side of any triangle is always greater than the difference in the length of the other two sides. Using this theorem, a set of landmarks are chosen and then the heuristic value for a particular node is taken as the maximum difference in the pre-computed shortest path-length of any landmark from that particular node and the target node [4]. The pre-computation for this heuristic to become a good lower bound is massive and depends upon the number of landmarks chosen as for each landmark the 'to' and 'from' distance for every other node in the graph needs to be stored or at least one of the two if the graph is undirected. It further depends on how isolated these landmarks are from each other; creating a set of the landmarks situated at the maximum distances from each other is an N-P Hard problem [7] as finding the maximum length path in a directed cyclic graph is an N-P Hard problem too.

4) Contraction Hierarchy: Here a contraction graph is created from the original graph where every node has a level assigned based on the sequence of contraction and this sequence is based on ordering the nodes by some measure of importance [8]. Contraction of a node deals with adding shortcuts if the shortest path between its neighbors goes via it; this step makes it the most efficient route planning technique but increases the pre-computation time. During the query, an upward graph is formed starting from the source with edges directed from the lower level to a higher level, and a backward graph from the target is also formed in the same manner. Dijkstra algorithm from the source and the target is applied to the upward and downward graph respectively.

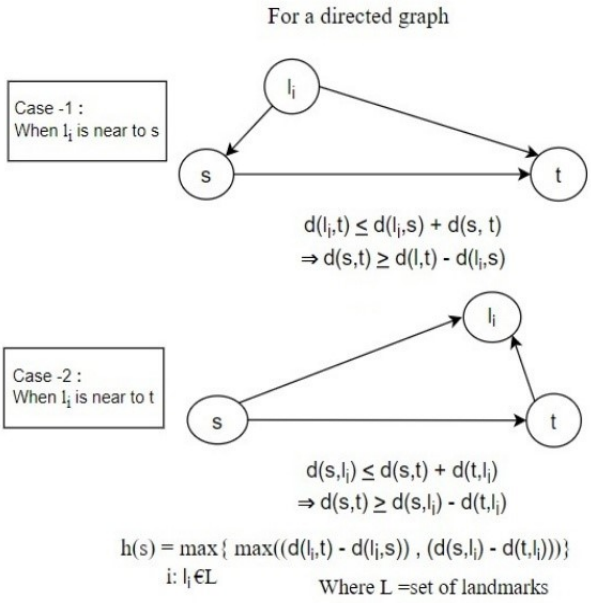


Fig. 1: Calculating the Heuristic Value for Node's' using ALT

From the set of common nodes settled in both the graphs '*S*':

$$dist(s, t) = min\{dist(s, v)+dist(v, t): v \in S; s, t \in V\} \quad (2)$$

The basis of this algorithm is the optimal node ordering (achieved using simple heuristics) [8], based on that forward and backward graphs are formed. Unidirectional Dijkstra turns out to be useless for the hierarchical approach.

The above methodology of applying search algorithms from both the source and target nodes comes under Bidirectional Search and is seen to be much efficient when compared to unidirectional search as the number of nodes expanded is significantly lesser.

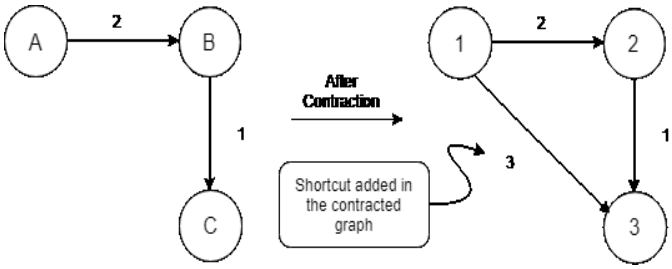


Fig. 2: Creating Contraction Graph (Here, Node Ordering is based on the Sequence of Node Contraction.)

5) Transit Node Routing: Transit Node Routing (TNR) is a technique and not a whole, independent algorithm in itself; for example one of the famous TNR variants uses Contraction Hierarchy pre-computations in its implementation. Its major drawback is the time it spends in

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:58:23 UTC from IEEE Xplore. Restrictions apply.

preprocessing which is significantly larger than even Contraction Hierarchy but this makes it the most efficient algorithm of its type. The underlying observation in TNR is: For multiple long distant routes with common start point and varying destination points, the initial portion of the shortest path route will always be one of the 'few' standard routes [9]. The nodes present at the periphery of this initial portion are called access nodes and a separate list of these nodes is maintained for both the source (say *X*) and the destination (say *Y*) also storing the distance from them respectively. Say $x \in X$ and $y \in Y$ then during query processing, the minimum of *dist(x, source) + dist(x, y) + dist(y, destination)* from every combination of *x* and *y* (total: $|X| + |Y|$) is taken as the shortest path from source to destination.

The most widely known variant of TNR uses Contraction Hierarchy for pre-computation [9]. CH orders the node in such a way that a node occurring in many shortest paths is assigned a higher level in the hierarchy. Hence a set of nodes above a certain threshold level is chosen for selecting the access nodes (called transit nodes). From the source node, a forward search in the upward graph is run and all the transit nodes settled during the run become access nodes. During this Dijkstra run, edges leaving access nodes remain unrelaxed. Similarly, a backward search in the downward graph from the destination node is run to collect its access nodes.

6) Multi-label Algorithms: The above routing algorithms only apply for a scalar number as the edge-weight that is the travel-time but there are also other important criteria that should be taken into consideration while calculating the most optimal path. These factors are discussed in the later section. All the above-mentioned algorithms can be transformed to consider these kinds of criteria but due to the dynamic nature of the road-network, the graphs are also dynamic and the weight also changes with time hence a mean value of these criteria is taken making these algorithms still static in nature.

7) Genetic Algorithm: Genetic algorithm (GA) falls in the range of meta-heuristic inspired by the process of natural selection. It belongs to the larger class of Biologically-inspired algorithms. In this paper, GA Algorithm is primarily used to generate high-quality solutions, enhance and explore the solution to route optimization problems [13].

Working: GA initiates its procedure with a population of candidates (also referred to as a chromosome). Each individual represents that it can be a feasible solution to the problem. Then, each individual is associated with some fitness value according to the fitness function. The fitness function is designed following the solution's quality [15]. The reproduction process generates the next generation of individuals by two distinct processes named 'repair/crossover and 'mutation'.

Need for GA: In maximum real-existence eventualities, the shortest path might not be a great course for the user. In fact, distinct customers may additionally opt for exclusive routes depending on the traffic congestion level, road incidents, and different time various quantities. Therefore, want to explore a multi-objective optimization algorithm that can deal with many distinctive targets simultaneously.

Crossover: It is a process that starts with two fittest individuals (viz. Parents). Also, crossover requires that they should have at least one different gene in both parents and afterward, the traditional method of one point crossover operation is simulated, and finally, any variant of the mutation function is applied on generated offsprings to remove feasible loops and add randomness in the solution [15].

Parent Selection: One of these three specific choice schemes for parent selection:

- Roulette wheel choice
- Rank-primarily based choice
- Match selection (with a tournament size of 2).

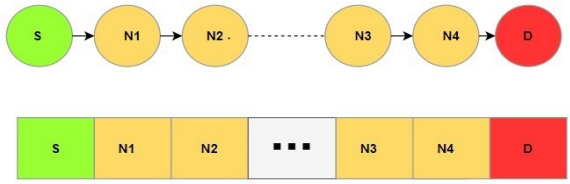


Fig. 3: Representation of a Chromosome

III. Enhancement Techniques for the Route Planning Algorithms

Through this part, the techniques that can be used to enhance the proficiency of the prevailing route planning algorithms is introduced for real and large network of roads based on the information obtained so far. Firstly, discussed the most crucial parameters that must always be considered while designing the algorithms for route planning and then mentioned some other parameters that can be implemented in their designing process to improve their performance on real road networks. After discussing these parameters, found and present some performance evaluation metrics to be used for evaluating the performance of various algorithms used in route planning for real and large road networks.

Now the input parameters are discussed, criteria for determination of the best route and evaluation metrics for route planning algorithms one by one in the next sections.

A. Crucial Parameters for route planning algorithms

Here, the most crucial parameters are discussed that must be always considered as parameters for designing the route planning algorithms. The parameters to be considered are as follows:

- Knowledge about the starting location: The beginning point for travel or the point on the map from where the journey commences.
- Knowledge about the destination location: The point on the map where the journey terminates or the end point for travel. Information regarding the objective for travelling is also useful.
- Features of road network: This information contains the details such as the number of lanes on that particular road, the present traffic scenario of the road, weather

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:58:23 UTC from IEEE Xplore. Restrictions apply.

conditions of the area, information about incidents such as accidents and road blockages etc.

- Vehicle characteristics and other insights: This information contains the details such as vehicle-related conditions including the size and speed of vehicle, the fuel remaining in the fuel tank, any specific condition related to traveller or the path to be chosen, the operating hours of vehicle etc.

Along with the aforementioned parameters, there are some other parameters that can be utilized in the process of designing the route planning algorithms such as the features of vehicle, i.e., its condition and physical specifications and also the information regarding the present traffic and future predictions of traffic on the road network [11]. These mentioned parameters assists in the process of choosing the best route.

For the implementation of routing algorithms on real and large road networks, there is a need to provide the current traffic conditions, the graph of road networks, the traffic prediction data, and the characteristics of vehicle as input parameters. These input parameters are processed in the routing algorithm and continuously updated with respect to the changing traffic conditions and occurrence of specific incidents.

The goal of routing algorithm is to determine the most optimal route to reach the target location (destination) from vehicle's current location (origin). This route is suitably changed in order to prevent such updates from affecting the selected travel path as well as making the route unavailable for variable duration and have to note that the features of vehicle is a very crucial input for route planning as vehicle's weight, its height and its vehicle class (e.g. bus, car, train) are crucial parameters to determine the best route in accordance to the physical capacities of the road networks (like lanes, flyovers, underpasses, type of highway etc.) and the driving regulations of our country.

B. Criteria for determination of the best route for these algorithms

In case needed to determine the best route from the start location to the destination location as per the request of driver, then first deduce the metric or the set of metrics which must be utilized in order to compare the different routes from a set of routes for that specific vehicle [11]. In order to measure and identify an optimal route by the objective function, the objective function to be composed of a single metric or a composed of a set of metrics applicable to any of the aforementioned algorithms is needed. In some studies, researchers have applied a wide variety of parameters in their routing algorithms. Now in the further section, presented the most valuable metrics that the route planning algorithms may utilize to find out the requested route as per our need.

- Time taken for travel: According to this metric, tend to look for the quickest route for journey instead of the smallest one (in terms of distance) among all the available routes. The quickest route implies a solution path that converges to the destination within the shortest time. Due to the dynamic nature of road and its associated parameters, the quickest route might not be the same as the smallest route. Furthermore, needed to update the quickest path continuously as the vehicle moves towards its destination and that too in accordance with the traffic conditions on road networks which change rapidly.

- Travel distance: In a road network, the values of the length of all roads have a specific value. When talk about finding the shortest path, it means interested in a path for a given vehicle through which it travels with the least distance from the start point to the destination point on its journey.

- Cost of travel: The cost of travel refers to all the expenses that have been incurred during the journey. It includes the cost of fuel used in addition to the total amount of toll taxes paid by the vehicle's driver on the route. This cost is affected by several factors such as the length of path to cover, the number of halting points on the route, the vehicle-type and the dimension and condition of road chosen for travel. So, to get a cost efficient route, routing algorithms can use this criterion, and provide a low cost route for reaching the destination location to the driver.

- Traffic conditions: The traffic conditions of the road network refer to the traffic congestion levels. If the traffic congestion on one route is less as compared to the other route, the vehicle must take the path having less traffic congestion, if the travel time for this path is also less, so as to reach the destination location in minimal time.

- Ease of driving: In a road network, the driver tends to choose a route which is smooth, safe and comfortable. This metric can be thought of a metric solely based on the driver's preference and several other parameters such as the dimensions of road, the number of traffic signals on road, the number of lanes and turns on road etc.

- Integration of two or more metrics: As per this criterion, two or more of the aforementioned criteria can be integrated with one another to act as a single metric to handle the preferences of a specific driver.

C. Evaluation metrics for route planning algorithms

Evaluation metrics are the certain parameters that are used for evaluating the performance of a particular algorithm. These metrics assist in differentiating between the different routing algorithms based on their performance measure. The four significant evaluation metrics that can be used for evaluating route planning algorithms are computational complexity, execution time, quality of the solution, and scalability.

- Computational complexity: Computational complexity refers to the computational performance of a particular route planning algorithms and the resources it requires for its execution. For instance, an algorithm may not be the best routing algorithm if it finds the most optimal route with a very high complexity, even if has comparatively low journey time, especially when the route needs to be continuously updated during the vehicle's journey.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:58:23 UTC from IEEE Xplore. Restrictions apply.

- Execution time: The execution time of these algorithms also serves as an evaluation criteria. The algorithms having overall less execution time for a variety of route is considered to be better than the others.
- Quality of the solution: In case to find multiple best routes, need to use a metric to find the best out of all the available routes. For this purpose, use this particular metric to compare the different solutions obtained by using different algorithms based on various parameters mentioned earlier, and then determining which of these algorithms provides the most optimal route by giving a value closest to the solution.
- Scalability: Scalability for a routing algorithm refers to the change in the performance and efficiency of a routing algorithm when the real network of roads is scaled up or down in terms of size. Hence, an algorithm which is efficient for a particular area might be inefficient or inapplicable when the area under consideration is increased in size.

IV. Experimental Results And Analysis

To fulfil the aim i.e. to present a comparison of various route planning algorithms, the above-stated algorithms namely-: Dijkstra, ALT, Contraction-Hierarchy, and Transit Node Routing algorithms (plus bidirectional variants) were implemented in Java on Eclipse IDE. On a simple weighted directed (cyclic) graph with no self-loop having a total of 422047 nodes and 1011274 edges obtained from parsing the OSM road-network data for the union territory of India 'Delhi', following line graphs (Execution Time (in ms) vs. Query number) were obtained when aforementioned algorithms were run for 420 shortest path queries between randomly chosen source and target nodes:

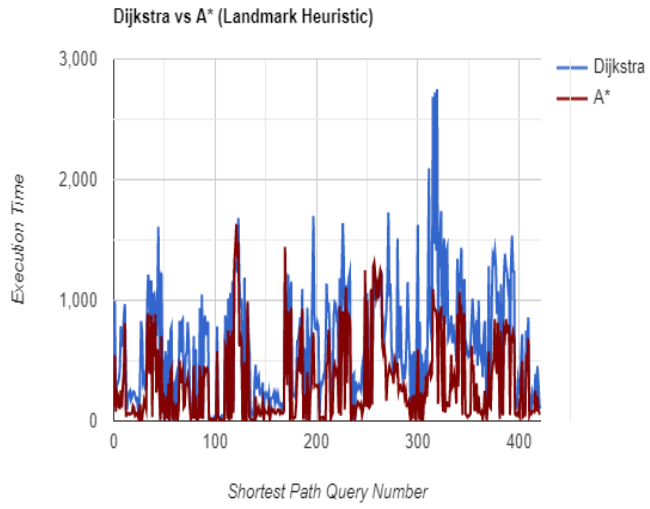


Fig. 4: A Comparison between Dijkstra and ALT based on Execution Time.

- Apparently, for most of the queries, ALT algorithm proved to be more efficient than the Dijkstra algorithm.
- On testing against various values of landmark-set size, the threshold for the given graph is 8 as beyond this value, no significant improvement in efficiency of the ALT algorithm was observed.

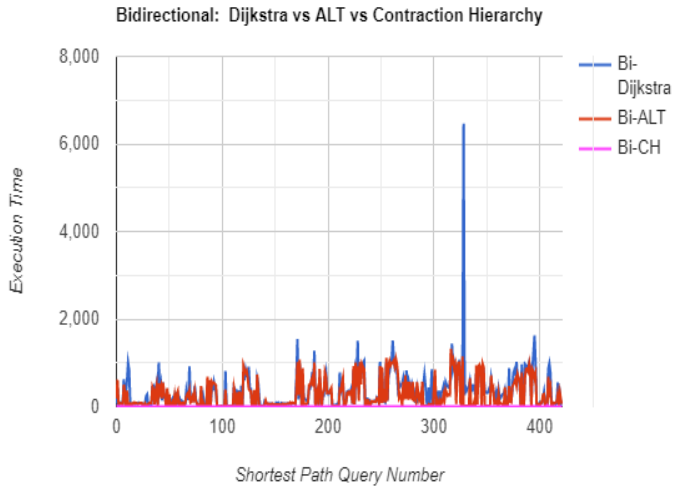


Fig. 5: A Comparison between Bidirectional Dijkstra, Bidirectional ALT and Bidirectional Contraction Hierarchy.

- Here, the blue peaks in the graph depict that the Bidirectional Dijkstra algorithm is least efficient amongst the three.
- As expected, Contraction Hierarchy took negligible time to calculate the optimal path for any of the queries.
- The difference between the execution time of Bidirectional Dijkstra and ALT was thin in many queries due to smaller path-length.

Fig. 5 shows the bidirectional variant of the three algorithms namely Dijkstra, ALT, and CH. When Bidirectional Dijkstra and ALT were run on the same 420 queries, results were difficult to interpret. Sometimes ALT performed better and other times Bidirectional Dijkstra had upper hand. But in the below graph (Fig. 6), the efficiency of Bidirectional Dijkstra emerged clearly for relatively longer path lengths where ALT took much longer times to process those queries.

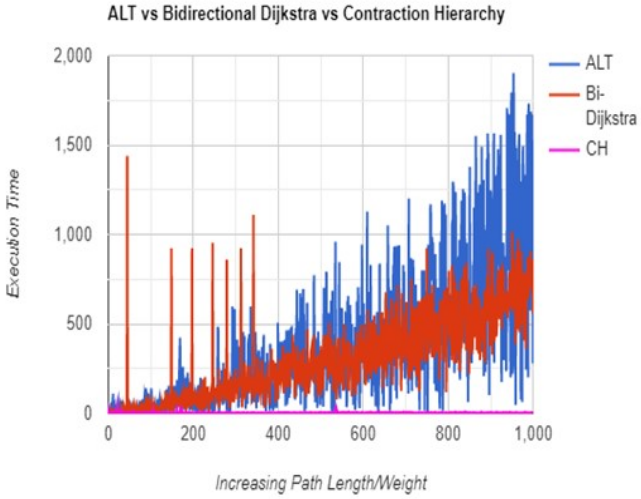


Fig. 6: A Comparison between Bidirectional Dijkstra, ALT and CH based on Increasing Path Length.

- The longest path length was estimated to be around 1300 km (by running on 10 different queries).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:58:23 UTC from IEEE Xplore. Restrictions apply.

- Bidirectional Dijkstra proved to be better than ALT when the path length was relatively larger with respect to the longest path length found in the graph [2] while CH outperforms all of them.

The below Figs. (Fig. 7 and 8) were obtained after running Transit Node Routing and Contraction Hierarchy on the simple directed weighted graph obtained from Andorra.osm file. Due to heavy preprocessing required by both the algorithms Andorra.osm was chosen to create a smaller graph. Andorra is a European Country with an area of only 468 km^2. From the bar graphs, it can be clearly seen that the preprocessing and query time are completely opposite of each other for TNR and CH. The time spent during preprocessing is compensated with the improvement seen during the shortest path query processing in the case of Transit Node Routing.

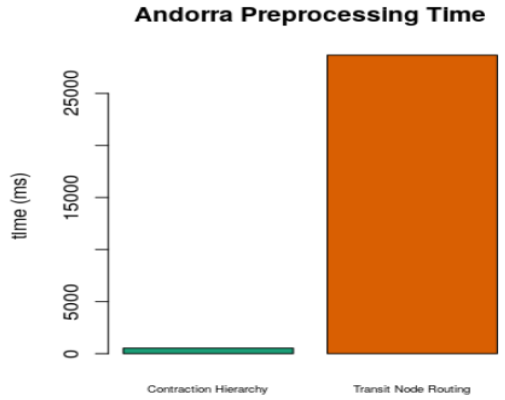


Fig. 7: Bar graph for comparing pre-computation time between CH (in green) and TNR (in orange)

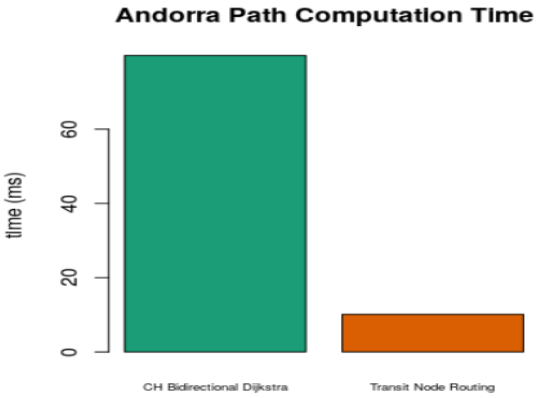


Fig. 8: Bar graph for comparing query execution time between CH (in green) and TNR (in orange)

V. Conclusion

It can be concluded that in order to create a more efficient route planning algorithm a compromise has to be made by doing more pre-processing. Transit Node Routing is the fastest route planning technique with approximately constant query processing time but with the highest pre-computation time. Contraction Hierarchy is one of the simplest ways to calculate the shortest path distance between two nodes in a graph and A* algorithms can also be fast with a good admissible heuristics while finding the shortest path is not always effective in real-time scenarios of the motorway systems as there are several parameters, the most important being the traffic that also contribute majorly in determining the optimal path.

References

[1] P. E. HART, N. J. NILSSON, AND B. RAPHAEL., "A FORMAL BASIS FOR THE HEURISTIC DETERMINATION OF MINIMUM COST PATHS IN GRAPHS." IEEE TRANS. SYST. SCI. AND CYBERNETICS, SSC-4(2):100–107, 1968.

[2] SAWLANI, SNEHA, "EXPLAINING THE PERFORMANCE OF BIDIRECTIONAL DIJKSTRA AND A* ON ROAD NETWORKS" (2017). ELECTRONIC THESES AND DISSERTATIONS. 1303.

[3] THIPPESWAMY.K,HANUMANTHAPPA.J., DR.MANJAIAH D.H. "A STUDY ON CONTRAST AND COMPARISON BETWEEN BELLMAN-FORD ALGORITHM AND DIJKSTRA'S ALGORITHM" (2015).

[4] A. V. GOLDBERG AND C. HARRELSON., "COMPUTING THE SHORTEST PATH: A* SEARCH MEETS GRAPH THEORY." PROCEEDINGS OF THE SIXTEENTH ANNUAL ACM-SIAM SYMPOSIUM ON DISCRETE ALGORITHMS, PAGES 156–165, 2005.

[5] K.ROHILA, P.GOUTHAMI, PRIYA M, "DIJKSTRA'S SHORTEST PATH ALGORITHM FOR ROAD NETWORK",IJIRCCE,VOL. 2, ISSUE 10, OCTOBER 2014,ISSN(ONLINE): 2320-9801

[6] DOMINIK SCHULTES, "ROUTE PLANNING IN ROAD NETWORKS",PAGE 22-29

[7] M. POTIMIAS, F. BONCHI, C. CASTILLO, A. GIONIS, "FAST SHORTEST PATH DISTANCE ESTIMATION IN LARGE NETWORKS" (2009)

[8] R. GEISBERGER, P. SANDERS, D. SCHULTES, D. DELLING, "CONTRACTION HIERARCHIES: FASTER AND SIMPLER HIERARCHICAL ROUTING IN ROAD NETWORKS" (2008)

[9] J. ARZ, D. LUXEN, P. SANDERS, "TRANSIT NODE ROUTING RECONSIDERED." (2013)

[10] LU FL, XIAO JX. , "AN IMPROVEMENT OF THE SHORTEST PATH ALGORITHM BASED ON DIJKSTRA ALGORITHM." NEW YORK: IEEE; 2010.

[11] VI NGOC-NHA TRAN,JOHN MURPHY, "A COMPARATIVE STUDY OF VEHICLES' ROUTING ALGORITHMS FOR ROUTE PLANNING IN SMART CITIES." CONFERENCE: 2012 FIRST INTERNATIONAL WORKSHOP ON VEHICULAR TRAFFIC MANAGEMENT FOR SMART CITIES (VTM).

[12] SURANGA NANAYAKKARA,D. SRINIVASAN,LAI WEI LUP, "A COMPARATIVE ANALYSIS OF MULTI CRITERA ROAD NETWORK, MUHAMMAD S. F. QURESHI, SYED MUJTABA SHAH, MR. GHARIB ISMAIL GHARIB AL-MATROUSHI" 2013.

[13] LINGARAJ, HALDURAI. (2016), "A STUDY ON GENETIC ALGORITHM AND ITS APPLICATIONS.INTERNATIONAL JOURNAL OF COMPUTER SCIENCES AND ENGINEERING." 4. 139-143.

[14] "ROUTE OPTIMIZATION TECHNIQUE: AN OVERVIEW", BALE, D. L. T. UGWU, C. NWACHUKWU, E. O., DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF NIGERIA.

[15] XAVIER GERMAN, "GENETIC ALGORITHM BASED ROUTE PLANNER FOR LARGE URBAN STREET NETWORKS." DOI:10.1109/CEC.2007.4425056 • SOURCE: IEEE XPLORE CONFERENCE: EVOLUTIONARY COMPUTATION, 2007

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:58:23 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Comparative Study of Static VAR Compensation Techniques – Thyristor Switched Reactor and Thyristor Switched Capacitor

Divyam
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
fastrack2030@gmail.com

Ayushi Saxena
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
ayushi.saxena1407@gmail.com

Bharat Singh
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
bharatsinghuae@gmail.com

J. N. Rai
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
jnrai@dce.ac.in

Corresponding Author: Bharat Singh, bharatsinghuae@gmail.com

***Abstract*— The paper makes a comparative study of the two Static VAR Compensation (SVC) techniques - Thyristor Switched Capacitor (TSC) and Thyristor Switched Reactor (TSR). These techniques are defined, examined and compared. The major constituent of the power flow should be active power and a sufficient and minimal amount of reactive power should be allowed to flow in the transmission line. For power system stability and reliability, there arises a need to compensate the reactive power flow. This Reactive Power Compensation has been done on a medium transmission line, which was modelled by scaling down to practically achievable parameters in a laboratory setup, for different load values. Thus, the paper also enables appropriate and relevant selection of the FACTS Controller based on the type and rating of load and also provides scope for implementation in the laboratory.**

Keywords—flexible AC transmission system, static VAR compensation, modelling of transmission line, thyristor switched reactor, thyristor switched capacitor

Acronyms

- ***FACTS (Flexible AC Transmission System)***
- ***SVC (Static VAR Compensation)***
- ***TSR (Thyristor Switched Reactor)***
- ***TSC (Thyristor Switched Capacitor)***

I. INTRODUCTION

FACTS is a system made up of power electronic components and power system components to improve the stability and controllability of the power transmission line system and improve its power transferring capability. The introduction of the power electronic devices such as the thyristor created new avenues for the invention of devices such as FACTS controllers. In short, the FACT system is utilised to improve the controlling capability of the power system transmission network by using power electronic devices and by introducing capacitive or inductive power in the power system.

In conventional AC power transmission networks, the capability of transferring AC power is restricted due to the effect of several factors such as voltage limit, short circuit current limit, transient stability limit, thermal limits etc. These factors provide the maximum limit for the electrical power that can be effectively transferred via the transmission lines without making any harm to the transmission lines and the electrical equipment involved. This is done typically by bringing about modifications in the power transmission system structure. Although, this is not an easy task in the modern power system. Thus, alternative method of getting highest power transmission capacity without any modifications in the power system structure is preferred. With the use of inconstant impedance electrical components such as inductors and capacitors, the entire energy or power is not able to reach the load from the source. Instead, a part of this is stored in such devices in the form of reactive power and reaches back to the source. This enables us to meet the reactive power demand of the load locally at the receiving end itself, and thus, the substantial quantity of power transmitted to the load or the active power is always lower than the apparent power. This also helps in maintaining the receiving end voltages at desirable values and increases the reliability of the power system. For optimal power transmission, the real power should always be equal to the total power or the power factor should be unity. This is where the role of reactive power compensation comes in achieving FACTS. [3]

II. STATIC VAR COMPENSATION

SVC is considered as the one of the earliest generations of FACTS controllers. The compensator is made up of a rapid thyristor regulating a shunt capacitive bank and an inductor to impart compensation in shunt. They are made up of shunt associated inconstant impedance components such as capacitor or inductor whose output can be modified via thyristor switches, to incorporate these inductive or capacitive reactance into the power line. They may also be located in the centre of the transmission line to enhance the maximal power transmission capacity and at the terminal of the transmission line to make up for the fluctuations arising out of different loading. The two types of SVCs are-

A. TSR

It is basically made up of a inductor which is shunt connected and the impedance of which is controlled using a Thyristor switch in a stepwise manner. TSR comprises of a reactor in series with bidirectional thyristor switches. This method is used to adjust the received reactive power by the load in a system with varying conditions. The difference in the sending and receiving reactive power leads to the difference in sending and receiving end voltages, TSR helps to reduce the difference between these two voltages and improves the voltage regulation of a power system. The inductor absorbs

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:54:43 UTC from IEEE Xplore. Restrictions apply.

the excess lagging VARs present in a capacitive system and maintains a balance of the reactive power, thus equalizing voltages at both the ends of the transmission line. [4]

B. TSC

It is basically made up of a capacitor which is shunt connected and the impedance of which is controlled using a Thyristor switch in a stepwise manner. TSC comprises of a capacitor in series with bidirectional thyristor switches. This method is used to adjust the received reactive power by the load in a system with varying conditions. The difference in the sending and receiving reactive power leads to the difference in sending and receiving end voltages, TSC helps to reduce the difference between these two voltages and improves the voltage regulation of a power system. The capacitor present in TSC provides the excess lagging VARs present in an inductive system and maintains a balance of the reactive power, thus equalizing voltages at both the ends of the transmission line. [5]

III. METHODOLOGY

A. Modelling of Transmission Line [1]

A Π-model of transposed medium transmission line with length 200 km, voltage level of 220 kV and the rating of 500 MVA is chosen for the analysis in order to inculcate the effect of both TSC and TSR and to help in their comparison. The transmission line model used is shown in Fig. 1 and Fig. 2. [8]

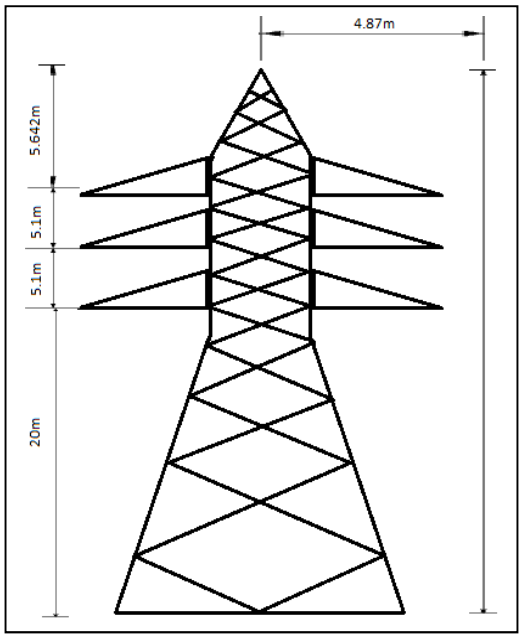


Fig. 1. Transmission Line Tower Design

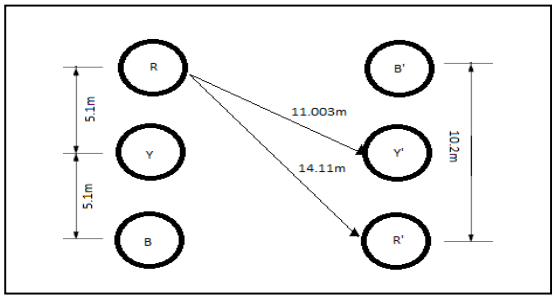


Fig. 2. Transposed Model for Transmission Line

Calculating Geometric Mean Distance (GMD), [7]

$D_{m1} = D_{m3} = [(5.1)(10.2)(9.75)](11.003)]^{1/4}$ m

$D_{m2} = [(11.003)(11.003)(5.1)(5.1)]^{1/4}$ m

Thus, $GMD = [D_{m1}D_{m2}D_{m3}]^{1/3} = 8.24$ m (1)

where D_{m1}, D_{m2} and D_{m3} are the mutual geometric mean distances for the three phases.

Calculating Geometric Mean Radius (GMR), [7]

$D_{s1} = D_{s3} = [(14.11)(0.7788 \times 7.5 \times 10^{-3})]^{1/2}$ m

$D_{s2} = [(9.75)(0.7788 \times 7.5 \times 10^{-3})]^{1/2}$ m

Thus, $GMR = [D_{s1}D_{s2}D_{s3}]^{1/3} = 0.27$ m (2)

where D_{s1}, D_{s2} and D_{s3} are the self-geometric mean distances for the three phases.

Inductance per phase of transmission line from (1) and (2),

$L_{ph} = \frac{\mu}{2\pi}\ln(\frac{GMD}{GMR})$ H/m = 0.683mH/km [7]

Capacitance per phase of transmission line from (1) and (2),

$C_{ph} = \frac{2\pi\varepsilon}{\ln(\frac{GMD}{GMR})}$ F/m = 16.26 nF/km [7]

B. Transmission Line Ratings

Length, L = 200 km
Voltage, V = 220 kV
Frequency, f = 50 Hz
Power, S = 500 MVA
L_{ph} = 0.683 mH/km
C_{ph} = 16.26 nF/km

C. Per Unit Calculations for Transmission Line Parameters

Current in line, $I = \frac{S}{V} = 2.2727$ kA

Base impedance, $Z = \frac{V}{I} = 96.8\ \Omega$

Total inductance for 200 km,

L = (0.683 x 200) mH = 136.6 mH

Total capacitance for 200 km,

C = (16.26 x 200) nF = 3.252 µF

$X_L = \omega L = 2\pi fL = 42.91415\ \Omega$

$X_C = \frac{1}{\omega C} = 978.81268\ \Omega$

X_L (p.u.) $= \frac{XL}{Z} = 0.44332$ p.u.

X_C (p.u.) $= \frac{Xc}{Z} = 10.1117$ p.u.

Y_C (p.u.) = 0.098895 p.u.

D. Scaling down Transmission Line Parameters for 5kVA, 220V model

Current in line, $I' = \frac{S\prime}{V\prime} = 2.2727$ kA

Base impedance, $Z' = \frac{V\prime}{I\prime} = 9.68\ \Omega$

X_L (actual) = X_L (p.u.) x Z' = 4.2913 Ω

L (actual) $= \frac{XL}{\omega} = 13.66$ mH

X_C (actual) = X_C (p.u.) x Z' = 97.88125 Ω

C (actual) = ω X_C = 32.52 µF

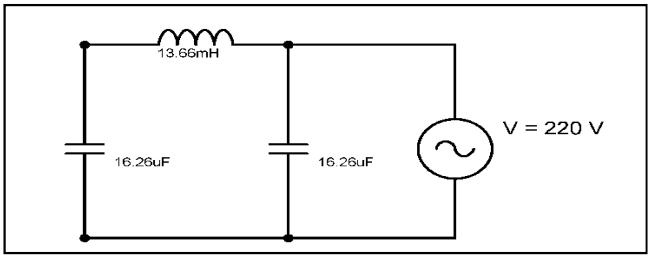


Fig. 3. Transmission Line Model at 5kVA, 220V

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:54:43 UTC from IEEE Xplore. Restrictions apply.

E. ABCD Parameters of Transmission Line [7]

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} \frac{YZ}{2}+1 & Z \\ Y\left(\frac{YZ}{4}+1\right) & \frac{YZ}{2}+1 \end{bmatrix}$$

$$= \begin{bmatrix} 1.02192 & j0.44332 \\ j0.099978 & 1.02192 \end{bmatrix}; \text{A} = 0^0, \text{B} = 90^0$$

F. Load Values

In order to achieve a complete and effective comparative analysis, both the leading and lagging loads are considered having a fixed power rating, i.e. 200 W, which is practically feasible to implement, for various power factor values.

TABLE I. LOAD VALUES FOR LAGGING LOADS

S.No.	Load Values	R (in Ω)	L (in H)
1.	0.1 pf lag, 200 W	2.42	0.0766
2.	0.2 pf lag, 200 W	9.68	0.1509
3.	0.3 pf lag, 200 W	21.78	0.2204
4.	0.4 pf lag, 200 W	38.72	0.2824
5.	0.5 pf lag, 200 W	60.5	0.3355
6.	0.6 pf lag, 200 W	87.11	0.3697

TABLE II. LOAD VALUES FOR LEADING LOADS

S.No.	Load Values	R (in Ω)	C (in µF)
1.	0.1 pf lead, 200 W	2.42	132.195
2.	0.2 pf lead, 200 W	9.68	67.123
3.	0.3 pf lead, 200 W	21.78	45.9613
.	0.4 pf lead, 200 W	38.72	35.8785
5.	0.5 pf lead, 200W	60.5	30.376
6.	0.6 pf lead, 200 W	87.11	27.4028

G. Equations Involved [7]

The real and reactive power at receiving end of the transmission line can be given as:

$$\text{P}_\text{R} = \left|\frac{VsVr}{B}\right| \cos(\beta - \delta) - \left|\frac{AVr^2}{B}\right| \cos(\beta - \alpha) \quad (1)$$

$$\text{Q}_\text{R} = \left|\frac{VsVr}{B}\right| \sin(\beta - \delta) - \left|\frac{AVr^2}{B}\right| \sin(\beta - \alpha) \quad (2)$$

For the transmission line under analysis, $\beta = 90^0$, $\alpha = 0^0$, $V_R = 1$ and $V_S = 1$, and P_R can be calculated from the load values used. Thus, δ can be calculated from (1). This value can be substituted in (2) to find Q_R.

Also, lagging VARs requirement of the transmission line with load is $|Q_R|$ and

$$Q_{cll} = - Q_R$$

Reactive power of load,

$$|Q_{load}| = P_R \tan \varphi$$

where cos φ is the power factor of the load.

For lagging loads, there is reactive power requirement and thus, a shunt capacitor is required for providing this reactive power and thus to compensate the reactive power.

Lagging VARs (in p.u.) provided by the shunt capacitor,

$$Q_c = |Q_{load}| - |Q_{cll}|$$

Thus, value of required capacitor,

$$\text{C}_\text{reqd} = \frac{|Qc| \times 5000}{\omega V^2} \text{ F} \quad (3)$$

For leading loads, there is a reactive power supply and thus these surplus lag VARs have to be absorbed by the shunt reactor and thus to compensate the reactive power.

Lagging VARs (in p.u.) provided by the shunt reactor,

$$Q_L = |Q_{load}| + |Q_{cll}|$$

Thus, value of required reactor,

$$\text{L}_\text{reqd} = \frac{\omega |Q_L| \times 5000}{V^2} \text{ H} \quad (4)$$

IV. SIMULATIONS

The model was simulated for all lagging, leading and no load values as in Table I and Table II. In the simulations, a triac has been used instead of back to back thyristor, which gives similar effect as two back to back thyristor with common gate to provide the bidirectional flow of power. Fig. 4, Fig. 5 and Fig. 6 show the snapshot of the simulation diagram for lagging load, no lead and leading load respectively.

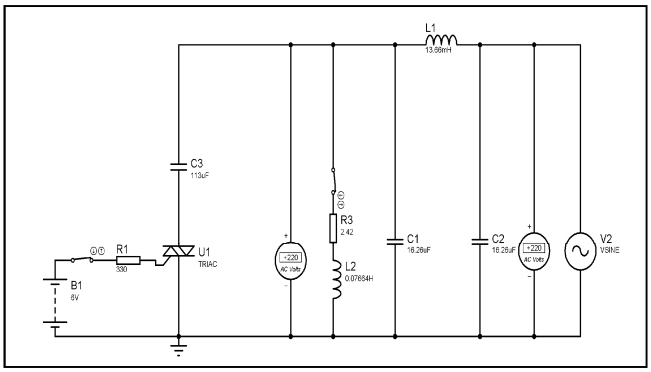


Fig. 4. Simulation snapshot for 0.1 pf lag, 200 W load using TSC

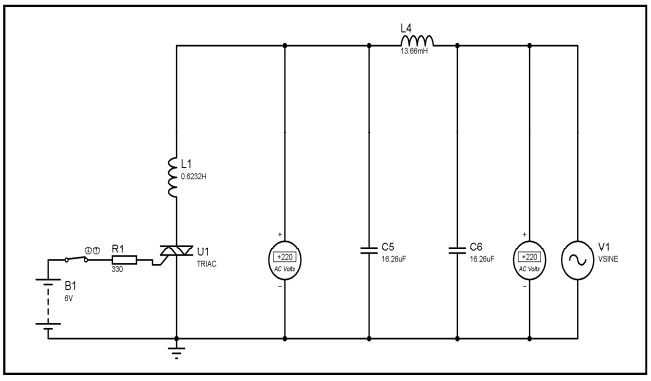


Fig. 5. Simulation snapshot for no load using TSR

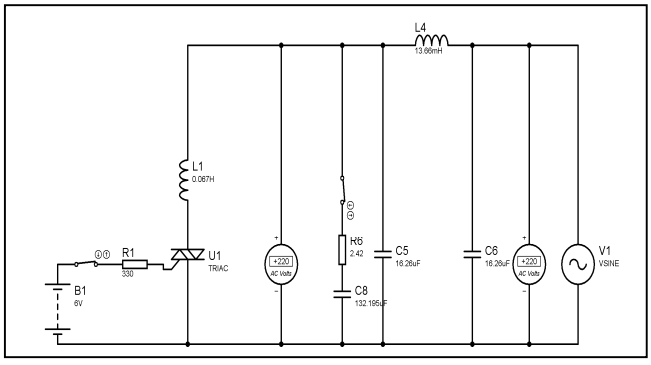


Fig. 6. Simulation snapshot for 0.1 pf lead, 200 W load using TSR

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:54:43 UTC from IEEE Xplore. Restrictions apply.

V. CALCULATIONS

The required $|Q_c|$ and $|Q_L|$ has been calculated for the load values and corresponding C_{reqd} and L_{reqd} has been calculated using (3) and (4) for each of the load values. After obtaining the required values of capacitor and inductor for compensating each of the loads, the voltage before and after compensation is noted through simulation.

TABLE III. CALCULATIONS FOR C_{REQD} FOR LAGGING LOADS

Load	Load Values		$\|Q_C\|$ (p.u.)	C_{reqd} (µF)	Compensation	
					Before	After
	R (Ω)	L (H)			V_R	V_R
0.1 pf lag, 200 W	2.42	0.0766	0.3481	113.0	191	220
0.2 pf lag, 200 W	9.68	0.1509	0.1461	48.062	207	220
0.3 pf lag, 200 W	21.78	0.2204	0.0773	25.449	213	220
0.4 pf lag, 200 W	38.72	0.2824	0.0418	13.762	216	220
0.5 pf lag, 200 W	60.5	0.3355	0.0194	6.4067	218	220
0.6 pf lag, 200 W	87.11	0.3697	0.0035	1.1622	220	220

TABLE IV. CALCULATIONS FOR L_{REQD} FOR NO LOAD

Load	$\|Q_L\|$ (p.u.)	L_{reqd} (H)	Voltage (V)			
			Before Compensation		After Compensation	
			VS	VR	VS	V_R
No Load	0.049445	0.6232	220	225	220	220

TABLE V. CALCULATIONS FOR L_{REQD} LEADING LOADS

Load	Load Values		$\|Q_L\|$ (p.u.)	L_{reqd} (H)	Compensation	
					Before	After
	R (Ω)	C (µF)			V_R	V_R
0.1 pf lead, 200 W	2.42	132.195	0.4477	0.067	276	220
0.2 pf lead, 200 W	9.68	67.123	0.2457	0.1253	248	220
0.3 pf lead, 200 W	21.78	45.9613	0.1769	0.1740	239	220
0.4 pf lead, 200 W	38.72	35.8785	0.1414	0.178	235	220
0.5 pf lead, 200 W	60.5	30.376	0.1190	0.2587	232	220
0.6 pf lead, 200 W	87.11	27.4028	0.1031	0.2987	230	220

VI. OBSERVATIONS

A. Lagging Loads

TABLE VI. OBSERVATIONS FOR LAGGING LOADS

pf	Q_{load} (VARs)	Q_C (VARs)
0.1 lag	1989.97	1740.975
0.2 lag	979.79	730.800
0.3 lag	635.96	386.965
0.4 lag	458.25	209.265
0.5 lag	346.41	97.415
0.6 lag	266.67	17.67

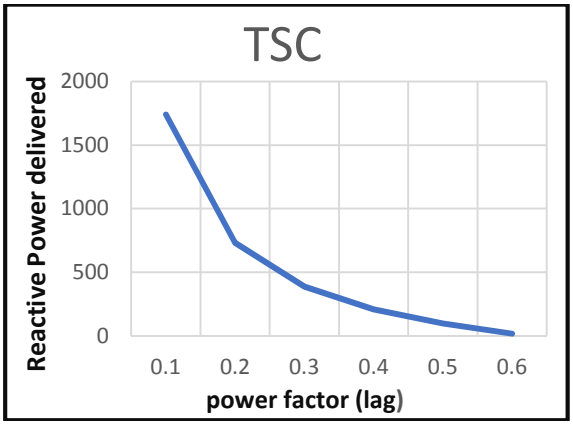


Fig. 7. Reactive Power Delivered Vs Power Factor for lagging loads

From Table VI and Fig. 7, as shown above, it is observed that for a given real power demand (200W), as the power factor of the load is increased from 0.1 to 0.2, the reactive power demand of the load decreases by 1010.17 VARs. Further, as the power factor increases from 0.5 to 0.6, the reactive power demand decreases by 79.74 VARs. Therefore, the requirement of the compensating reactive VARs to be supplied by the TSC decreases as the power factor increases.

As the power factor improved to 0.6 pf lag then the reactive power demand of the lagging load became almost equal to the lagging power supplied by the capacitive transmission line and hence the requirement of the compensation is almost negligible (17.67 VARs), and is depicted by the graph in Fig. 7, tending to zero.

B. No Load

Under no load conditions, it is observed from Table IV, that for proper compensation, an inductor is required. This proves that a medium transmission line, having a length of 200km has a dominating capacitive nature, and due to this capacitance effect, surplus reactive power is injected in the transmission line and is responsible for the increase in the voltage level of receiving end. This is phenomenon of increase in the receiving end voltage due to capacitance effect of the transmission line is known as Ferranti Effect.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:54:43 UTC from IEEE Xplore. Restrictions apply.

C. Leading Loads

TABLE VII. OBSERVATIONS FOR LEADING LOADS

pf	Q_{load} (VARs)	Q_L (VARs)
0.1 lead	1989.97	2238.97
0.2 lead	979.79	1228.75
0.3 lead	635.96	884.995
0.4 lead	458.25	707.252
0.5 lead	346.41	595.405
0.6 lead	266.67	515.66

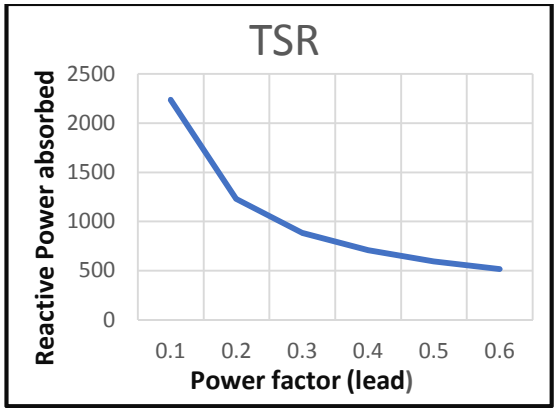


Fig. 8. Reactive Power Absorbed Vs Power Factor for leading loads

From Table VII and Fig. 8 as shown above, it is observed that for a given real power demand (200W), as the power factor of the load is increased from 0.1 to 0.2, the reactive power absorbed decreases by 1010.2 VARs. Further, as the power factor increases from 0.5 to 0.6, the reactive power absorbed decreases by 79.74 VARs. Therefore the requirement of compensating reactive VARs to be absorbed by the TSR decreases as the power factor increases.

VII. CONCLUSION

The paper presented the comparison of FACTS Static VAR Controllers, i.e. TSC and TSR on a medium transmission line for varying loads (No Load, Lagging and Leading loads) and with the help of modelling and mathematical calculations, in order to compensate the reactive power, thereby attaining equal sending end and receiving end voltage levels. We can also conclude that the above chosen transmission line depicted Ferranti Effect and hence TSR is required to compensate this effect. The method employed to achieve the improvements outlined are, introduction of a TSR to absorb the reactive power and TSC to deliver the reactive power demand. This improves the capacity of an already built transmission line and is used to meet the increasing load demands, by locally meeting the reactive power demands at the receiving end itself, and hence achieving the goal of flexible transmission and voltage regulation.

The above comparative study can be experimentally implemented and verified in a laboratory setup. The accuracy of the experimental results in the laboratory setup may vary on the availability of the exact capacitor and inductor rating. Though the above study has been implemented for a medium transmission line, in future, the study can be extended for short and long transmission lines as well.

ACKNOWLEDGMENT

We would like to thank Prof J.N. Rai, for his constant motivation and guidance during the project. Without his timely help and support, the completion of this paper wouldn't have been possible. We would also like to thank the publisher of the research papers from where we have taken reference.

REFERENCES

[1] Mr.Vinayak Gaikwad, Suyash Tripathi, Yash Wanjari, Ankesh Thakre, Diksha Dakare, Jayant Shendre, "Laboratory Setup For Long Transmission Line", International Research Journal of Engineering and Technology (IRJET), Volume: 04 Issue: 03, Mar -2017.

[2] Jai Damania, Karuna Nikum, Shoeb Shaikh, Hardik Save, "Flexible AC Transmission System using TCR", International Journal of Engineering Research & Technology (IJERT) 2017.

[3] Amaranath Sanyal, Archita Ghosh, Raju Basak, Sourish Sanyal, Subhasree Chatterjee, Deepayan Biswas "An Article On Flexible AC Transmission System" IRJET Volume 3, Issue 5, May 2016.

[4] R.T Kalapad, "Review on FACTS devices integrated with TSR" International Journal of Engineering Research and Applications (IJERA), October 2015.

[5] Aashish Goyal, Neharika Kapil, Sheila Mahapatra, "Implementation of Thyristor Switched Capacitors for Power Factor Improvement" Advance in Electronic and Electric Engineering. ISSN 2231-1297, Volume 4, Number 3 (2014).

[6] Rudra Rameshwar, "Efficient, Reliable and Cost Effective Transmission of Electrical Energy using FACTs", National Power System Conference (NPSC), 2004.

[7] "Power System Analysis" , John Grainger, William Stevenson, Jr.

[8] R. Saravanan, "Transmission Line Design Concept", SlideShare.

[9] Rasal Madhuri B., Tagad Bharati T., Surwase Komal R., Prof. Tapre.P.C, "Ferranti Effect Overcome By Using Thyristor Switched Reactor (TSR) In FACTS" IJARIIE-ISSN(O)-2395-4396, Vol-5 Issue-3 2019.

[10] "Modern Power System Analysis", Fourth Edition, D. P. Kothari, I. J. Nagrath.

[11] Jyotideep Deka, Kamaldeep Sharma, Krishan Arora, "Study Of 132 Kv Transmission Line Design And Calculation Of Its Parameters", International Journal for Electrical and Electronics Engineers, Vol. 8, Issue-01, January-June, 2016.

[12] Rishabh Jain , Sunil Kumar Sahu , C Venkata Srikar , Rohit , Suvit Chitnis , Mr. Mousam Sharma, "FACTS (Flexible AC transmission system) by using TSR (Thyristor Switch Reactance)", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 6, Issue 4, April 2017.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:54:43 UTC from IEEE Xplore. Restrictions apply.

Air Quality, Atmosphere & Health
https://doi.org/10.1007/s11869-020-00881-z

Comparative study on air quality status in Indian and Chinese cities before and during the COVID-19 lockdown period

Aviral Agarwal[1] · Aman Kaushik[1] · Sankalp Kumar[1] · Rajeev Kumar Mishra[1]

Received: 26 May 2020 / Accepted: 15 July 2020
© Springer Nature B.V. 2020

Abstract

Amidst COVID-19 pandemic, extreme steps have been taken by countries globally. Lockdown enforcement has emerged as one of the mitigating measures to reduce the community spread of the virus. With a reduction in major anthropogenic activities, a visible improvement in air quality has been recorded in urban centres. Hazardous air quality in countries like India and China leads to high mortality rates from cardiovascular diseases. The present article deals with 6 megacities in India and 6 cities in Hubei province, China, where strict lockdown measures were imposed. The real-time concentration of $PM_{2.5}$ and NO_2 were recorded at different monitoring stations in the cities for 3 months, i.e. January, February, and March for China and February, March, and April for India. The concentration data is converted into AQI according to US EPA parameters and the monthly and weekly averages are calculated for all the cities. Cities in China and India after 1 week of lockdown recorded an average drop in $AQI_{PM2.5}$ and AQI_{NO2} of 11.32% and 48.61% and 20.21% and 59.26%, respectively. The results indicate that the drop in AQI_{NO2} was instantaneous as compared with the gradual drop in $AQI_{PM2.5}$. The lockdown in China and India led to a final drop in $AQI_{PM2.5}$ of 45.25% and 64.65% and in AQI_{NO2} of 37.42% and 65.80%, respectively. This study will assist the policymakers in devising a pathway to curb down air pollutant concentration in various urban cities by utilising the benchmark levels of air pollution.

Keywords Air quality index · China · COVID 19 · India · NO_2 · $PM_{2.5}$

Introduction

In the present time, with the emergence of rapid globalisation and urbanisation, megacities in developing nations are facing severe health issues due to ambient air pollution. According to WHO (World Health Organization), seven million people die each year because of exposure to polluted air (UN Environment Programme 2018). Numerous epidemiological studies in the past two decades have highlighted outdoor air pollution as a cause of various respiratory diseases such as asthma, premature deaths and cardiovascular diseases. These have been identified as primary causes of mortality. In such cases, the population living in the vicinity of major roadways in metropolitan cities suffers the most (Park et al. 2020). In urban areas, 80% of people live in concentrations exceeding the WHO limits (Błaszczyk et al. 2017). Motor-vehicle emitted compounds in urban areas which include carbon monoxide (CO); nitrogen oxides (NO_x); coarse (PM_{10}), fine ($PM_{2.5}$), and ultrafine ($PM_{0.1}$) particle mass, black carbon, polycyclic aromatic hydrocarbons and benzene which are found in elevated concentrations as reported by Venkatram and Schulte (2018). Moreover, studies suggest that particulate matter and NO_2 levels are higher in cities with greater transportation activity and urban backgrounds (Rodríguez et al. 2016). PM_{10} and $PM_{2.5}$ are the two primary particulate matters monitored all over the world. However, $PM_{2.5}$ possesses a higher health risk as compared with PM_{10} because of its high retention time and ability to penetrate deep into the lungs and enter the bloodstream (US EPA 2018). The WHO ambient air quality guidelines suggest an annual mean $PM_{2.5}$ concentration limit of 10 μg/m^3 and 25 μg/m^3 for the 24-hourly mean. The NO_2

✉ Rajeev Kumar Mishra
rajeevkumarmishra@dtu.ac.in

Aviral Agarwal
agarwalaviral1999@gmail.com

Aman Kaushik
kakukaushik@gmail.com

Sankalp Kumar
sankalp.vibhu@gmail.com

[1] Department of Environmental Engineering, Delhi Technological University, Delhi 110042, India

Published online: 23 July 2020

limit is 40 μg/m^3, and 200 μg/m^3 for the annual and 1-h mean, respectively (World Health Organization 2005).

At the dawn of twenty-first century, developing economies like India and China are undergoing rapid industrialisation and modernisation, which are leading to hazardous levels of air pollution similar to the Industrial Revolution in Europe. It is well-understood that megacities in both countries like Beijing, Shenyang, Taiyuan, New Delhi, Mumbai, and Chennai are the world's most polluted cities (Zhu 2005). The primary sources of air pollution in India have been identified as vehicular emissions, industrial emissions, coal combustion, biomass burning, road dust, and refuse burning (Pant and Harrison 2012). Likewise, the poor air quality in China is a matter of global concern. The air pollution caused by transportation and industries is a serious environmental issue in urban settlements, and 50% of the PM in the urban air comes from traffic emissions (Li et al. 2017). Kumar and Joseph (2006) analysed ambient and kerb site air pollution correlation of PM_{10}, $PM_{2.5}$, and NO_2 in Mumbai, India. The results indicated a strong correlation between $PM_{2.5}$ and NO_2 at the ambient site due to vehicular emission as a result of high traffic density. A 2016 report estimated that only 3% of the Chinese population and less than 1% of the Indian population have exposure to $PM_{2.5}$ concentrations complying with WHO guidelines (IEA OECD 2016), although, an overall monotonic decrease in air pollutants was recorded in China from 2015 to 2018 (Fan et al. 2020). The existing levels of ambient $PM_{2.5}$ and NO_2 are above the safe limit. A 2017 report on global air pollution stated that China and India contribute to 52% of global $PM_{2.5}$—attributable deaths (1.525 million deaths) (Health Effects Institute 2019). In the past decades, both countries have been pro-active towards the efforts reducing air pollution. However, no long-term solution has been identified yet. Numerous academic studies have been conducted in both nations regarding the growing air pollution and its health effects. Kumar and Mishra (2018) conducted an assessment of major air pollutants at 36 transport corridors in Delhi, India, and the results of the study concluded that 31 corridors had "severe" and "very poor" AQI, and high traffic volume in most corridors is characterised by traffic-induced human health risks. It has been found that high levels of ambient $PM_{2.5}$ and NO_2 increase the risk of cardiovascular diseases and lung cancer in humans (Liu et al. 2018; Siddique et al. 2010).

In late December 2019, there was an outbreak of a highly contagious disease caused by the novel coronavirus, SARS-CoV-2. The first case emerged from Wuhan City, Hubei Province, China. The disease has been identified as Coronavirus disease (COVID-19). It's an outbreak, recognised as a "Pandemic" by WHO, has been extensively worldwide and exponential with more than 200 countries and territories reporting 3,267,936 cases and 234,703 deaths (7.18%) as of April 30, 2020 (WHO 2020). Individuals with underlying health problems, weak immunity, and the elderly are most likely to become extreme cases (Chen et al. 2020a). The critical sources of infection are patients infected with the novel coronavirus and those with asymptomatic infection (Wang et al. 2020a). Studies indicate correlation between the long-term exposure to air pollutant and COVID 19 death rate. Cities with hazardous air quality face a serious threat from the pandemic (Wu et al. 2020b; Conticini et al. 2020). Therefore, in the absence of a vaccine or treatment available for COVID-19, there has been a coordinated global response of imposing "lockdown" measures on citizens. As of now, more than a third of the worldwide population is under restriction. India recorded 34,867 COVID-19 cases (as of 30-04-2020) and a nationwide lockdown was imposed in India on March 25, 2020 for 21 days (MoHFW 2020). The lockdown constrained people from stepping out of their homes. Transport services, road, air, and rail, were suspended along with institutions and industrial establishments except for essential goods and services (Jain and Sharma 2020) and has been extended up to May 3, 2020. In parallel, 82,862 cases (as of 30–04-20) were recorded in China and 82% of these cases were recorded in Hubei Province (National Health Commission of the People's Republic of China 2020).

In Hubei province, Wuhan was regarded as the epicentre (Zhang et al. 2020) of the virus. On January 23, 2020, Hubei province, China, was kept under community quarantine with the shutdown of public transport, educational institutes, business centres, parks, and other social contacts to slow down the spread of COVID-19 (Wilder-Smith and Freedman 2020).

Further restrictions in Hubei province were lifted on March 23, 2020.With the implementation of lockdown and other federal restrictions in various countries around the globe, a visible reduction in air pollution is found in megacities. This study is aimed at quantifying and analysing the reduction in air pollution due to the lockdown imposed in two overpopulated and highly polluted countries of the world, viz. China and India, to determine the effect of lockdown on the air quality in an urban environment. The results of this study will help in gauging the ability of a full lockdown on reducing air pollution. Further, it will help in devising a response plan for unforeseen episodes of the high level of air pollution in urban environments.

Methodology of the study

Site selection

For the present study, 6 cities have been selected, each from India and Hubei Province, China. These locations are selected based on the availability of historical air pollution data, population density, monitoring station network, and the number of positive COVID-19 cases per million people. Selected cites

with their population, the number of monitoring stations taken into account, COVID-19 cases per million people, their geographical coordinates, and start and end date of lockdown are given in Table 1.

Parameters for analysis

For the analysis of the effect of the lockdown imposed by the governing authorities on the air quality, $PM_{2.5}$ and NO_2 are selected as parameters of the study. Both of these pollutants have a direct relationship with various anthropogenic activities that were restricted due to the lockdown (US EPA 2018; Ministry for the Environment New Zealand 2020). Hence, analysing these parameters assist in espying the effect of lockdown on the air quality of the selected locations.

Data collection and interpretation

During the data collection, 24-h average concentration (μg/m^3) data is taken for $PM_{2.5}$, and hourly average concentration data is taken for NO_2 (ppb) from respective EPAs of the locations selected. The data is collected in China for 13 weeks starting from January 1, 2020 except for Wuhan for which data was collected for 15 weeks since the lockdown was imposed till April 8, 2020. In India, data is collected for 13 weeks starting from February 1, 2020 to April 30, 2020. For the years 2016–2019, the data for Hubei Province, China has been collected from January 1 to March 31, and similarly, for India, the data has been collected from February 1 to April 30.

Weekly average data of $PM_{2.5}$ and NO_2 for the mentioned months has been calculated for the selected monitoring stations. The average value of the $PM_{2.5}$ and NO_2 concentration in a city is calculated by taking an average of all the monitoring stations selected, located at various distant locations throughout the city. The average value of the concentration of $PM_{2.5}$ and NO_2 is converted to individual AQI ($AQI_{PM\ 2.5}$ and AQI_{NO2}) by using the protocol suggested by US EPA for reporting the air quality data using the Air Quality Index (AQI) (Mintz 2012).

To analyse the changes in the $AQI_{PM2.5}$ and AQI_{NO2} levels, for each city, various drop percentages are calculated. The immediate drop percentage is calculated by the difference in average $AQI_{PM2.5}$ and AQI_{NO2} of the weeks before and after the lockdown was enforced. The final $AQI_{PM2.5}$ and AQI_{NO2} drop percentages are calculated by the difference in average $AQI_{PM2.5}$ and AQI_{NO2} of the week before lockdown and the last week when lockdown restrictions were lifted. Five-year and 1-year $AQI_{PM2.5}$ drop percentages in the year 2020 are calculated for January, February, and March in China and February, March, and April in India. It has been calculated by the difference in average $AQI_{PM2.5}$ of months of years 2016 and 2019 to the same months of 2020 for 5-year and 1-year drop percentages, respectively.

Results and discussion

With the parameters of immediate and final $AQI_{PM2.5}$ and AQI_{NO2} drop percentages, and 5 year and I year drop percentages of $AQI_{PM2.5}$, analysis for the cities in China and India has been done followed by a comparative assessment between the two countries.

Air quality analysis for the selected cities of China

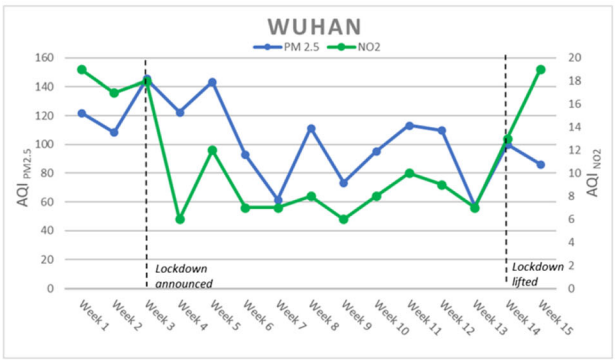
Due to a large number of reported cases, Wuhan and its neighbouring cities (Huanggang and Ezhou) implemented a lockdown on January 23, 2020 followed by several cities on

Table 1 General Information of the Selected Cities (Office of the Registrar General and Census Commissioner 2011; National Bureau of Statistics of China 2010; Central Intelligence Agency 2018; MoHFW 2020; National Health Commission of the People's Republic of China 2020)

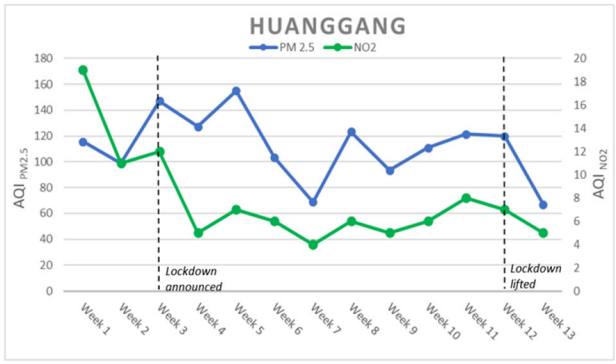
Country	City	Population (in million)	No. of stations monitored	COVID-19 cases per Million	Geographical coordinates
China	Xiangyang	5.89	5	199.22	30° 48′ 01″ N 110°23′11″ E
	Jingzhou	0.97	3	1624.47	30°13′35" N 111° 47′ 18″ E
	Huanggang	6.628	2	438.55	30° 24′ 16″ N 114° 42′ 49″ E
	Xiaogan	5.17	3	679.26	31° 03′ 41″ N 113° 25′ 37″ E
	Wuhan	8.11	5	6204.90	30° 42′ 07″ N 113° 46′ 52″ E
	Yichang	4.37	5	213.18	30° 41′ 49″ N 110° 48′ 01″ E
India	Delhi	18.62	5	501.03	29° 03′ 55″ N 76° 06′ 09″ E
	Lucknow	3.12	5	88.30	26° 52′ 53″ N 80° 41′ 49″ E
	Kolkata	4.98	5	240.51	23° 04′ 44″ N 87° 17′ 22″ E
	Mumbai	13.80	5	1279.98	19° 10′ 38″ N 72° 23′ 50″ E
	Chennai	5.16	5	1153.44	13° 07′ 10″ N 79° 44′ 05″ E
	Jaipur	6.42	4	234.51	27° 03′ 33″ N 75° 18′ 19″ E

January 24, 2020 (Wu et al. 2020a). Wuhan, having the highest number of cases than any other city in China enforced lockdown with strict federal orders restraining anthropogenic activity to minimal level (Lu 2020), which entailed the highest immediate AQI_{NO2} (69.35%) and $AQI_{PM2.5}$ drop (15.95%) among the selected cities. In contrast, Xiangyang recorded the lowest immediate and final AQI_{NO2} drop of − 3.22% and − 16.40%, respectively. This trend is observed since Xiangyang is one of the most industrialised cities in central China, and due to the high demand for PPE kits and testing equipment, industries were working at double shifts to meet the demand (Hubei Provincial People's Government 2020).

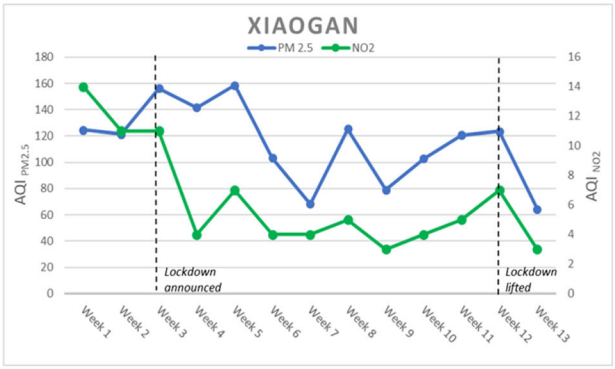
The average immediate and final AQI_{NO2} drop recorded is 48.61% and 26.64%, respectively. It can be concluded that the final AQI_{NO2} drop percentage is lower than the immediate AQI_{NO2} drop percentage; as the lockdown progressed, citizens had to get out of their homes for necessary essential commodities. Every city experienced a drop in $AQI_{PM2.5}$ (Fig. 1a, b, c, d, e, and f); on an average, the immediate $AQI_{PM2.5}$ drop is 11.28%; subsequently, the average final $AQI_{PM2.5}$ drop is 26.37%. Before lockdown was implemented, every city had $AQI_{PM2.5}$ within unhealthy for sensitive groups (101–150) and unhealthy (151–200) range according to US EPA standards. Nevertheless, the $AQI_{PM2.5}$ level



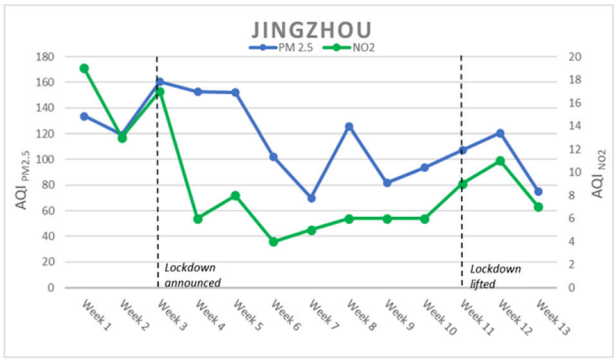
a



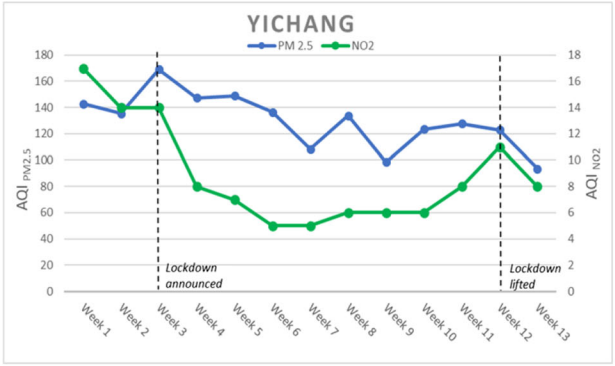
b



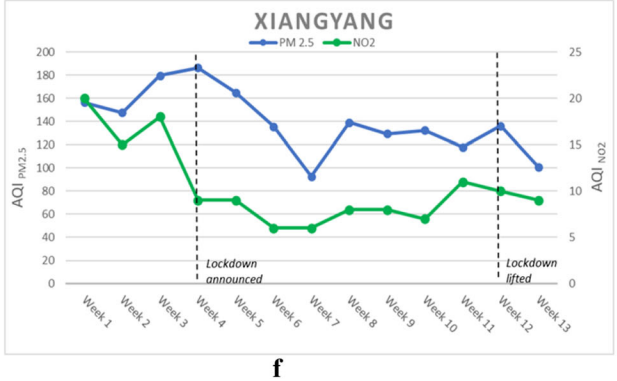
c



d



e



f

Fig. 1 Weekly averages of $AQI_{PM2.5}$ and AQI_{NO2} for cities of China

reduced and reached the moderate category during the lockdown period. Furthermore, for January, February, and March 2020, the cities recorded the lowest $AQI_{PM2.5}$ levels in 5 years (Fig. 2a, b, c, d, e, and f). As the virus started to spread in Hubei province in January, the citizens avoided leaving their homes as a self-precautionary measure. As the lockdown was implemented from January 23, 2020 in various cities, a drastic drop has been recorded in $AQI_{PM2.5}$ levels

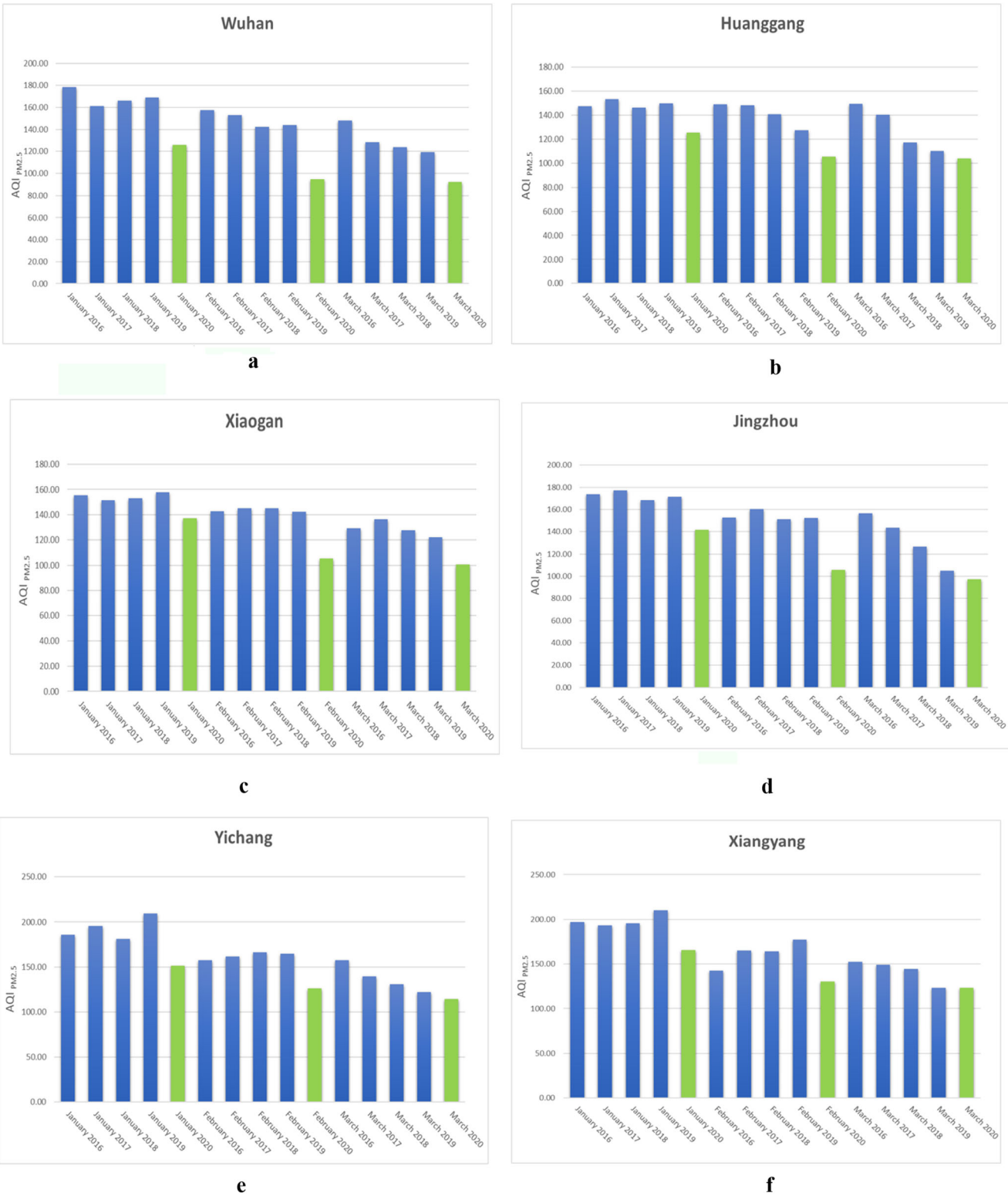


Fig. 2 Past 5 years' averages of $AQI_{PM2.5}$ for cities of China

from week 3 to week 4 (Fig. 1a, b, c, d, e, and f). The $AQI_{PM2.5}$ level remained constant through 2016–2019, to a certain degree for January and February. The highest 5-year and 1-year $AQI_{PM2.5}$ drop were recorded in Wuhan: 39.96% and 34.21%, respectively, while the lowest was found in Xiangyang: 8.53% and 26.55%, respectively. The average 5-year $AQI_{PM2.5}$ drop for January, February, and March, are 18.14%, 25.74%, and 29.18%, respectively. Subsequently, the average 1-year $AQI_{PM2.5}$ drop for January, February, and March are 20.17%, 26.31%, and 9.97%, respectively. It can be recorded that the 5-year and 1-year $AQI_{PM2.5}$ drop percentages are comparable in all months except March.

Additionally, meteorological conditions have an essential influence on the variations of $PM_{2.5}$ and NO_2 concentrations in the ambient environment (Agarwal et al. 2006). The cities Wuhan, Huanggang, and Xiaogan recorded a spike of $AQI_{PM2.5}$ and AQI_{NO2} levels (Fig. 1a, b, and c) in week 5 which led to a gradual decline in the forthcoming weeks of $AQI_{PM2.5}$ and AQI_{NO2}. The increase in the AQI levels was due to low precipitation recorded in week 5 in the three cities; total average precipitation for the three cities for week 4 and 5 were 42.6 mm and 0.34mm, respectively. In the subsequent weeks, rainfall intensity increased, which led to the drop in $AQI_{PM2.5}$ and AQI_{NO2} levels; the total average rainfall in weeks 6 and 7 for the three cities were 19.94 mm and 22.07 mm. In week 8, it has been observed from the Fig. 1a, b, c, d, e, and f that there is an increase in $AQI_{PM2.5}$ and AQI_{NO2} levels. Wuhan experienced 4.2mm precipitation in Week 8, as compared with 28.6mm and 39.4mm precipitation in Week 7 and Week 9, respectively. Hence, the abrupt increase in $PM_{2.5}$ and NO_2 in the cities is due to the low precipitation received in central China in Week 8 (19th–25th, February 2020). In the 13th week, a sharp decrease in $AQI_{PM2.5}$ and AQI_{NO2} levels is recorded. The mean rainfall in six cities in the 12th week is 14.23 mm. In contrast, the 13th week recorded heavy precipitation in all cities with mean average rainfall as 66.36 mm.

Air quality analysis for the selected cities of India

India enforced a nationwide lockdown from March 24, 2020, to May 3, 2020, after successive extensions as a preventive measure against COVID-19 pandemic. As the lockdown was implemented from March 24, 2020, a drastic drop has been recorded in $AQI_{PM2.5}$ levels of all selected cities from week 7 to week 8 (Fig. 3a, b, c, d, e, and f). Maharashtra, western peninsular state of India, has recorded the most cases of COVID-19 and deaths, 12,296 and 521, respectively (MoHFW 2020), and among the six megacities of India. Its capital, Mumbai, has shown the highest immediate drop of both AQI_{NO2} and $AQI_{PM2.5}$, i.e. 76.28% and 34.02%, respectively. Kolkata recorded the highest final $AQI_{PM2.5}$ drop (76.67%), and Lucknow recorded the least immediate drop in $AQI_{PM2.5}$ (6.47%) partly due to negligible precipitation in week 8. Chennai experienced the least immediate and final drops of AQI_{NO2}, which are 32.14% and 20.95%, respectively. It can be understood because of Chennai having an already low value of AQI_{NO2} (Fig. 3a) in week 7 and the weeks before the lockdown (Table 2).

The 6 cities experienced an average immediate $AQI_{PM2.5}$ drop of 20.21%, and an average final $AQI_{PM2.5}$ drop of 37.42%. Each one of the six Indian cities in the study recorded an immediate and final AQI_{NO2} drop with the average immediate AQI_{NO2} drop of 59.26% and an average final AQI_{NO2} drop of 65.80%. It has shown an overall drop in both $AQI_{PM2.5}$ and AQI_{NO2}. The average 5-year $AQI_{PM2.5}$ drop for the months of February, March, and April is recorded as 16.05%, 26.68%, and 37.51%, respectively; subsequently, average 1-year $AQI_{PM2.5}$ drop for the months of February, March, and April is 3.48%, 17.98%, and 27.06%, respectively. Chennai recorded the highest 5-year drop and 1-year drop in $AQI_{PM2.5}$ in April 2020 as 59.79% and 42.90%, respectively. All cities, except Mumbai in April 2020 and Chennai in March 2020, recorded the lowest $AQI_{PM2.5}$ levels in March and April 2020 as compared with the past 5 years. Mumbai is the only Indian city in the study to have shown a 1-year rise (1.09%) in an average $AQI_{PM2.5}$ in April 2020 (Fig. 4a, b, c, d, e, and f). A spike in $AQI_{PM2.5}$ was recorded (Table 3) between weeks 9 and 10 in New Delhi. It is as a result of reported fireworks incidents recorded on April 5, 2020, the day-wise $AQI_{PM2.5}$ levels of week 9 and 10 are given in Fig. 3b (The Indian Express 2020).

Furthermore, meteorological factors have an essential factor in the reduction and increase of $PM_{2.5}$ and NO_2 concentrations in the ambient environment. It can be recorded from Fig. 3c, in Kolkata, that $AQI_{PM2.5}$ and AQI_{NO2} have decreased augmented by heavy precipitation in the weeks 11, 12, and 13 with 29.20 mm, 80.90 mm, and 60.20 mm, respectively. However, in the preceding weeks 9 and 10, there was no precipitation. On the other hand, New Delhi, Lucknow, and Jaipur recorded an increase in $AQI_{PM2.5}$ in week 11 due to high surface winds in northern India due to dust storms from western India according to the Ministry of Environment and System of Air Quality and Weather Forecasting and Research (SAFAR) (Fig. 3b, e, and f) (ANI News 2020).

Comparative analysis and discussion between China vs India

An entire month lockdown was implemented in February 2020 in Hubei province, China; likewise in India, the month of lockdown was April 2020. The 1-year drop for February in China comes out to be 26.31%, whereas, for April in India, it is 26.06%. Hence, it can be deduced from the results that an entire month lockdown in urban centres results in a drop of around 26% in $AQI_{PM2.5}$ if compared with previous year



Fig. 3 Weekly averages of $AQI_{PM2.5}$ and AQI_{NO2} for cities of India

Table 2 Chennai weekly AQI_{NO2} averages (CPCB-CCR 2020; Mintz 2012)

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13
AQI_{NO2}	4	5	4	4	4	3	3	2	3	4	3	3	2

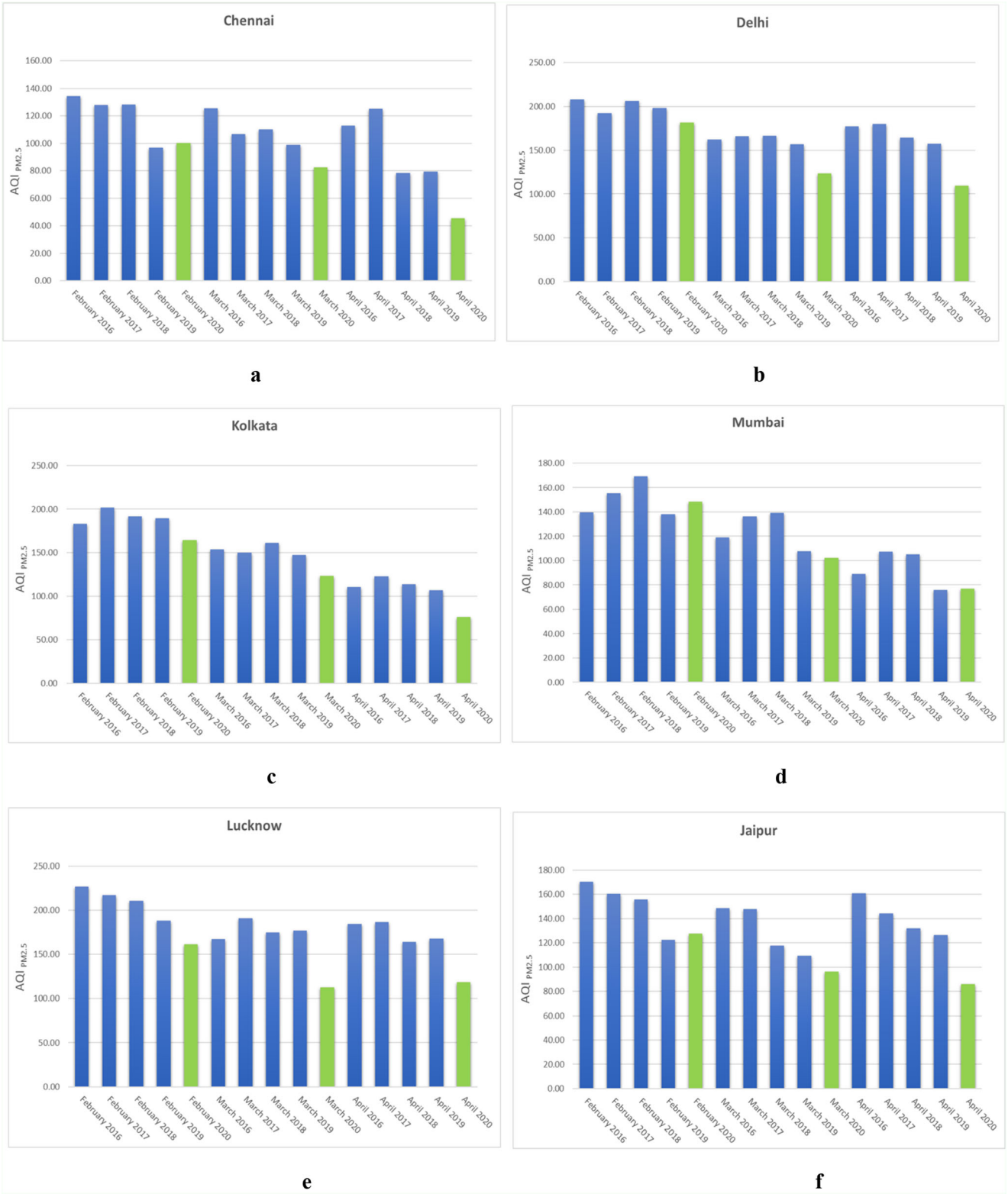


Fig. 4 Past 5 years' averages of $AQI_{PM2.5}$ for cities of India

value. Several academic studies have been conducted to study the relationship between the local meteorological factors and concentration of various pollutants (viz. $PM_{2.5}$ and NO_2) (Guo et al. 2017). The prime factors which influence the concentration of $PM_{2.5}$ and NO_2 have been identified as precipitation, ambient temperature, wind speed, and relative humidity.

Table 3 Delhi daily AQI_{NO2} and $AQI_{PM2.5}$ averages (CPCB-CCR 2020; Mintz 2012)

	April 1, 2020	April 2, 2020	April 3, 2020	April 4, 2020	April 5, 2020	April 6, 2020	April 7, 2020	April 8, 2020	April 9, 2020	April 10, 2020	April 11, 2020	April 12, 2020	April 13, 2020
AQI_{NO2}	7	9	10.75	12.33	11	8	9.5	11	13	11.67	8.33	12	12.33
$AQI_{PM2.5}$	92.75	77.2	88.2	97.75	116	137	94.67	90.67	93.5	124.25	122.25	113	128.75

However, the 12 selected cities tend to show a tremendous reduction in the concentration of $PM_{2.5}$ and NO_2 due to lockdown enforcement. It reflects that cutting down on anthropogenic sources of various pollutants can be useful in reducing the AQI.

Huanggang, China, has a population of 6.62 million, and the $AQI_{PM2.5}$ and AQI_{NO2} of Huanggang before the lockdown were 147.1 and 12, respectively. After the first week of implementation of lockdown, there was found a decrease of 13.40% and 58.54% in $AQI_{PM2.5}$ and AQI_{NO2}. Likewise, Jaipur, India, has a population of 6.42 million and the $AQI_{PM2.5}$ and AQI_{NO2} of Jaipur before the lockdown was implemented were 115.7 and 15, respectively. After the first week of the implementation of lockdown, a decrease of 27.50% and 60.37% in $AQI_{PM2.5}$ and AQI_{NO2} was recorded. Both cities have a comparable population and immediate AQI_{NO2} drop, but the immediate $AQI_{PM2.5}$ drop differs by 14.1%. The lowest week average $AQI_{PM2.5}$ recorded in Huanggang and Jaipur was recorded 68.7 and 73.4, respectively. Both the cities' $AQI_{PM2.5}$ has dropped down from unhealthy for sensitive groups to moderate air quality. Within 4 weeks of implementation of lockdown, Huanggang experienced a drop of 53.30% and 65.85% in week average $AQI_{PM2.5}$ and AQI_{NO2}. Furthermore, Jaipur saw a drop of 36.56% and 67.49% in week average $AQI_{PM2.5}$ and AQI_{NO2} in just 2 weeks.

For a holistic view, the cities selected in India recorded an average immediate $AQI_{PM2.5}$ and AQI_{NO2} drop of 20.21% and 59.26%, respectively. In contrast, on the other hand, the cities in China recorded an average drop of 11.32% and 48.61%, respectively. After 6 weeks of implementing the lockdown, cities in India recorded an average drop in $AQI_{PM2.5}$ and AQI_{NO2} of 37.42% and 65.80%, respectively, while cities in China recorded a drop of 42.54% and 56.67% respectively. From these results, it can be inferred that the drop in $PM_{2.5}$ is rather gradual as compared with the sudden drop in NO_2 concentrations throughout the cities.

As shown in Fig. 5, the drop in $AQI_{PM2.5}$ of coastal cities (viz. Chennai, Mumbai, and Kolkata) is relatively more significant than inland cities. The exceptional drop in $AQI_{PM2.5}$ in the coastal cities is vastly due to the coastal winds which are very prominent in these cities. Previous studies conducted in these coastal cities show that coastal regions show a significant drop in $PM_{2.5}$ in the morning as compared with inland regions (Chen et al. 2020b; Gupta et al. 2004). The three coastal cities incorporated in the present study record an average immediate $AQI_{PM2.5}$ drop of 24.96%, and the final average $AQI_{PM2.5}$ drop was found as 54.90%. On the other hand,

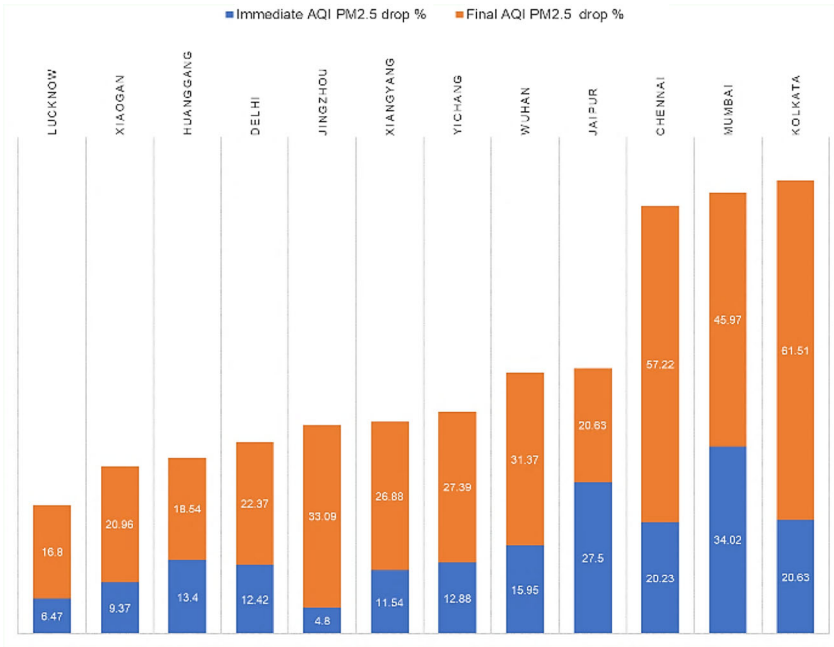


Fig. 5 Immediate and final Drop percentages of $AQI_{PM2.5}$ for all cities

the inland region cities recorded an average immediate $AQI_{PM2.5}$ drop of 12.70% and the final average $AQI_{PM2.5}$ drop of 24.23%. The drop in average immediate and final $AQI_{PM2.5}$ of cities in inland regions is moderate as compared with the drop in coastal regions.

Drop-in AQI_{NO2} also shows similar trends as $AQI_{PM2.5}$ (Fig. 6). The coastal cities recorded a much higher percentage of drop in AQI_{NO2} as compared with cities in inland regions. The coastal cities, except for Chennai, show an exceptionally high drop in AQI_{NO2}. Mumbai and Kolkata recorded an immediate AQI_{NO2} drop of 76.28% and 55.70%, respectively, whereas the final AQI_{NO2} drop is 92.58% and 76.67%, respectively. However, Chennai is a coastal city that recorded a much lower percentage drop in AQI_{NO2} as compared with the other two coastal cities, AQI_{NO2} of Chennai was already at a record low values between 3 to 5 before even lockdown was implemented. Xiangyang recorded an increase in AQI_{NO2} levels after the implementation of lockdown. Xiangyang, being a heavily industrialised city, had industries that were operating during the lockdown to produce essential medical equipment.

Conclusion

Both the nations followed different protocols for implementing lockdown in each country, although the lockdown in both of the nations was found effective in declining the rate of spread of COVID-19 cases (Wang et al. 2020b; Barkur et al. 2020), and it played a significant role in reducing the air pollution to record low values. The significant findings of the study are as follows:

- In China, the week before the lockdown was enforced, 4 out of 6 cities had an $AQI_{PM2.5}$ in the unhealthy category. Wuhan and Huanggang were found in unhealthy for sensitive group category. In the last week of lockdown, 5 out of 6 cities were found to be in unhealthy for sensitive group category except for Wuhan, which was found in the moderate category.
- In India, the week before the lockdown was enforced, 5 out of 6 cities had an $AQI_{PM2.5}$ that is unhealthy for sensitive group category except Chennai that was found under the moderate category. In the sixth week of lockdown, all cities were found in the good and moderate category except Delhi and Lucknow, which were found to be in unhealthy for sensitive group category.
- Meteorological factors are an essential factor in order to address pollutant concentration in ambient environment. Henceforth, meteorological should be taken into account before the execution of a response plan to mitigate pollution in urban cities around the world.
- For all 12 cities, a gradual decline has been recorded in $AQI_{PM2.5}$ levels in subsequent lockdown weeks. The mean immediate and final $AQI_{PM2.5}$ drops are 15.76% and 31.89%, respectively. However, in the case, AQI_{NO2} levels, a sharp decline has been recorded in the first week of lockdown. The mean immediate and final AQI_{NO2} drops are 53.93% and 46.22%, respectively.
- The coastal cities (viz. Chennai, Mumbai, and Kolkata) recorded a more significant decline in $AQI_{PM2.5}$ and AQI_{NO2} as compared with the other inland region cities.

The lockdown implemented in various regions around the world provided us with a unique opportunity to identify the

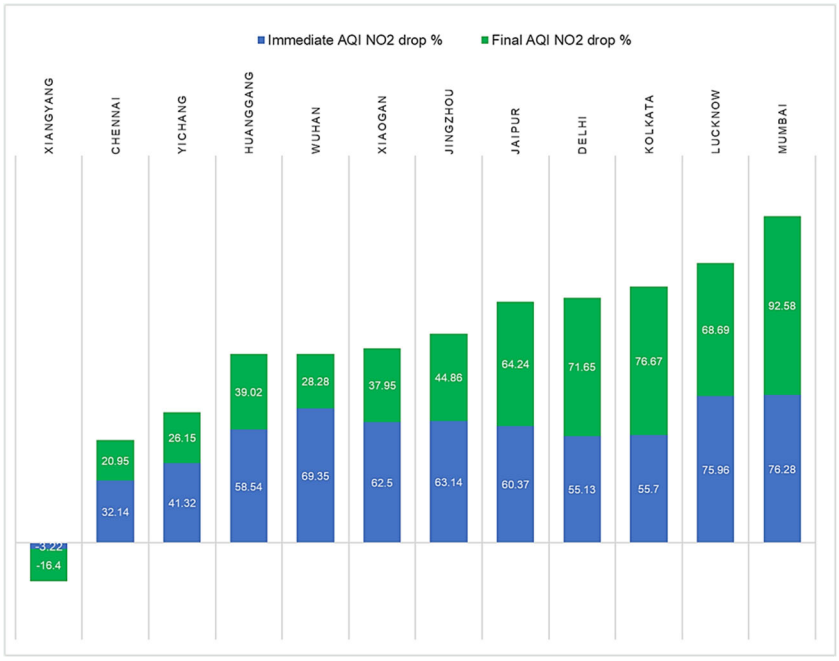


Fig. 6 Immediate and final Drop percentages of AQI_{NO2} for all cities

benchmark levels of pollutants in various urban cities around the world. The findings of the study will assist the governing authorities and policymakers to calibrate a proper response plan to bring down the ever-increasing pollution levels in various developing urban regions across the globe, especially China and India.

Acknowledgements Authors are incredibly grateful to Advance Air & Acoustics Research Laboratory, Delhi Technological University, Delhi (India), for encouraging to conduct this research and providing all the facility to compile this work.

Funding information This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of data and material The data used in the current study have been taken from respective EPA and is available in public domain.

Compliance with ethical standards

Competing interests The authors declare that they have no known competing financial interest or personal relationships that would have appeared to influence the work reported in this paper.

References

Agarwal R, Jayaraman G, Anand S, Marimuthu P (2006) Assessing respiratory morbidity through pollution status and meteorological conditions for Delhi. Environ Monit Assess 114:489–504. https://doi.org/10.1007/s10661-006-4935-3

ANI News (2020) Delhi's air quality may deteriorate due to western dust storms: SAFAR. https://www.aninews.in/news/environment/delhis-air-quality-may-deteriorate-due-to-western-dust-storms-safar20200415085141/. Accessed 1 Apr 2020

Barkur G, Vibha, Kamath GB (2020) Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: evidence from India. Asian J Psychiatr 51:102089. https://doi.org/10.1016/j.ajp.2020.102089

Błaszczyk E, Rogula-Kozłowska W, Klejnowski K, Kubiesa P, Fulara I, Mielżyńska-Švach D (2017) Indoor air quality in urban and rural kindergartens: short-term studies in Silesia, Poland. Air Qual Atmos Health 10:1207–1220. https://doi.org/10.1007/s11869-017-0505-9

Central Intelligence Agency (2018) The World Factbook. https://www.cia.gov/library/publications/the-world-factbook/fields/344rank.html. Accessed 18 Apr 2020

Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y, Xia J, Yu T, Zhang X, Zhang L (2020a) Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 395:507–513. https://doi.org/10.1016/s0140-6736(20)30211-7

Chen Y, Wild O, Conibear L, Ran L, He J, Wang L, Wang Y (2020b) Local characteristics of and exposure to fine particulate matter (PM2.5) in four Indian megacities. Atmos Environ X 5:100052. https://doi.org/10.1016/j.aeaoa.2019.100052

Conticini E, Frediani B, Caro D (2020) Can atmospheric pollution be considered a co-factor in extremely high level of SARS-CoV-2 lethality in northern Italy? Environ Pollut 261:114465. https://doi.org/10.1016/j.envpol.2020.114465

CPCB-CCR (2020) Central control room for air quality management – all India. https://app.cpcbccr.com/ccr/#/caaqm-dashboard-all/caaqm-landing/data. Accessed 16 Apr 2020

Fan H, Zhao C, Yang Y (2020) A comprehensive analysis of the spatio-temporal variation of urban air pollution in China during 2014–2018. Atmos Environ 220:117066. https://doi.org/10.1016/j.atmosenv.2019.117066

Guo J, Xia F, Zhang Y, Liu H, Li J, Lou M, He J, Yan Y, Wang F, Min M, Zhai P (2017) Impact of diurnal variability and meteorological factors on the PM2.5 - AOD relationship: implications for PM2.5 remote sensing. Environ Pollut 221:94–104. https://doi.org/10.1016/j.envpol.2016.11.043

Gupta A, Patil RS, Gupta S (2004) Influence of meteorological factors on air pollution concentration for a coastal region in India. Int J Environ Pollut 21:253. https://doi.org/10.1504/ijep.2004.004193

Health Effects Institute (2019) State of Global Air 2019. Special Report. Health Effects Institute, Boston

Hubei Provincial People's Government (2020) Hubei Enterprises. http://en.hubei.gov.cn/business/enterprises/. Accessed 3 Apr 2020

IEA OECD (2016) Energy and Air Pollution: World Energy Outlook Special Report 2016

Jain S, Sharma T (2020) Social and travel lockdown impact considering coronavirus disease (COVID-19) on air quality in megacities of India: present benefits, future challenges and way forward. Aerosol Air Qual Res 20:1222–1236. https://doi.org/10.4209/aaqr.2020.04.0171

Kumar R, Joseph AE (2006) Air pollution concentrations of PM2.5, PM10 and NO2 at ambient and Kerbsite and their correlation in Metro City – Mumbai. Environ Monit Assess 119:191–199. https://doi.org/10.1007/s10661-005-9022-7

Kumar A, Mishra RK (2018) Human health risk assessment of major air pollutants at transport corridors of Delhi, India. J Transp Health 10:132–143. https://doi.org/10.1016/j.jth.2018.05.013

Li S, Feng K, Li M (2017) Identifying the main contributors of air pollution in Beijing. J Clean Prod 163:S359–S365. https://doi.org/10.1016/j.jclepro.2015.10.127

Liu H, Tian Y, Song J, Cao Y, Xiang X, Huang C, Li M, Hu Y (2018) Effect of ambient air pollution on hospitalization for heart failure in 26 of Chinas largest cities. Am J Cardiol 121:628–633. https://doi.org/10.1016/j.amjcard.2017.11.039

Lu D (2020) Inside Wuhan's lockdown. New Sci 245:7. https://doi.org/10.1016/s0262-4079(20)30234-7

Ministry for the Environment New Zealand (2020) Nitrogen dioxide. https://www.mfe.govt.nz/air/specific-air-pollutants/nitrogen-dioxide. Accessed 5 Apr 2020

Mintz D (2012) Technical assistance document for the reporting of daily air quality – the air quality index. EPA-454/B-12-001

MoHFW, Ministry of Health and Family Welfare (2020) COVID-19 INDIA. https://www.mohfw.gov.in/. Accessed 2 Apr 2020

National Bureau of Statistics of China (2010) Census Data http://www.stats.gov.cn/english/Statisticaldata/CensusData/. Accessed 5 May 2020

National Health Commission of the People's Republic of China (2020) April 30: Daily briefing on novel coronavirus cases in China. http://en.nhc.gov.cn/2020-05/16/c_80207.htm. Accessed 30 Apr 2020

Office of the Registrar General & Census Commissioner, India (2011) 2011 Census Data. http://censusindia.gov.in/2011-Common/CensusData2011.html. Accessed 22 Apr 2020

Pant P, Harrison RM (2012) Critical review of receptor modelling for particulate matter: a case study of India. Atmos Environ 49:1–12. https://doi.org/10.1016/j.atmosenv.2011.11.060

Park S, Allen RJ, Lim CH (2020) A likely increase in fine particulate matter and premature mortality under future climate change. Air Qual Atmos Health 13:143–151. https://doi.org/10.1007/s11869-019-00785-7

Rodríguez MC, Dupont-Courtade L, Oueslati W (2016) Air pollution and urban structure linkages: evidence from European cities. Renew Sust Energ Rev 53:1–9. https://doi.org/10.1016/j.rser.2015.07.190

Siddique S, Ray MR, Lahiri T (2010) Effects of air pollution on the respiratory health of children: a study in the capital city of India. Air Qual Atmos Health 4:95–102. https://doi.org/10.1007/s11869-010-0079-2

The Indian Express (2020) Indian outrage on social media after people light fireworks, lead processions after lights off. https://indianexpress.com/article/trending/trending-in-india/primisniter-narendra-modi-9pm9minutes-coronavirus-6349647/. Accessed 5 May 2020

UN Environment Programme (2018) Air pollution: know your enemy. https://www.unenvironment.org/news-and-stories/story/air-pollution-know-your-enemy. Accessed 2 Apr 2020

US EPA (2018) Particulate Matter (PM) Basics. https://www.epa.gov/pm-pollution/particulate-matter-pm-basics. Accessed 12 Mar 2020

Venkatram A, Schulte N (2018) Urban transportation and air pollution. Elsevier Science Publishing Co Inc, s.l.

Wang P, Chen K, Zhu S, Wang P, Zhang H (2020a) Severe air pollution events not avoided by reduced anthropogenic activities during COVID-19 outbreak. Resour Conserv Recycl 158:104814. https://doi.org/10.1016/j.resconrec.2020.104814

Wang S-X, Wang Y, Lu Y-B, Li J-Y, Song Y-J, Nyamgerelt M, Wang X-X (2020b) Diagnosis and treatment of novel coronavirus pneumonia based on the theory of traditional Chinese medicine. Journal of Integrative Medicine 18:275–283. https://doi.org/10.1016/j.joim.2020.04.001

WHO (2020) Coronavirus Disease (COVID-19) Situation Reports. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports. Accessed 30 Apr 2020

Wilder-Smith A, Freedman DO (2020) Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak. J Travel Med 27. https://doi.org/10.1093/jtm/taaa020

World Health Organization (2005) WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulphur dioxide: global update 2005: summary of risk assessment (No. WHO/SDE/PHE/OEH/06.02)

Wu JT, Leung K, Leung GM (2020a) Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 395:689–697. https://doi.org/10.1016/s0140-6736(20)30260-9

Wu X, Nethery RC, Sabath BM, Braun D, Dominici F (2020b) Exposure to air pollution and COVID-19 mortality in the United States: a nationwide cross-sectional study. https://doi.org/10.1101/2020.04.05.20054502

Zhang L, Wang B, Zhou J, Kirkpatrick J, Xie M, Johri AM (2020) Bedside focused cardiac ultrasound in COVID-19 infection from the Wuhan epicenter: the role of cardiac point of care ultrasound (POCUS), Limited Transthoracic Echocardiography and Critical Care Echocardiography. J Am Soc Echocardiogr 33:676–682. https://doi.org/10.1016/j.echo.2020.04.004

Zhu Z (2005) Political economy of China and India: dealing with air pollution in the two booming economies. Historia Actual Online 7:123–132

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

J. Inst. Eng. India Ser. B
https://doi.org/10.1007/s40031-020-00472-7

ORIGINAL CONTRIBUTION

Control and Design of Seven-Level Cascaded Multilevel Converter for Transformerless Large-Scale Photovoltaic Integration

Jyoti Kulkarni[1] · **Narendra Kumar**[1] · **Bhim Singh**[2]

Received: 15 December 2019 / Accepted: 14 July 2020
© The Institution of Engineers (India) 2020

Abstract In this work, the design and control of seven-level H-bridge multilevel converter are carried out for the grid integration of large photovoltaic (PV) system without bulky line frequency transformer, as the multilevel converter provides a direct integration to the medium voltage grid. This system is a single-stage system to reduce cost and control complexity. Its control uses a decoupled current control with synchronous reference frame (SRF) and phase locked loop (PLL) with low switching frequency phase-shifted pulse width modulation (PWM) technique. It reduces the switching losses and hence improves the efficiency. The outer loop of control comprises of a separate maximum power point technique (MPPT) controller for each PV array to overcome the unpredictable irradiance level variations, ambient temperature, shading effect and other relevant factors. The improved perturb and observe (P&O) MPPT algorithm is used to track the maximum power from the PV array to give the accurate results. A significant reduction in total harmonic distortion (THD) in the grid current because of the increased number of steps in output AC voltage waveform of multilevel converter and consequently reduction in size of the filter component are achieved. Simulative experiments are carried out in MATLAB/Simulink, and the performance is validated in accordance with the IEEE-519 standard.

Keywords MPPT · PV generation · SRF-PLL · Phase-shifted PWM · VSC · THD

Introduction

The fuels used in conventional power plant are depleting and polluting in the nature. Hence, to get clean, ever available, reliable and environment-friendly electricity generation, the usage of renewable energy sources, like solar and wind, is increasing since last two decades. Lots of research are being done on power converters to interface these renewable sources to the grid, and multilevel converter has been perceived as gen-next technique to connect renewable energy sources with the grid. Multilevel converters are advantageous over the traditional two-level converter due to increased scalability, lower harmonics content, reduced filter components, low switching frequency, less voltage stress across switching device and an increased efficiency. Many multilevel topologies like diode clamped, flying capacitor and cascaded H-bridge are reported in the literature. Out of which, cascaded H-bridge (CHB) multilevel converter is the best suitable configuration for photovoltaic (PV) generation as multistring PV plant naturally provides the isolated DC source for each bridge of the CHB converter [1–5].

For grid interfacing of PV system, single-stage (requires only DC–AC converter) as well as double-stage PV systems (require both DC–DC converter and DC–AC converter) is reported in the literature [6–18]. This system is a single-stage H-bridge multilevel converter to eliminate DC–DC stage, for reducing the cost and also control complexity in the system.

As multilevel converter provides the direct connection to the medium voltage grid, hence eliminating the need

✉ Jyoti Kulkarni
jyotikulkarni0803@yahoo.com

[1] Department of Electrical Engineering, Delhi Technological University, Main Bawana Road, New Delhi, Delhi 110042, India

[2] Department of Electrical Engineering, Indian Institute of Technology, New Delhi, Delhi, India

Published online: 27 July 2020

of large size and heavyweight transformer for grid interfacing, only a low frequency transformer is sufficient for providing galvanic isolation to limit leakage current [13].

Extracting maximal power from the solar PV array is also a challenge. Researchers have tried to track the maximum power using various approaches known as "MPPT techniques." Different techniques for implementing the concept of MPPT have been developed and reported in the literature [19]. In this system, separate improved P&O algorithm is conceived due to good accuracy and deployed for tracking the maximum power from each PV panel [20, 21] and each controlled by separate controller to overcome the unpredictable irradiance level variations, ambient temperature, shading effect and other relevant factors [18].

Its control comprises the decoupled current controller with SRF-PLL and PWM control for the modulation. A number of multicarrier PWM techniques are discussed in [22]. The low switching frequency phase-shifted PWM technique is employed in this work, which reduces the switching losses and hence improves the efficiency. With an increase in the number of steps (corresponding to the number of levels achieved) in the AC voltage output waveform in the multilevel converter, the THD in output grid current waveform is reduced, which also contributes to reduced size filter component. Moreover, a low switching frequency used in given work produces low acoustic noise, which eliminates the need of high value of an inductive filter; hence, only a low-value inductive filter is sufficient for the filtrating. The reduction in size of filter further results in reduction in the cost of system [18, 22, 23].

Many methods for synchronized connection, i.e., "integration" of large solar PV plant to the grid, are reported in the literature [9–11]. In [11], multipulse converters are used in large solar PV grid-connected system to minimize harmonics in the voltage waveform resulting in good power quality, lower switching losses and good reliability in comparison with the two-level topologies, but the good designing of transformer and interfacing magnetic is required for proper operation of multipulse-based system. The authors in [10] have reported "a novel scheme suitable for high-voltage and large-capacity photovoltaic power stations."

Some topologies for grid integration of single-phase converter are discussed in [12–15]. Here, a 3-Φ, 1.48 MW, 3.3 kV and 50 Hz power–frequency grid-integrated solar photovoltaic plant interfaced through H-bridge cascaded seven-level multilevel converter with phase-shifted PWM operated at 500 Hz switching frequency is proposed. The performance of the system is analyzed on MATLAB/Simulink.

System Description

The system as shown in Fig. 1, is a 1.48 MW, 3.3 kV, 50 Hz, 3-Φ seven-level cascaded H-bridge (CHB) converter with low switching frequency of 500 Hz.

In cascaded multilevel converter, the number of levels is equal to (2 * number of H-bridge per phase + 1) [23], and hence, seven-level converter consists of nine H-bridges, three H-bridges in each phase and is fed by the separate PV array. A distinct MPPT algorithm is needed for each PV array.

Design of System

The system is designed for 1.48 MW power, 50 Hz frequency, 3.3 kV voltage, and it includes the design of PV array, the DC-link capacitor and DC-link voltage, other specifications of the system, which are given in Table 1.

Selection of V_{dc} Voltage

In case of L number of level converter, for each individual H-bridge connected to the PV array, the DC-link voltage (V_{dc}) is calculated as in [23],

$$V_{rms} = 0.612 * (L-1) * V_{dc} \qquad V_{dc} = 898.692\,\text{V}. \tag{1}$$

V_{dc} is taken as 905 V.

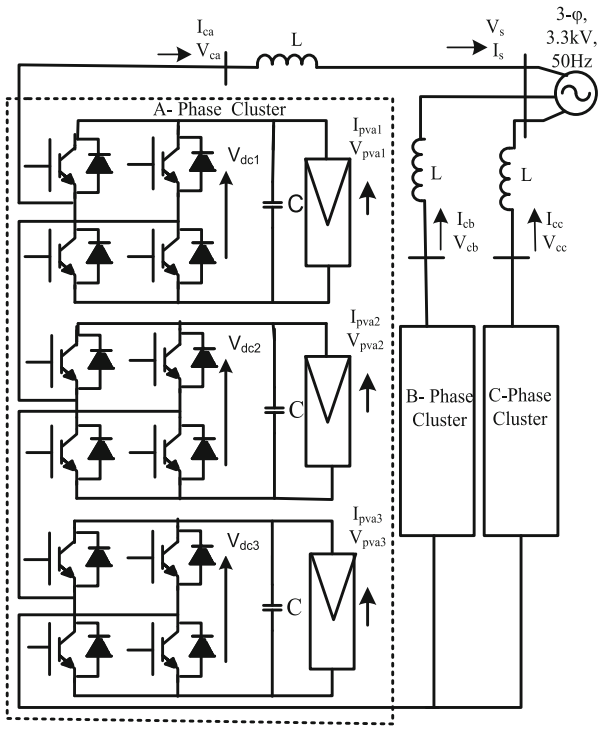


Fig. 1 Seven-level grid-interfaced converter PV system

Table 1 Component specifications

Component	Value
Power	1.48 MW
Voltage (V_{rms})	3.3 kV
Frequency	50 Hz
Switching frequency	500 Hz
Inductor (L)	5.4 mH
No of parallel paths in SPV array (N_p)	23
No of series paths in SPV array (N_s)	30
Voltage of PV array at maximum power (V_{mmp})	30.2 V
Current of PV array at maximum power (I_{mmp})	7.95 A
Open-circuit voltage of PV array (V_{oc})	37.5 V
Short-circuit current of PV array (I_{sc})	8.61 A
DC-link voltage (V_{dc})	905 V
DC-link capacitor (C_{dc})	14,500 μF

Design Calculations for Selection of PV Array

The required power rating of each PV array for a seven-level CHB converter consisting of nine PV array for 1.48 MW power grid is calculated as,

$$P_{mmp} = \frac{(\text{power rating of system})}{(\text{number of total PV array})} = 1.48\,\text{MW}/9 = 164.44\ \text{kW} \quad (2)$$

To obtain the PV array of 164.44 kW power (P_{mmp}) and V_{dc} of 905 V, a Sharp ND H230Q2 model is selected from PV array library, and numbers of PV modules in parallel and series are selected as,

$$N_s = V_{dc}/V_{mmp} = 30 \quad (3)$$

$$N_p = P_{mmp}/\left(N_s * I_{mmp} * V_{mmp}\right) = 23 \quad (4)$$

Selection of Capacitor for DC-Link

The DC-link voltage (V_{dc}) for each H-bridge connected to the PV array is obtained in accordance with the principle of conservation of energy for V_{dc} recovery in 5 ms and 1.2 overloading factor a, as [24],

$$\begin{aligned} {}^1\!/_2 * C_{dc}\left(V_{dc}^2 - V_{dc1}^2\right) &= K_1 * 3 * V * a * I * t \\ C_{dc} &= (0.1 * 164444 * 0.005)/ \\ &\quad (0.5 * (905 * 905 - 898.7 * 898.7)) \\ C_{dc} &= 14471.5\,\mu\text{F} \end{aligned} \quad (5)$$

It is selected as, $C_{dc}(C)$ = 14,500 μF.

Control Algorithm

The used algorithm for the control of the PV system connected to the three-phase AC grid is shown in Fig. 2. The maximum power developed by the PV array is tracked using an improved P&O algorithm. The decoupled current controller having a feedforward term, separate controller for each DC-link capacitor, SRF and PLL is used for controlling the voltage source converter (VSC) [25].

MPPT Algorithm

As shown in Fig. 3, by using an improved P&O MPPT algorithm, the maximum power is extracted from PV array and a suitable V_{dc_ref} is generated in this work. In this improved P&O MPPT algorithm, the detail of variation in current (ΔI_{pv}) is also considered along with the variation in voltage (ΔV_{pv}) and power (ΔP_{pv}) as in the P&O algorithm. In case of insolation change, normal P&O technique fails to differentiate between variation in the power on account of insolation variation and due to the perturbation, which is taken care of by improved P&O algorithm [20, 21].

Control of VSC

The voltage measured from each DC link is compared with the individual V_{dc_ref} given by each MPPT algorithm, as shown in Fig. 4. The error voltage (difference between the two) is processed by a proportional–integral (PI) controller to provide the active current loss component I^*_d, through the DC link, which is given as,

$$I_d^*(k) = I_d^*(k-1) + K_p\{V_e(k) - V_e(k-1)\} + K_i * V_e(k) \quad (6)$$

where V_e(k) and $V_e(k-1)$ are the error voltages at given sampling instants k & $(k-1)$ and K_p & K_i are proportional and integral gain constants. The sum of all reference DC-link currents along with the feedforward term is compared with the direct current component obtained by Park transform from the grid current as,

$$\begin{bmatrix} I_d \\ I_q \\ I_0 \end{bmatrix} = \frac{2}{3}\begin{bmatrix} \sin(\theta) & \sin\left(\theta - \frac{2\pi}{3}\right) & \sin\left(\theta + \frac{2\pi}{3}\right) \\ \cos & \cos\left(\theta - \frac{2\pi}{3}\right) & \cos\left(\theta + \frac{2\pi}{3}\right) \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}\begin{bmatrix} I_a \\ I_b \\ I_c \end{bmatrix} \quad (7)$$

The grid detection angle (θ) for Park transformation is obtained from three-phase PLL. I^*_q, the reference quadrature current (set to zero for feeding grid current at unity power factor) is compared with the quadrature current component obtained from Park transformation.

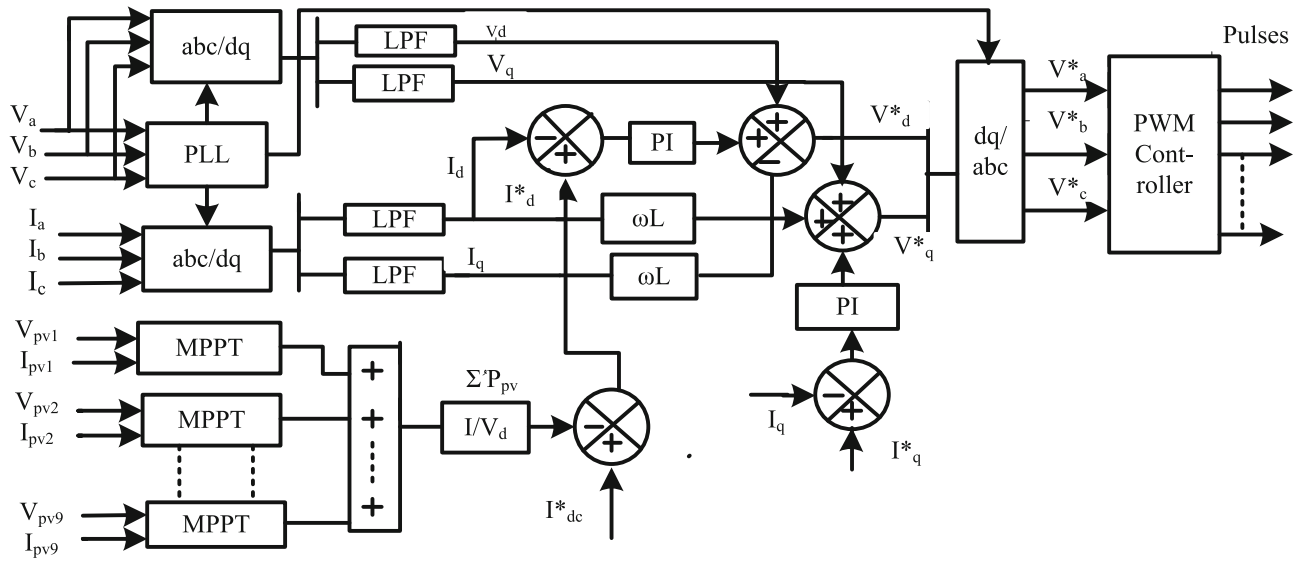


Fig. 2 Control algorithm

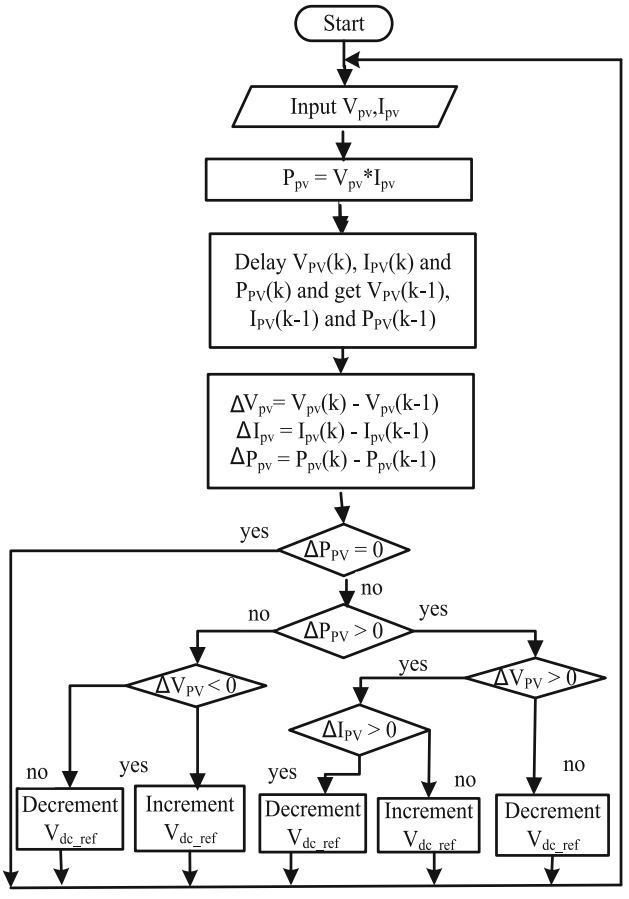


Fig. 3 Flowchart for improved P&O MPPT algorithm

Similarly, using Park transformation, direct and quadrature components of grid voltages (V_d & V_q) are calculated as follows.

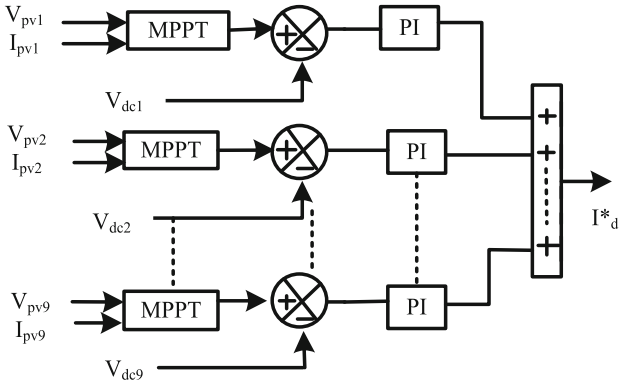


Fig. 4 Control algorithm for generating I^*_{dc}

$$\begin{bmatrix} V_d \\ V_q \\ V_0 \end{bmatrix} = \frac{2}{3} \begin{bmatrix} \sin(\theta) & \sin\left(\theta - \frac{2\pi}{3}\right) & \sin\left(\theta + \frac{2\pi}{3}\right) \\ \cos(\theta) & \cos\left(\theta - \frac{2\pi}{3}\right) & \cos\left(\theta + \frac{2\pi}{3}\right) \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} V_a \\ V_b \\ V_c \end{bmatrix} \tag{8}$$

The reference direct-axis voltage and quadrature-axis voltages are calculated as follows,

$$V_d^*(k) = V_d^*(k-1) + K_{p2}\{I_{de}(k) - I_{de}(k-1)\} + K_{i2} * I_{de}(k) + V_d - I_q^* \omega L \tag{9}$$

$$V_q^*(k) = V_q^*(k-1) + K_{p3}\{I_{qe}(k) - I_{qe}(k-1)\} + K_{i3} * I_{qe}(k) + V_q + I_d^* \omega L \tag{10}$$

where $I_{de} = (I^*_d - I_d)$, $I_{qe} = (I^*_q - I_q)$ and ωL is the coupling reactance. K_{p2}, K_{i2}, denote the proportional and

integral gains of direct-axis current controller and K_{p3} and K_{i3} are the proportional and integral gains of quadrature-axis current controller. The reference voltages (modulating signal) for the PWM controller are obtained using inverse Park transform.

In PWM controller, phase-shifted modulation technique is used, in which, for seven-level (L) converter, six (i.e., $(L-1)$), triangular carrier signals are generated [23]. The phase shift (φ_{cr}) between two adjacent carrier waves is given by

$$\begin{aligned}\varphi_{cr} &= 360^\circ/(L-1)\\ &= 360^\circ/6 = 60^\circ\end{aligned} \tag{11}$$

Each carrier (V_{cr1}–V_{cr6}) is compared with V_{ref} signal for generating pulses as shown in Fig. 5.

Results and Discussion

The system, shown in Fig. 1, is designed and simulated in MATLAB/Simulink, and the results are discussed as follows.

Response of System

The solar power incident on the PV array set equal to 1000 W/m^2 and temperature 25 °C for steady-state performance of the system and dynamic performance of the system is studied when insolation level is changed from 1000 to 400 W/m^2 at 0.6 s. The system performance is validated by the simulative results consisting of the grid voltages (V_{sabc}), grid current (I_{sabc}), AC phase voltage of VSC (V_{cbc}), PV voltage (V_{pv}), SPV current (I_{pv}), the power (P), PV voltage & current for array 1 of A-phase (V_{a1pv} & I_{a1pv}), PV voltage & current for array 1 of B-phase (V_{b1pv} & I_{b1pv}), PV voltage & current for array 1 of C-phase (V_{c1pv} & I_{c1pv}) as depicted in Figs. 6a and b and 7a–c. The maximum active power transfer is evident at a power factor of unity.

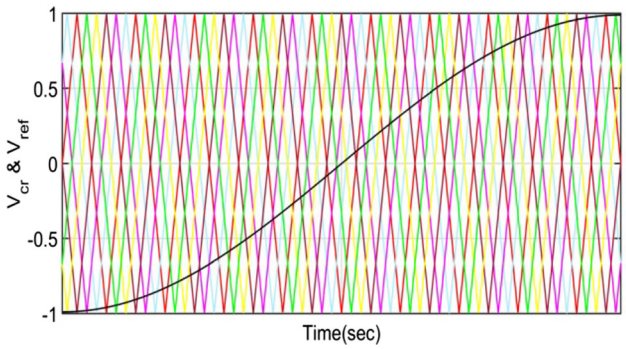


Fig. 5 Phase-shifted PWM scheme

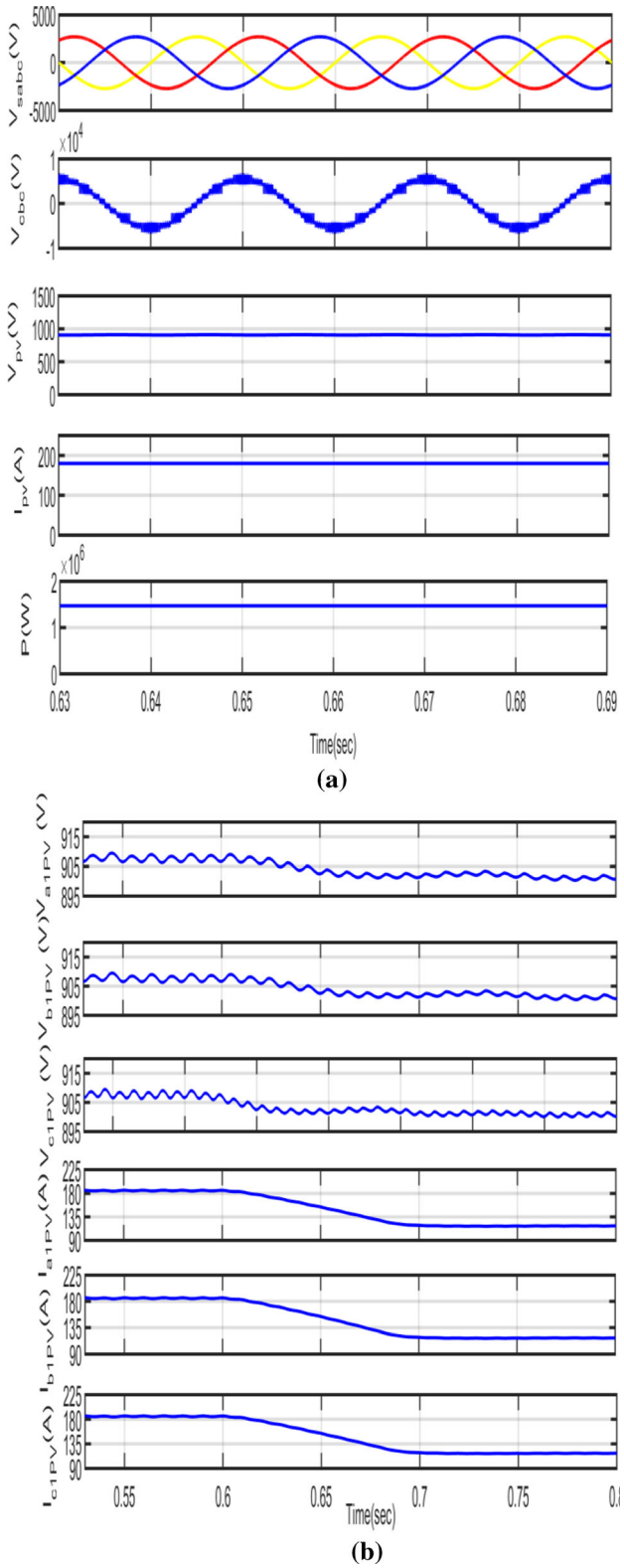
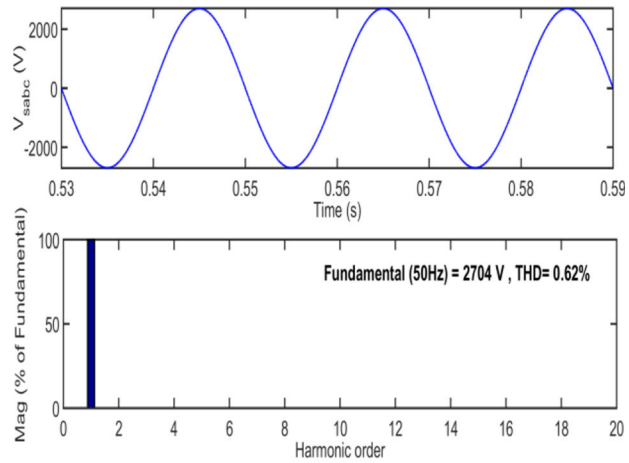
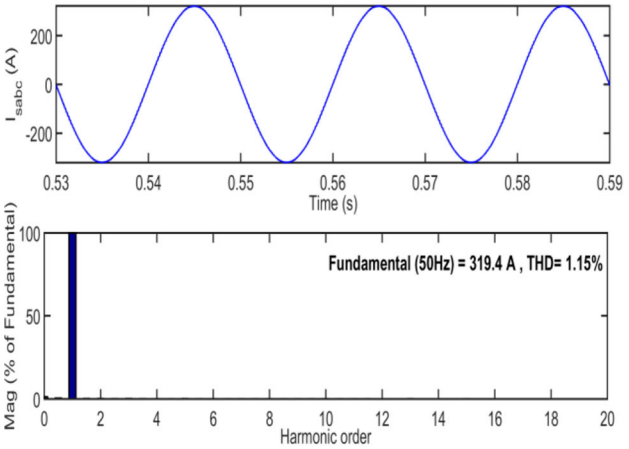


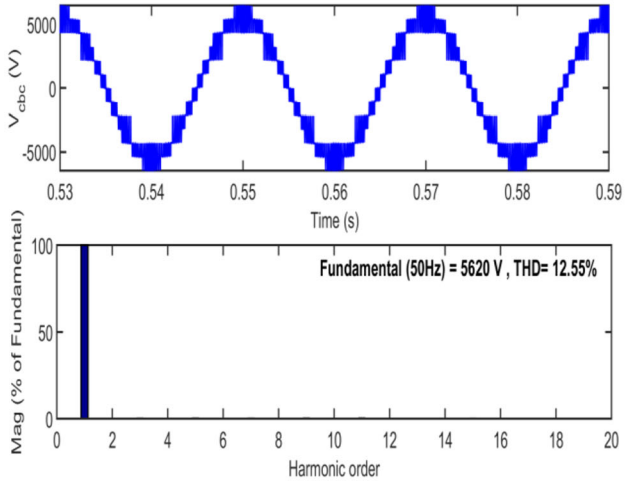
Fig. 6 **a** Steady-state performance at insolation level 1000 W/m^2, **b** dynamic performance when insolation level is changed from 1000 to 400 W/m^2 at 0.6 s for seven-level CHB grid-interfaced system



(a)



(b)



(c)

Fig. 7 Waveform and harmonic spectra of **a** line voltage, **b** line current and **c** converter line voltage at 1000 W/m^2

Improvement in Power Quality

Figure 7 shows the waveforms for the line voltage, line current (I_a) and converter line voltage (V_{cbc}) and harmonic spectra along with THD levels at 1000 W/m^2 levels. The THD values of 0.62%, 1.15% and 12.55% at insolation of 1000 W/m^2 and temperature of 25°, respectively, are observed for line voltage, line current and converter line voltage. The THD of the supply current is observed to be within the limits set by the IEEE-519 standard [26]. A minor variance is seen, though the power factor is also close to unity.

Conclusion

The design and control of single-stage transformerless seven-level H-bridge cascaded multilevel converters have been carried out with the decoupled current controller based on SRF-PLL and phase-shifted PWM at lower (500 Hz) switching frequency for large-scale photovoltaic grid-integrated plant. The low switching frequency reduces the switching losses, and hence, an improvement in the efficiency is achieved. An enhanced power quality is observed because of increased steps in output multilevel converter voltage waveform as compared to a two-level converter. It is attributed to the reduction in the supply current THD and reduction in size of filter components. Only an inductive filter is used for filtrating, which further results in reduction in the cost. The performance of the system is observed satisfactory at 1000 W/m^2 insolation level and 25° temperature and has been validated in accordance with the IEEE-519 standard [26].

References

1. F. Blaabjerg, R. Teodorescu, M. Liserre, A.V. Timbus, Overview of control and grid synchronization for distributed power generation systems. IEEE Trans. Ind. Electron. **53**(5), 1398–1409 (2006)
2. S. Kouro, M. Malinowski, K. Gopakumar, J. Pou, L. Franquelo, B. Wu, J. Rodriguez, M. Pérez, J. Leon, Recent advances and industrial applications of multilevel converters. IEEE Trans. Ind. Electron. **57**(8), 2553–2580 (2010)
3. J. Rodriguez, J.S. Lai, F.Z. Peng, Multilevel inverters: a survey of topologies, controls, and applications. IEEE Trans. Ind. Electron. **49**(4), 724–738 (2002)
4. R. Doga, M. Pathak, Recent trends in solar PV inverter topologies. Sol. Energy **183**, 57–73 (2019)
5. B. Xiao, L. Hang, J. Mei, C. Riley, L.M. Tolbert, B. Ozpineci, Modular cascaded H-bridge multilevel PV inverter with distributed MPPT for grid-connected applications. IEEE Trans. Ind. Appl. **51**(2), 1722–1731 (2015)
6. T. Kerekes, R. Teodorescu, M. Liserre, C. Klumpner, M. Sumner, Evaluation of three-phase transformerless photovoltaic inverter

topologies. IEEE Trans. Power Electron. **24**(9), 2202–2211 (2009)
7. M.R. Islam, Y. Guo, J. Zhu, A multilevel medium-voltage inverter for step-up-transformer-less grid connection of photovoltaic power plants. IEEE J. Photovolt. **4**(3), 881–889 (2014)
8. B. Singh, I. Hussain, Grid integration of single stage solar PV power generating system using 12-pulse VSC, in *2014 IEEE 6th India International Conference on Power Electronics (IICPE)*, Kurukshetra (2014), pp. 1–6
9. I. Hussain, B. Singh, Investigations on solar PV grid interfaced power generating system using two-level twelve-pulse double bridge converter, in *2014 9th International Conference on Industrial and Information Systems (ICIIS)*, Gwalior (2014), pp. 1-6
10. C. Zhang, S. Du, Q. Chen, A novel scheme suitable for high-voltage and large-capacity photovoltaic power stations. IEEE Trans. Ind. Electron. **60**(9), 3775–3783 (2013)
11. I. Hussain, B. Singh, Grid integration of large capacity solar PV plant using multipulse VSC with robust PLL based control, in *6th IEEE Power India International Conference (PIICON)*, Delhi (2014), pp. 1–6
12. N. Mishra, P. Kant, B. Singh, An asymmetric seven levelmultilevel converter for grid Integrated systems, in *2018 8th IEEE India International Conference on Power Electronics (IICPE)*, Jaipur, India (2018), pp. 1–6
13. S. Kuncham, K. Annamalai, S. Nallamothu, Single phase two stage seven-level power conditioner for photovoltaic power generation system. IEEE J. Emerg. Sel. Top. Power Electron. **8**(1), 794–804 (2020)
14. B. Singh, N. Mishra, P. Kant, Power quality improvement in single phase five level cascaded grid interfaced systems, in *2018 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES)*, Chennai, India (2018), pp. 1–6
15. R. Gonzalez, E. Gubia, J. Lopez, L. Marroyo, Transformerless single-phase multilevel-based photovoltaic inverter. IEEE Trans. Ind. Electron. **55**(7), 2694–2702 (2008)
16. B. Singh, M. Kandpal, I. Hussain, Grid integration of single stage SPV system using asymmetric cascaded 7-level VSC, in *39th National Systems Conference (NSC)*, Noida (2015), pp. 1–6
17. A. Acharya, M. Rico, D. Sera, R. Teoderscu, L. Norum, Performance analysis of medium-voltage Grid Integration of PV plant using modular multilevel converter. IEEE Trans. Energy Conv. **34**(4), 1731–1740 (2019)
18. A. Elmelegi, M. Aly, E. Ahmed, A. Alharbi, A simplified phase shift PWM-based feedforward distributed MPPT method for grid-connected cascaded PV inverters. Sol. Energy **187**, 1–12 (2019)
19. B. Subudhi, R. Pradhan, A Comparative Study on maximum power point tracking techniques for photovoltaic power systems. IEEE Trans. Sustain. Energy **4**(1), 89–98 (2013)
20. A. K. Singh, I. Hussain, B. Singh, An improved P&O MPPT algorithm for single stage three-phase grid integrated solar PV system, in *IEEE 7th Power India International Conference (PIICON)*, Bikaner (2016), pp. 1–6
21. M. Killi, S. Samanta, Modified Perturb and observe MPPT algorithm for drift avoidance in photovoltaic systems. IEEE Trans. Ind. Electron. **62**(9), 5549–5559 (2015)
22. B.P. McGrath, D.G. Holmes, Multicarrier PWM strategies for multilevel inverters. IEEE Trans. Ind. Electron. **49**(4), 858–867 (2002)
23. B. Wu, *High-Power Converters and AC Drives* (IEEE, 2006)
24. B. Singh, A. Chandra, K. Al-Haddad, *Power Quality: Problems and Mitigation Techniques* (Wiley, Hoboken, 2015)
25. R. Teodorescu, M. Liserre, P. Rodriguez, *Grid Converters for Photovoltaic and Wind Power Systems* (Wiley, Hoboken, 2011)
26. IEEE Recommended Practices and Requirements for Harmonic Control in Electrical Power Systems, IEEE Std 519-1992 (1993), pp. 1–112, 9

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/342523500

COVID-19 Global Pandemic: Impact on Management of Supply Chain

Article · April 2020

DOI: 10.46338/IJETAE0416

CITATIONS
0

READS
55

2 authors, including:



Rakesh Kumar
Delhi Technological University
7 PUBLICATIONS **0** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Performance Measurement of TQM in Indian FMCG Industry View project

Project Population Statistics View project

All content following this page was uploaded by Rakesh Kumar on 12 July 2020.

The user has requested enhancement of the downloaded file.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 10, Issue 04, April 2020)

COVID-19 Global Pandemic: Impact on Management of Supply Chain

Rakesh Kumar[1], R.S. Mishra[2]
[1,2]*Department of Mechanical Engineering, Delhi Technological University, Delhi, India-110042*

***Abstract*— To understand more about the global impact of (corona virus disease, detected in December, 2019, hereafter named) COVID-19 on supply chain management, a set of questionnaire is designed which addressed four pertinent themes focusing on product classification including export product, import product, elective product and emergency product. The questionnaire consisted of ten multiple choice questions including three open-ended questions (appendix A). The questionnaire was targeted at those organizations which were practicing supply chain management function as a core or a department of organization and was distributed online in interest groups using various social media platforms. The questionnaire was open from 26th March 2020 to 20th April 2020 and did not collect any identifiable data. Valid responses were represented as categorical data and presented in percentages.**

***Keywords*— COVID-19; supply chain management, product classification**

I. INTRODUCTION

The global downturn's speed and harshness due to COVID- 19 have significant implications for the supply chains of global manufacturers. That kind of volatility wreaks havoc on traditional supply chain planning: the process for determining production levels, raw-material purchases, transport capacity, and other vital factors, largely by examining historical patterns of demand. Indeed, the forecasting challenge is particularly acute because in many upstream industrial settings, as supply partners along the chain anticipate that demand will fall, the supply chain appears to be decoupling from downstream consumption—the focus of most forecasting models. In fact, risk managers and other executives fail to anticipate the effects, both negative and positive, of events that occurred. At first glance, for instance, a thunderstorm in a distant place wouldn't seem like cause for alarm. It was initially reported to the World Health Organization (WHO) on December 31, 2019 about Coronavirus disease 2019 (COVID-19) identified amid an outbreak of respiratory illness cases in Wuhan City, Hubei Province, China. On January 30, 2020, the WHO declared the COVID-19 outbreak a global health emergency. On March 11, 2020, the WHO declared COVID-19 a global pandemic.

The appeal of Janata Curfew on March 22,2020 by Hon'ble Prime Minister of India that the nation is set to observe an unprecedented shutdown and on March 24,2020 India imposed a 21-day lockdown, which was announced, comes after a call from the UN health agency, WHO, for the country to take "aggressive action".

What can companies do to prepare themselves for this sort of situation? True, there's no easy formula for anticipating the way risk cascades through a company or an economy. But we've found that executives who systematically examine the way risks propagate across the whole value chain—including competitors, suppliers, distribution channels, and customers—can foresee and prepare for second-order effects more successfully. Especially China, Korea, Japan, Germany and the US are also at of global value chains, so their woes will produce 'supply-chain contagion' in virtually all nations, Richard Baldwin and Beatrice Weder di Mauro apprehended this fact in their e-book on Economics in the Time of COVID-19. These nations are central to the global supply chains in many manufactured goods. The near term impacts—missed shipments, border-crossing delays, and communications problems—are easily understood. Less clear is the longer-term effect of increased uncertainty on lead times and demand and what that means for supply chain strategy. "Risk is something you can predict. If you flip a coin, over time you know that half will be heads and half will be tails. The nature of choice between heads and tails involves a risk because you can bet with these probabilities. So, in conditions of risk, you can distribute your investment into certain proportions against different outcomes that you may expect," Kim said (Minyoung Kim, associate professor of strategy and international business at the University of Kansas)."But uncertainty means there's nothing you can do about it beforehand. You know neither the outcomes nor the probability. I would call that the 'unknown unknown.' This could be a natural disaster or Brexit. If you knew, you might prepare against it. But you don't know that you don't know." The new uncertainty COVID-19 will require companies to rethink their supply chain strategy. For many, that effort begins with a reassessment of their demand-based strategy, which in recent years has been touted as the preferred approach to supply chain management.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 10, Issue 04, April 2020)

Observing the scenario of Indian market S. Srinivasamurthy, Research Director, Enterprise Solutions and ICT Practices, IDC India says, "While the actual impact of COVID-19 on India market will be evident by middle of 2020, we expect a slowdown in terms of discretionary IT spending, contract renewals and new deals getting signed as enterprises recalibrate by cost structure in coming months. Existing project executions have also taken a hit due to travel restrictions in place. Corporates across the country implementing alternative ways of working, i.e. work from home, it is generating a parallel corporate line that demands to be connected from where they want, when they want and to who they want. But the different sets of challenges that the IT teams within the organization are grappling with - how to secure data and assets from cyber threats. Companies can meet most challenges noted above with a sound supply chain strategy, with IT as the key enabler. Though the complexities of India's supply chain may appear overwhelming, understanding and mastering them is a critical success determinant for an organization attempting to serve customers in India. An efficient review of supply chain design will help better position companies in what is becoming an increasingly competitive marketplace.

Supply chain in Indian Economy

Supply chain costs in India represent as much as 13 percent of the Gross Domestic Product (GDP). This is almost twice the percentage in developed countries: in the U.S., supply chain costs amount to 8.5 percent of the GDP. ARC Advisory Group categorizes supply chain challenges in India into two groups: demand side challenges and supply side challenges. Demand side supply chain challenges in India relate to price and variety. With 28 States and 7 Union Territories, differing considerably in dress, food tastes, customs, traditions and purchasing habits, it's almost impossible for a single manufacturer to address the wide variety of consumer requirements. Thus, it is important for manufacturers to partner with local distributors and retailers who can help develop products that will appeal to local consumers, at appropriate price points. When it comes to COVID-19's immediate aggregate demand shock (i.e purchases drop) is observed, for which, two aspects are to distinguish: real-world and emotional. Real-world since some consumers are or will be prevented from getting to stores, so their demand disappears from the market. Similarly, some home delivery services are suspended, so goods and consumers are approaching together less frequently.

Emotional since – as happened in the wake of the Global Crisis – consumers and firms tend to embrace a 'wait-and-see' attitude when faced with massive Knightian uncertainty (the unknown-unknowns) of the type that COVID-19 is now presenting to the world. This, quite naturally, makes trade flows susceptible to demand shocks (purchases drop) and supply shocks (production drop). The supply shock aspects of COVID-19 – factory closures, travel bans, border closings and the like – will reduce exports across the border. Many vendors have been successful in selling "no-frills," standardized products to rural segments in India only. Big Bazaar, Hindustan Unilever, Maruti, Tata Motors and Hero have offerings relevant to this price-sensitive segment of the population. Other successful examples of localized low-price products include Siemens' pacemakers, Godrej's Chotukool refrigerator, and General Electric's ECG. Supply side supply chain challenges in India mainly relate to poor infrastructure, complex tax infrastructure, weak distribution system, fragmented market, and lack of technology adoption. Richard Baldwin and Beatrice Weder di Mauro as responsible editor mentioned in their recent e-book on Economics in the Time of COVID-19, the disruptions that companies, individuals and governments are experiencing imply that globalisation and integration may be at risk from such health shocks i.e. COVID 19. Firms will probably take into account the lesson they are learning that global supply chains can be abruptly broken by this health shock.

II. Literature Review

Businesses dependent on global sourcing are facing hard choices in crisis management amid the supply chain disruptions due to COVID-19. The impact of the coronavirus pandemic on global supply chains is a major disruption, along the lines of having an earthquake or a tsunami. This is an unprecedented type of disruption. The uncertainties ahead swing between extremes. The main target of managing the supply chain is to realize and neutralize the uncertainties in the supply chain. Uncertainties in the supply, process and demand are recognized to have a major impact on the manufacturing function. As this paper focuses on the impact of COVID-19 on the supply chain management, the literatures were reviewed nearby similar context of reputed journals related to the flexibility, agility and pro-activeness of the supply chain.

Angkiriwang, R. et al. (2014) in his study obtain insights into the typology of uncertainty and relevant strategies adopted by manufacturing companies to achieve better supply chain flexibility.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 10, Issue 04, April 2020)

Strategies are classified into reactive (buffering) and proactive (redesigning). A framework is developed that links supply chain uncertainty, as result the two types of strategies for achieving supply chain flexibility and the relevant objectives to be achieved. The research is endorsed by four case studies and are compared in terms of uncertainty typology and strategies being adopted to improve supply chain flexibility. supply chain flexibility (SCF) entails the implicit requirement of flexibility within and between all partners in the chain (Duclos *et al.*, 2003). Vickery, S.(1999) to establish the linkage, examined dimensions of supply chain flexibility and their relationships with environmental uncertainty, business performance, and functional interfaces. Prater et al. (2001) used five case studies to show the techniques of better trade-off between vulnerability (uncertainty) and supply chain agility by creating a link between the two called supply chain exposures. They considered and found two main problems of European industries, that is, in accurate forecasting and in developing an efficient and effective supply chain (SC) structure. The work described exposure as the degree to which an agile supply chain must be overextended (i.e. vulnerable) and, therefore, should be restructured or adjusted to address the issues in international SCs. To obtain high service levels at a low cost for emergency supply options in terms of lateral transshipments and direct deliveries, Alfredsson, P. et al. (1999) consider a two-echelon inventory system. An analytical model is used to calculate relevant performance measures. Comparison of the results of the model with the results of other models indicates that the combined use of lateral transshipments and direct deliveries can lead to significant cost savings. Agrawal, S. et al. (2019) emphasized about the segregation of products into different classified categories is essential for reverse supply chain also.

Like COVID-19, the earthquake and tsunami that struck Japan in March 2011 directly impacted over 27000 businesses whose production, warehousing, and retail facilities were destroyed or disabled by the natural disaster. Supply chain disruption is defined MacKenzie, C. A (2014) as a disruptive event that causes production difficulties for multiple suppliers and in which at least two of these suppliers deliver non-identical goods or services to at least two competing firms. The severity of a supply chain disruption can be measured by the number of entities that encounter difficulties in receiving or delivering materials or goods due to an unplanned event (Craighead et al., 2007).

A number of authors have tried to develop a framework for relationship between the typology of uncertainty and the type of strategies chosen for better supply chain. The study proposed in this article follows real time approach of the disruption COVID-19. The survey based work analyzes what supply chain firms should do during a severe supply chain disruption under uncertain conditions.

III. Motivation For The Research

The current COVID-19 global effect greatly motivated for this work. As India goes into a 21-day lockdown, there is some confusion arises about the e-commerce delivery of essential goods. According to lockdown rules, delivery of essential goods and groceries from e-commerce sites is allowed. Services like Grofers, BigBasket and Milkbasket allege their delivery people have been beaten up on the road. Facing difficulties, Flipkart has temporarily shut down service. Others too vary(www.indiatoday.in). The three East Asian manufacturing giants – China, South Korea, and Japan – account for over 25% of US imports, and over 50% of US imports of computer and electronics products. Apparel and footwear companies are particularly vulnerable to East Asian supply disruptions. The Economist magazine pointed to the vulnerability of the electronics industry with its practice of keeping very lean inventory levels and the lack of alternative sources for many electronic components. The optics sector is likewise highly exposed. The heart of the outbreak, Hubei province, is known as China's 'optics valley' since so many firms manufacturing fiber optic components are located there (these are essential inputs for telecoms networks). The automobile sector, especially in East Asia, is already significantly disrupted by ruptures in international supply chains. For example, a shortage of parts coming from China has forced Korean carmaker Hyundai to shut all its car plants in Korea. The Japanese firm Nissan closed a factory in Japan temporarily. The shock has even reached Europe. Fiat-Chrysler has recently warned that it could soon halt production at one of its European factories. Jaguar Land Rover, a UK-based auto company, announced it might to run out of parts from the end of February,2020. To stave this off, it had flown in emergency supplies from China in suitcases [7].

IV. Problem Description

The basis used by marketers to classify the products are: (i) On the basis of durability and tangibility of the product (ii) On the basis of Consumer Shopping Habits and (iii) On the basis of use in industry.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 10, Issue 04, April 2020)

Traditionally products have various classifications, consumer products and industrial/business products is based on use is classified. Consumer product is further classified as: convenience products, shopping products, specialty products and unsought products and industrial product is classified as: material and parts (these goods enter the product directly), installation, accessory equipment, operating supplies. In general, the consumer goods directed to ultimate end-users, but the buyers seek product and marketing is important for it, but the industrial products used to produce other products and services, and usually vendors seek buyers. Based on durability it's classified as durable and non-durable and few classified products on the basis of tangibility, tangible and non-tangible also. But It is vital to design a product classification method in order to support global supply planning and improve the supply chain performance.

A. Product classifications

With respect to the present study and to measure the impact of COVID-19 on SCMs, the products are classified on basis of exigency as export products, import products, elective products and emergency products. Categorization of the product is important one for those organizations which were dealing with supply chain, the basis they can maintain it in their data base and exploit it at the time of need. This categorization helps to do planning accordingly by evaluating its gravity to the present situation of the demand.

Export products: those products which is manufactured by manufacturer for the specifically export purposes.

Import products: The product which will arrive from other country.

Elective products: Also called optional product which have substitute for the customer.

Emergency products: In this category those products will lie which are critically necessary for providing care for people. Viz. medical devices, medical products, productive equipment, masks, screeners, disinfectants and so on.

Companies use to handle complicated supply chain operation but lacks in product classification system. It is vital to design a product classification method in order to support global supply planning and improve the supply chain performance. That part of this research for future work. Also, Product Classification System is supposed to be applied in supply planning prioritization, inventory management and workforce balancing, etc.

The things which are considered as mechanism for the propagation of economic shocks, or economic contagion are: Products (tangible), people, services, know-how(skill), financial capital, foreign direct investment (FDI), international banking, and exchange rates. The flow of global economies dependencies is connected by cross-border by these mechanisms, and product is considered at top priority among all the above. Companies are learning the lesson that global supply chains can be abruptly broken by a health shock and will adjust accordingly. Financial intermediaries and regulators are likely to incorporate pandemic shocks into their future risk assessments and stress tests.

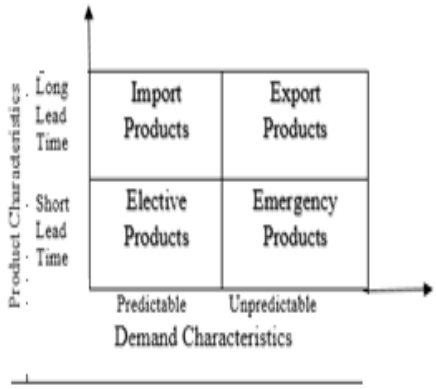


Figure 2. Characteristics of Export, Import, Elective and Emergency Product

V. Data Collection And Analysis

A. Questionnaire development

A questionnaire was developed based on the practical practice of supply chain in various organizations proposed for the study and aimed at gathering information related to dealing with the basis of classification of the product and its impact in context to Indian supply chain industry. With the help of available literature and telephonically talk with the experts the questionnaire is developed and suggestions were incorporated and questionnaire was improved before distribution for data collection. Survey form was developed with the help of google forms. The survey questionnaire is available in Appendix A. The survey was sent to the top management person who was handling these organizations via mail as well as in person asking people to participate in the survey. The given deadline for data collection for the responses was of 26th March to 20th April 2020.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 10, Issue 04, April 2020)

B. Questionnaire administration

The questionnaire was self-administered, and the survey for the study with target population of supply chain industries were conducted through google form, online. The survey link was sent on email addresses of the respondents with a cover letter. Nowadays, web-based survey is being widely used by researchers because of less time required, ease of availability, real time data compilation, and lower cost than other surveys. With the help of google search engine, the dedicated organization were searched with their e-mail address to consult for the survey. Justdial.com helps more to lean the time of searching to those companies. To cover the entire India, the companies of Delhi, Kolkata, Mumbai, Chennai, Bangalore, Hyderabad and Ahmedabad were selected, besides these cities Pune, Goa, Nagpur, Coimbatore, Indore, Nashik, Vadodara, Jaipur, Mysore, Vizag, Chandigarh and Surat, many companies are there which associated with supply chain operations. The questionnaires were sent to the respondents in single phases. 300 forms were sent to different organizations through google forms and responses were analyzed. Total 130 responses were received 35 were incomplete so discarded. An effective response rate of 43.3% was achieved which is found adequate for the present study.

C. Respondent's profile and non-response bias

Due to time constraint, in single phases, total 300 e-mails were sent to the selected supply chain dedicated organizations for the survey. Overall, 130 responses were found suitable for the study. Participating organizations in the survey belong to (Original Equipment Manufacturer) OEM, Fast Moving Consumer Goods (FMCGs) and SCM consultant. Respondents who have participated in the survey represent different levels of management including top position to middle level and executive level employees and mostly belong to logistics department, marketing & sales, corporate strategy, and supply chain departments. Non-response bias with web-based survey was accessed by splitting the sample into two groups based on receiving date of the responses. Early response group consisted of 67 responses and late response group consisted of 63 responses. The t-test was performed on the difference in early response group and late response group. It yielded no statistically significant difference between the mean of two groups ($p > 0.05$). In addition, statistical tests show that there is no significant difference among samples.

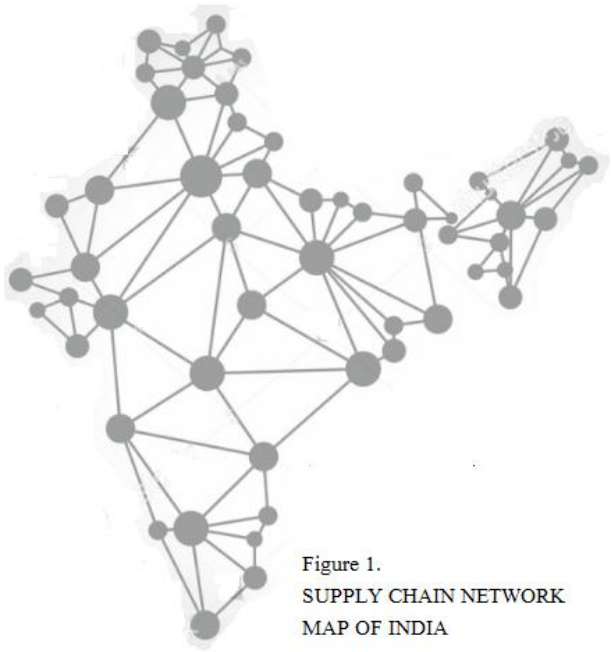


Figure 1.
SUPPLY CHAIN NETWORK MAP OF INDIA

VI. Result And Discussion

To understand more about the global impact of COVID-19 on supply chain management, we designed a questionnaire that addressed four pertinent themes including export product, import product, elective product and emergency product. The questionnaire consisted of 7 multiple choice questions and three open-ended questions (appendix A). The questionnaire was targeted at supply chain organizations and was distributed online in interest groups using various social media platforms. The questionnaire was open from 26th March to 20th April 2020 and did not collect any identifiable data. Valid responses were represented as categorical data and presented in percentages.

Impact studies with individual SCM organization data

This section describes a new wave of research that is based on samples of organization collected precisely to study strength of chain of supply with the help of product classification. The sampling was designed to examine organizations in different product category zones and to select organizations within country across a wide range of product. The survey instrument was designed to measure accessibility of products in places without barriers and hindering when situation is unfavorable.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 10, Issue 04, April 2020)

The instrument collected information about the choices that organizations made: which product to ship, which product to hold and which inputs to purchase. Precise Products classification give solutions and support decision-making for Global Supply Planning. For example, higher priority of products requires shorter lead time of supplying. The crisis never took place for time with the organization, but the plan was put in place and organization feel secure knowing that should it happen, they would be safe. For all the effort, resources, and consideration put into designing the most efficient manufacturing facilities and systems, it should be a natural next step to protect them by developing crisis response plans. Fisher is an acknowledged pioneer in the classification of supply chain strategies, claiming that there are basically two types of supply chain strategies, physically efficient and market responsive (Harris, 2007).

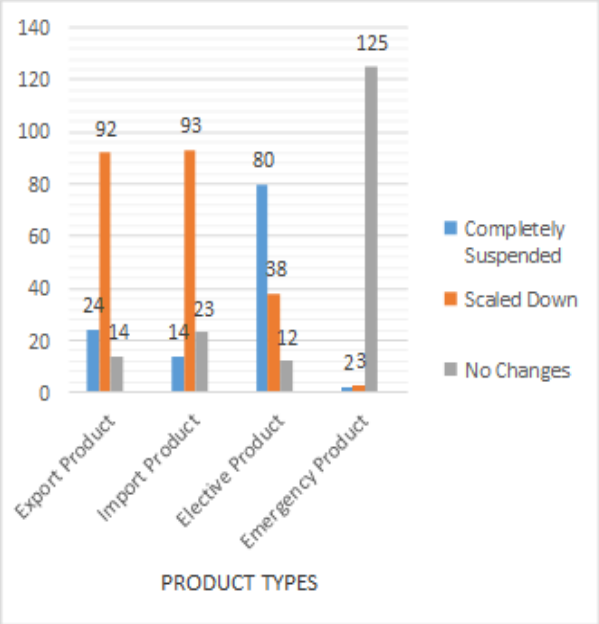
A total of 130 individual responses were obtained from the supply chain firms. Of these, 32(24.6%) were from the Delhi, 21 (16.1%) were from Kolkata, 10 (07.6%) were from Mumbai, 17 (13.0%) were from the Chennai, 17 (13.0%) were from Bangalore,19 (14.6%) were from Hyderabad and 14 (10.7%) were from Ahmedabad. Of the 130 respondents, 85(65.3%) were working Supply Chain as a part of OEM. When asked about export product, 102 (78.4%) replied that export product orders were either suspended or scaled down. When asked about import product, 118 (90.7%) replied that import product were either suspended or scaled down. With regards to elective products, 120 (92.3%) replied that at least some, if not all elective products orders were cancelled. The majority of orders cancelled were due to its availability of substitute product or not felt essential and emergency by the customer. On the contrary, 125 out of 130 respondents (96.2%) are still delivering all emergency products (Figure 1). For a large majority of respondents (113 out of 130, 87.0%), face-to-face multidisciplinary or team meetings have been reduced or cancelled entirely, and have been replaced by online messaging platforms, video conferencing, teleconferencing or email discussions. The rate of response also reflects the effect of lockdown.

As only 13% response from Mumbai and also we got news that Maharashtra counts highest death due to COVID-19. The impact of COVID-19 on supply chain services has been far-reaching. The majority of respondents have reported a reduction or cessation of export products orders and elective product orders, whilst continuing with emergency products.

Numbers from group purchasing organization Premier and data software companies show the supply chain gaps in getting needed personal protective equipment (PPE) to those on the front lines. Data on increasing orders of PPE in the hospital setting and the arc of COVID-19 calls compared to a typical flu season show the pandemic in a new light. Healthcare systems, distributors and purchasing organizations, along with state and federal governments, rely on real-time data to manage supplies and model their equipment and staffing needs. so manufacturers and their complex webs of suppliers, who in many cases lack that same experience or large-scale backup plans. The result: a chaotic scramble resulting in production delays, product shortages, and higher prices.

E-commerce giant Amazon has announced that it will only deliver high-priority products. In view of the challenges faced by the delivery executives, the Karnataka government on Monday issued a notice stating that e-commerce and home delivery fall under essential services and hence should be exempted from the restrictions imposed on citizens to curb the spread of the coronavirus. Following the suit, Maharashtra too issued an order stating that the e-commerce services such as food, medicine, etc should not be included in the restrictions that are imposed to contain the virus. Grofers founder had tweeted his apology to the customers for not being able to deliver their products on time. He stated “We @grofers are having a lot of trouble getting essentials to people who need them as our warehouses are asked to shut, and trucks and delivery partners are being stopped by the police. We apologize to our customers and are working hard to find a solution.” Flipkart also suspended its services temporarily.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 10, Issue 04, April 2020)



REFERENCES

[1] Reina Angkiriwang, Nyoman Pujawan & Budi Santosa. 2014. Managing uncertainty through supply chain flexibility: reactive vs. proactive approaches. Production & Manufacturing Research, 2:1, 50-70, DOI: 10.1080/21693277.2014.882804

[2] Buyurgan, N., Zhang, S., & Okyay, H. K. (2019). Methodical analysis of inventory discrepancy under conditions of uncertainty in supply chain management. International Journal of Logistics Systems and Management, 32(2),

[3] Patrik Alfredsson, Jos Verrijdt. (1999). Modeling Emergency Supply Flexibility in a Two-Echelon Inventory System. Management Science45(10):1416-1431.

[4] Richard Baldwin and Beatrice Weder di Mauro, "Economics in the Time of COVID-19", A CEPR Press VoxEU.org eBook, ISBN: 978-1-912179-28-2.

[5] Prater, E., Biehl, M. and Smith, M.A. (2001). International supply chain agility: tradeoffs between flexibility and uncertainty. International Journal of Operations and Production Management, Vol. 21 Nos 5/6, pp. 823-839.

[6] Vickery, S. nee, Calantone, R., & Droge, C. (1999). Supply Chain Flexibility: An Empirical Study. The Journal of Supply Chain Management, 35(3), 16–24.

[7] Duclos, L.K., Vokurka, R.J. and Lummus, R.R. (2003). A conceptual model of supply chain flexibility. Industrial Management & Data Systems, Vol. 103 No. 6, pp. 446-456.

[8] McKibbin, Warwick J. and Fernando, Roshen. (2020). The Global Macroeconomic Impacts of COVID-19: Seven Scenarios (March 2, 2020). CAMA Working Paper No. 19/2020.

[9] MacKenzie, C. A., Barker, K., & Santos, J. R. (2014). Modeling a severe supply chain disruption and post-disaster decision making with application to the Japanese earthquake and tsunami. IIE Transactions, 46(12), 1243–1260.

[10] Craighead, C.W., Blackhurst, J., Rungtusanatham, M.J. and Handfield, R.B. (2007) The severity of supply chain disruptions: Design characteristics and mitigation capabilities. Decision Sciences, 38(1), 131–156.

[11] Harris, G.A. (2007). Alignment of Supply Chain Strategy with Product Characteristics. Ph.D. diss. The University of Alabama in Huntsville.

[12] Agrawal, S., Singh, R. K. (2019). Analyzing disposition decisions for sustainable reverse logistics: Triple Bottom Line approach. Resources, Conservation & Recycling, 150 (2019),104448

[13] https://www.idc.com/url.do?url=/getdoc.jsp?containerId=prAP46181620&position=5&transactionId=74540281&term=&page=1&perPage=25

[14] https://www.supplychainbrain.com/articles/15619-indias-supply-chain-is-improving-but-still-has-a-long-way-to-go

[15] https://phys.org/news/2020-04-global-uncertainty-businesses-flexibility-efficiency.html?utm_source=TrendMD&utm_medium=cpc&utm_campaign=Phys.org_TrendMD_1

[16] https://www.economist.com/finance-and-economics/2020/02/22/covid-19-presents-economic-policymakers-with-anew-sort-of- threat?cid1=cust/ednew/n/bl/n/2020/02/

[17] https://www.justdial.com/Delhi/Supply-Chain-Management-Companies/nct-10808773

Appendix A

Questionnaire

Name of the organization:
_____
Year of establishment:
_____
Annual Turnover (in Rupees):
_____
Number of Employees:
_____

1. What is your country of practice of supply chain

2. Is your company is dealing with Supply Chain as a part of OEM or as a consultant company?

a. Yes b. No c. Other (please specify)

3. How has COVID-19 affected your export product order demand?

a. export product order completely cancelled

b. export product order scaled down i.e. non-urgent products postponed/cancelled

c. export product order demands operate as per usual

d. Other (please specify)

4. How has COVID-19 affected your import product demand?

a. import product completely cancelled

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 10, Issue 04, April 2020)

b. import product scaled down i.e. reduction in number of orders

c. import product demand as per usual

d. Other (please specify)

5. How has COVID-19 affected your elective product demand?

a. All elective product demands are cancelled

b. Some elective product demands are cancelled

c. No elective product demands are cancelled

d. Other (please specify)

6. If elective product demands have been cancelled, what type of products are they? Leave blank if no elective product demands have been cancelled.

7. How has COVID-19 affected your emergency product demand?

a. All emergency product orders are rejected

b. Only some emergency product orders are rejected

c. No emergency product orders are rejected

d. Other (please specify)

8. If emergency product orders have been rejected, what type of products are they? Leave blank if no emergency product orders have been rejected.

9. How has COVID-19 affected face-to-face multidisciplinary meetings?

a. All face-to-face multidisciplinary meetings are cancelled

b. Some face-to-face multidisciplinary meetings are cancelled

c. No face-to-face multidisciplinary meetings are cancelled

d. Other (please specify)

10. What modality has replaced face-to-face multidisciplinary meetings? Leave blank if your face-to-face multidisciplinary meetings have not been cancelled. (You can choose more than 1 option)

a. Teleconference

b. Video conference

c. Online messaging platforms i.e. WhatsApp, Telegram, etc

d. Email e. Other (please specify)

About Author(s)

Mr. Rakesh Kumar received Bachelors in Industrial Engineering and Master's in Mechanical Engineering with specialization in Computer Integrated Manufacturing from India. He is currently a assistant professor at Delhi Technological University in Department of Mechanical Engineering. His research interest includes *Total Quality Management, Reliability Engineering, Operations and Supply chain, Fuzzy Logic, AI and Decision Science*.

R.S. Mishra (Professor) received the Ph.D. degree in Mechanical engineering from IIT Delhi, India and currently at Delhi Technological University serving since 1997. He having 150 research papers in peer reviewed in International journals and 75 research papers in proceedings of International and National conferences. Prior to joining DTU, he served 09 years at different engineering colleges of India, viz, REC Hamirpur, PAU Ludhiana (Punjab)and HAU, Hisar (Haryana.).His research area includes *Green Technology, Refrigeration & Air Conditioning, Metal Cutting Technology, Total Quality Management, Solar Energy Technology, Bio Fuels, power Plant engineering*.

View publication stats

Proceedings of the Fourth International Conference on Trends in Electronics and Informatics (ICOEI 2020)
IEEE Xplore Part Number: CFP20J32-ART; ISBN: 978-1-7281-5518-0

Deep Learning Approach for Arbitrary Image Style Fusion and Transformation using SANET model

[1]Pratibha Rathi
Department of Computer Science & Engineering
Delhi Technological University
Delhi, India
pratibharathi96@gmail.com

[2]Pranav Adarsh
Department of Computer Science & Engineering
Delhi Technological University
Delhi, India
pranavadarsh95@gmail.com

[3]Dr. Manoj Kumar
Department of Computer Science & Engineering
Delhi Technological University
Delhi, India
mkumarg@dce.ac.in

***Abstract*—For real-time applications of arbitrary style transformation, there is a trade-off between the quality of results and the running time of existing algorithms. Hence, it is required to maintain the equilibrium of the quality of generated artwork with the speed of execution. It's complicated for the present arbitrary style-transformation procedures to preserve the structure of content-image while blending with the design and pattern of style-image. This paper presents the implementation of a network using SANET models for generating impressive artworks. It is flexible in the fusion of new style characteristics while sustaining the semantic-structure of the content-image. The identity-loss function helps to minimize the overall loss and conserves the spatial-arrangement of content. The results demonstrate that this method is practically efficient, and therefore it can be employed for real-time fusion and transformation using arbitrary styles.**

Keywords—image processing; deep learning; neural style transfer; computer vision; arbitrary image stylization; SANET; ASPM; convolutional neural networks.

I. INTRODUCTION

In the recent few years, various research works contributed to improve and accelerate the development of deep learning methods for continuous reformations of style transformation procedures. Arbitrary style transformation is the process of fusing new artistic styles with the content of the given image to create impressive artworks. NST was first introduced and explained by Leon A. gatys in 2015. Based on this idea presented by Gatys, many NST applications like a deep dream, prisma, and ostagram, etc. have been released to generate amazing artworks. The method proposed by Gatys [1] [2] follows an iterative procedure for the optimization of the generated image, which makes it computationally expensive and slow. However, the quality of the generated output is quite impressive. After that, numerous models have been proposed based on the optimization of the employed model. Johnson et al. [3] and Ulyanov et al. [4] [5] proposed fast offline approaches for generating output in a single feed-forward-pass, but their model is limited in the number of styles, and the quality of the result is also reduced. For fusing arbitrary styles using a single model, some ASPM approaches have been introduced. The swap model [6] was stated that employs the patch match procedure to substitute the features of content by the nearest match of characteristics of the style. Li et al. [7] specified multilevel stylization by recursively employing whitening-color-transformation to enhance the quality of output and to protect the content-structure. AdaIN [8] model changes the mean-variance of the content-image to match with style-image for transforming features. Avatar-net [9] model contains a style-decoder model based on patch-match and an hourglass network for adapting styles on multiscale. To solve the trade-off between content and style losses, this paper presents a new loss function identity-loss and a network consist of an integration of two SANET models for preserving both content-structures and style-characteristics. The arrangement of the paper is as follows. Below in section 2, background information of style transformation methods is given, which covers the explanation of some IBAR methods along with their drawbacks. Section 3 provides information about the existing neural methods of style transformation and their categorization into online and offline divisions. Section 4 illustrates the analysis of optimization approaches and expansion areas of NST by providing some detailed information about them. Section 5 demonstrates the implementation results and its comparison with other existing models. Finally, section 6 summarizes the conclusion.

II. BACKGROUND

In this section, the background information of the approaches used in style transformation before the evolution of deep learning is explained. The techniques in the taxonomy of image-based artistic rendering (IBAR) were employed for the stylization of two-dimensional images. First, the procedure of

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:25:38 UTC from IEEE Xplore. Restrictions apply.

some IBAR [10] methods are described and then their drawbacks are reviewed.

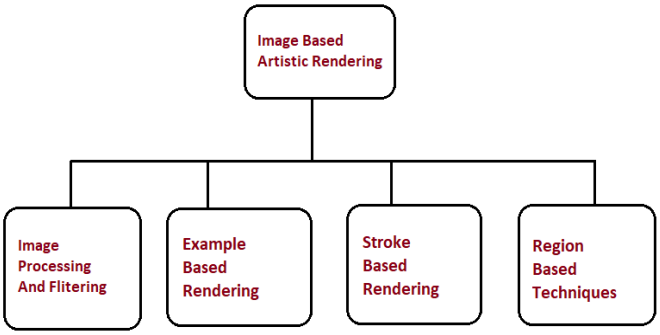


Fig.1: Categorization of IBAR techniques

A. *Stroke-based rendering*

It is the process of integrating portraits, paintings, and artworks by providing the strokes or marks on a digital canvas using brushes, lines, and tiles [11]. For converting a photograph into a particular style, virtually some brush strokes are superimposed at some positions on a digital canvas. It can create conventional arts based on the brush, for instance, an oil painting. This procedure generally starts from a source picture by sequentially adding strokes to modify that picture and ultimately composing a nonphotorealistic image that looks like the original photograph carrying an exquisite style. The limitation of SBR methods is that all SBR approaches are intended to create a single distinct style and therefore is not proficient for composing arbitrary styles. It illustrates the non-flexibility of this approach.

B. *Region-based techniques*

RBT is used to accommodate the effects based on the content in regions. Previous RBT algorithms [12] deteriorate the shape of structures in managing the arrangement of strokes. The latter method introduces an IBAR procedure [13] based on regions to alter the semantic geometry of photographs for generating creative effects. This algorithm replaces parts of the picture by many canonical patterns for constructing uncomplicated shape rendering results. In this approach, a source image and target photograph are distributed into many areas using the method based on graphs. The semantic conformity is estimated between them based on colors, patterns, and texture. Ultimately, the colors and textures are transferred for each area by employing a patch match method. Drawbacks of RBT methods are similar to SBR, all RBT approaches are intended to create a single distinct style and therefore is not proficient for composing arbitrary styles. It represents the non-flexibility of this approach.

C. *Image-processing and filtering*

Usually, the IPF approach for rendering is simple to implement and effective in application. IPF techniques [14] apply to real-time rendering and accommodate multicore CPU and GPU. Bilateral and Gaussian filters are employed to compose the effects of cartoons. Kuwahara filter is an exciting range of filters for preserving edges. It works relatively well on pictures of high contrast. These filters eliminate features in sharp regions and preserve shape outlines in areas of low contrast. Some other filters are morphological, diffusion, and shock filter. Gradient-domain techniques are also utilized in which the result is interpreted by creating a gradient field. The drawback of IPF is that style variety, and diversity is inadequate.

D. *Example based techniques*

EBR [15] understands the mapping within the prototype pairs containing the source photograph and the corresponding styled artwork. The learned mapping is employed to style arbitrary pictures. EBR can work in two categories by performing texture transformation and color transformation. The proposed framework named image analogy is a supervised method containing two stages, namely the design stage and application stage. In the design stage, it intends to study the analogous mapping between the pairs of source pictures and target outputs. The data used for training in image-analogy contains combinations of input-pictures along with the corresponding stylized output-pictures. In the application phase, learned mapping is applied to new target photographs to produce relevant analogous results. The main drawback of this method is the unavailability of training data. This method fails to capture the structure of images accurately and uses only low-level features of the picture.

III. NEURAL STYLE TRANSFER

NST is a machine learning technique that takes two images in input, one of which is the content-input image and another one is called a style-input image, for example, the artwork in a famous painting and blend them to transform the input images into an output image that looks like the picture called content-image but it will be painted similarly to the picture called style-image. Neural style transfer uses deep learning and its convolution neural networks to perform this amazing task. It is much propitious for artists and widely used by them to impose their artwork on digital photographs. It was first introduced and explained by Leon A. Gatys in 2015. After the Gayts, several endeavors are in progress related to the style transfer and its algorithms to perform training and learning faster as much as possible and to extend the procedure of style transfer techniques from static images to videos, audios, and other mediums. Neural methods of style transfer are broadly divided into two categories. One of which is the online method and the other one is the offline method.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:25:38 UTC from IEEE Xplore. Restrictions apply.

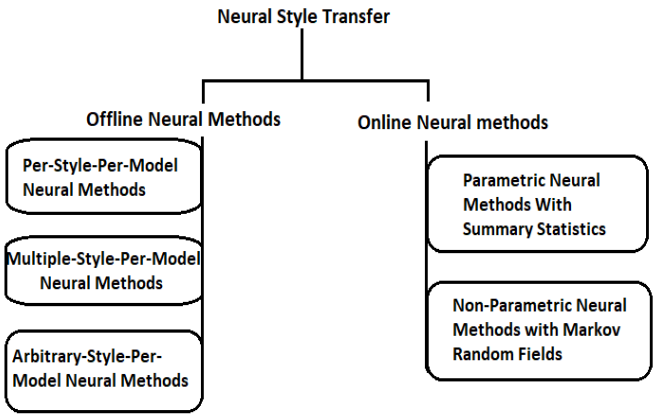


Fig.2: Categorization of NST techniques

A. *Online neural methods*

One of the famous applications in the field of style transfer is Deep-dream, which is based on the online approach of NST. The online method focuses on the optimization of the generated output image but not on the improvement of the applied model. This approach employed several iterations for enhancing the quality of the output image. Due to the multiple iterations, this procedure is much time consuming and computationally expansive. Therefore it is also known as the slow neural method for a style transformation. The process is to extract the content information and style representation from input images and then apply this extracted information to the CNN model for recombining both components to generate the artistic output. Gradient descent algorithm is used for the optimization of the result by minimizing the value of loss function. These online methods differ in the approach of applying the content and style representation to the CNN model so categorized into two classes.

1) Parametric online neural methods: These methods employed spatial representation summary and statistics to generate artworks. Gatys et al. [1] [2] in 2015 proposed an algorithm that applies deep CNN model VGG 19 for extracting structure, shape, and texture, etc. related information of content and style features from preprocessed input pictures. This information is spatially modeled in different layers of VGG 19 and used for the reconstruction of target artwork. VGG nineteen network is composed of sixteen layers of convolution and five layers for pooling, where higher layers are handled to represent the content characteristics. Style-image input is passed through the VGG network, and the gram matrix is employed to store style characteristics. The overall loss is a linear combination of both content and style losses balanced by parameters calculated through the one thousand iterations of gradient descent. Content loss is computed in terms of mean squared difference. The α and β hyperparameters provide the appropriate weights to both components of loss for balancing the equation of total loss function.

$$L_{total} = \alpha L_{content} + \beta L_{style} \quad (1)$$

$$L_{content}(C, G, L) = 0.5 \sum_{ij} \left(a[L](C)_{ij} - a[L](G)_{ij}\right)^2 \quad (2)$$

$$L_{GM}(S, G, l) = \frac{1}{4N_l^2 M_l^2} \sum_{ij} \left(GM[l](S)_{ij} - GM[l](G)_{ij}\right)^2 \quad (3)$$

$$L_{style}(S, G) = \sum_{l=0}^{L} w_l \times L_{GM}(S, G, l) \quad (4)$$

Here S, C, G, and GM are used to represent style-image, content-image, output generated-image, and gram-matrix, respectively. Activation of layer L is used for the calculation of content-loss. N_l is the number of channels in the feature map, and M_l is the product of the height and width of the feature map. After that demystifying approach [16] of NST was introduced that employs a different method for the statistical representation of style features. In this procedure, the gram matrix was replaced by MMD. It states that the minimization of maximum-mean-dispersion is comparable to the matching of gram-matrices. Several different methods for the representation of style were used, which include Gaussian, linear, polynomial and batch-normalization, etc. In 2017, Lapstyle NST was presented [17], which added a laplacian-loss function to calculate the overall loss. This laplacian-loss preserves the low-level fine details of the content characteristics and is aimed to reduce the distortion of edges, shapes, structure, contour, and colors of original input when converted to artistic output.

2) Non-parametric online neural methods: In 2016, Li and wand [18] proposed a method of style transformation that employed a combination of Markov-random-field models and deep CNN for synthesizing two-dimensional artworks. This approach works well with both photo-realistic and non-photo-realistic transformation of style. For minimizing loss, it uses limited memory BFGS algorithm. This algorithm matches style on the level of neural-patches. Hence, it is suitable for preserving the fine details of the structure. They formed a new function for calculating the style-loss and content-loss by calculating Euclidean distance between patches using the energy function of MRF-neural-patch-matching. The characteristic of this algorithm is that it produces considerably well-resulting artwork for photo-realistic styles, especially if the shape, structure, and panorama of both content-image and style-image are alike. Though, it usually doesn't work well if there is a great difference in view and arrangement of structures of content-image and style-image because the patches in both pictures could not be paired accurately.

B. *Offline neural methods*

Earlier explained the online method uses the iterative approach of gradient descent for improving the quality of generated output. Therefore it consumes a large amount of computation time when image size is considerably large. To overcome this limitation of speed, offline neural methods were introduced. This approach is much speedy compared to the previous procedure because it generates the output artwork in a single feed-forward-pass by optimizing the adopted model. As it mainly focuses on the optimization of the applied model, hence it is also known as the model-optimization-based offline method of NST. This offline approach is further divided into three categories based on the number of styles generated by the model in its output artwork.

1) The offline neural model generating single style-PSPM: For adopting an individual style, this approach optimizes the employed model. Johnson et al. [3] proposed a parametric method of training a feed-forward-network for the task of style transformation. This method replaces the per pixel losses by perceptual losses computed using high-level features of

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:25:38 UTC from IEEE Xplore. Restrictions apply.

images. The system contains two components of the network, which are image-style transformation network and loss-network. Image-style transformation network is a deep residual CNN that can convert input images to an artistic output by understanding the mapping. It is trained using stochastic-gradient-descent for minimizing the weighted combination of loss-functions, which is computed by employing loss-network. Perceptual losses measure the similarities and dissimilarities of images more precisely compared to per pixel losses. Another similar approach proposed by Ulyanov et al. [4] uses multi-scale architectures that can be trained faster and result in a smaller loss in content and texture characteristics and better visual quality while working with comparatively few parameters. These are very lightweight models that are capable of generating artworks comparable to the quality of output produced by Gatys et al., and their speed is almost 100 times faster. However, these are also limited in considering the small details and depth information. After that, a new advance concept was introduced, which is known as instance-normalization [5]. It means normalization is applied to every single image in place of the batch of images. It converges faster, learns quickly, and significantly improves the visual quality of generated stylized artwork. This model can be trained rapidly, reduces the overall loss, and enhances the diversity of network output.

The non-parametric approach of single-style per model offline neural methods uses MGAN [19]. It is a markovian model applied to generative adversarial networks for learning the mapping between different representations of the same content-image. MGAN utilized the imagenet dataset for adversarial-training. It successfully preserved the coherent texture of compact content-images, and the speed of synthesis is remarkably fast. MGAN network consists of two subnetworks, namely discriminator and generator, for improving the model in iterations. This method lacks in transferring non-texture style, such as facial-features of two different face-pictures. Facial-features cannot be interpreted as texture because these require the understanding of expression, poses, and gender, etc. semantic characteristics.

2) The offline neural model generating multiple styles-MSPM: Above mentioned PSPM procedures can generate artistic output hundred times speedier than previous offline neural methods based on image-optimization. But it needs to train several network models individually to generate output for every distinct style-image. It exhibits inflexibility and excessive time consumption in training for each style. For reducing the irrelevant time consumption during the separate training of various models in PSPM, a new approach was introduced for consolidating multiple-styles into a single model, which is known as MSPM. In 2017, Dumoulin et al. [20] proposed that it is possible to model different styles at the same time by using the same parameters in the convolutional network. For incorporating different styles, shifting and scaling of parameters is required in the layers of instance normalization, which is called conditional-instance-normalization. CIN is an alteration of style-transformation networks that is uncomplicated, scalable, flexible, and efficient for incorporating diversity in output artwork. By consolidating affine parameters of different styles, it can be applied to incorporate multiple-styles in one output artwork.

After this style-bank method [21] was introduced, that is an aggregate of various convolution filter-banks. One style is expressed by every filter-bank explicitly in that combination. For converting a picture to a specific style, the corresponding filter-bank is turned over the intermediate feature-embedding generated by a single autoencoder. While keeping a fixed autoencoder, it is possible to carry out incremental-learning for appending a new style by training a new filter-bank. This style-bank fusion approach works in two ways, linear and region-based. In linear fusion, the style-bank layer is fed with a linear combination of multiple styles. While in region-based fusion, content-image is disintegrated into some disjoint regions using automatic k-means clustering, and then combined style transformation can be conducted using more than one style simultaneously. But the size of the model expands as the number of styles learned by it increases, which is a drawback. To eliminate this limitation, Zhang and Dana [22] explored the capabilities of a single CNN model VGG 16 for consolidating both content and style in one network. The Multiple-style-generative network was introduced, which can control the size of brush strokes in the real-time transfer of style. In MSG-Net, the comatch layer was used to represent styles in two-dimensional ways. For the comatch layer, it is possible to differentiate and train end to end, but tough to conserve the fine details of structures.

3) The offline neural model generating arbitrary styles-ASPM: The third type of offline neural method is ASPM that intends to build one model for all styles. It states that a single model can be trained to transfer any arbitrary style. ASPM is also divided into two types. Parametric-ASPM employed spatial representation summary and statistics to generate artworks while nonparametric-ASPM applies MRF for texture-modeling.

Style-Swap: Chen and Schmidt [6] proposed the first algorithm for nonparametric-ASPM in which they introduced the concept of style-swap for training an inverse network to generate the stylized output for any arbitrary style. At first, a set of patches having sufficient overlap for both content and style activations are extracted for computing in a previously trained VGG network. Each content patch is matched with style patches to find the closest and similar match for swapping. Then content activation is reconstructed by averaging areas of overlapping. This approach of one-model-for-any-style is more flexible than the previous methods but generates less appealing results in which the content is preserved thoroughly, whereas the style is not reflected adequately.

AdaIN: In parametric-ASPM, Huang, and Belongie [8] modified conditional-instance-normalization (CIN) to adaptive-instance-normalization (AdaIN) instead of training a network for parameter prediction. The above-mentioned style swap method takes a considerable amount of time and occupies much memory, but AdaIN is a simplistic approach with the low cost of computation. For the given input pictures, AdaIN modifies the mean value and variance value of both content-image and style-image for matching. AdaIN layer efficiently

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:25:38 UTC from IEEE Xplore. Restrictions apply.

transfers the feature-characteristic statistics for performing style transformation. After that, a decoder is used for inverting the output of the AdaIN layer, and the VGG encoder is utilized for measuring loss. It is a good parametric-ASPM approach for performing stylization in real-time, but it is insufficient in the generalization of new styles. Therefore, it is challenging to synthesize intricate style patterns consist of fine details and complex structures.

WCT: Li et al. [7] changed the AdaIN-layer of the above-described model by a combination of the whitening-color-transformation network. The objective is to protect the arrangement and composition of content. Whitening-transformation receives content-activations from the encoder and then produces a filtered representation of content-image. After that, coloring-transformation is applied for consolidating the style-patterns into the filtered content design, and the decoder provides stylized output artwork. This method is good in generalizing distinct styles as learning is not required of every style, but not useful in delivering sharp-features and small strokes-details.

Avatar-net: It blends the images based on the semantic-arrangement of content and employs multi-scale stylization in a single feed-forward pass [9]. It is better than WCT that requires several recursive feed-forward passes. Avatar-net uses a style-decorator and an hourglass-network. Patch-based style decorator module is used to decorate the content features with style patterns while retaining the semantic properties of content. The hourglass network is employed for a multi-scale adaptation of style. But its output depends on the patch size, and therefore usually, it is unable to express the local and global styles simultaneously.

SANET Model: It stands for the style attentional network that can blend style-patterns in the content-image efficiently and flexibly [23]. It uses a kernel that can learn similarities rather than fixed kernels. This neural-network model modifies the self-attention mechanism to understand the mapping between content and style features. For synthesizing artworks, this model takes two images in input, which are content-image I_c and style-image I_s. By applying the encoder module, which is a previously trained VGG nineteen network, both the input-images are encoded into an intermediate representation. After encoding both input-images, both the obtained feature-maps F_c and F_s, are applied to two distinct style attentional networks to produce the combined feature-map F_{cs}.

$$F_c = E(I_c) \tag{5}$$

$$F_s = E(I_s) \tag{6}$$

$$F_{cs}^{r41} = SANET_{r41}(F_c, F_s) \tag{7}$$

$$F_{cs}^{r51} = SANET_{r51}(F_c, F_s) \tag{8}$$

$$F_{csc}^{r41} = F_c + W_{cs} F_{cs}^{r41} \tag{9}$$

$$F_{csc}^{r51} = F_c' + W_{cs}' F_{cs}^{r51} \tag{10}$$

$$F_{csc} = conv_{3\times3}\left(F_{csc}^{r41} + upsamle\left(F_{csc}^{r51}\right)\right) \tag{11}$$

$$I_o = D(F_{csc}) \tag{12}$$

Here F_{csc}^{r41} and F_{csc}^{r51} are the synthesized feature-map values generated by applying 1xl conv-net to the output feature-maps of SANETs. F_{csc} is the integrated output feature-map achieved by applying 3x3 conv-net to the resulting feature-maps obtained by relu_4_1 SANET and upsampled relu_5_1 SANET. The combined output of these SANETs is concatenated by passing through a symmetric-decoder, which produces the final output-image I_o. This method is both effective and practically efficient. It flexibly performs style transformation using the loss function that unites traditional style-reconstruction-losses and the newly introduced identity-loss.

$$L = w_c L_c + w_s L_s + L_{id} \tag{13}$$

$$L_c = \left\| \overline{E(I_o)^{r41}} - \overline{F_c^{r41}} \right\|^2 + \left\| \overline{E(I_o)^{r51}} - \overline{F_c^{r51}} \right\|^2 \tag{14}$$

$$L_s = \sum_{i=1}^{L} \left\| \mu(\varphi_i(I_o)) - \mu(\varphi_i(I_s)) \right\|^2 + \left\| \sigma(\varphi_i(I_o)) - \sigma(\varphi_i(I_s)) \right\|^2 \tag{15}$$

$$L_{id} = w_{id1}(\|I_{cc} - I_c\|^2 + \|I_{ss} - I_s\|^2) + w_{id2} \sum_{i=1}^{L} \begin{pmatrix} \|\varphi_i(I_{cc}) - \varphi_i(I_c)\|^2 \\ + \|\varphi_i(I_{ss}) - \varphi_i(I_s)\|^2 \end{pmatrix} \tag{16}$$

Here L_c is content-loss, L_s is style-loss and w_c, w_s are the weights assigned respectively. $\overline{F_c^{r41}}$ and $\overline{F_c^{r51}}$ are channel-wise-normalized mean, variance feature-map values of content-image. $\overline{E(I_o)^{r41}}$ and $\overline{E(I_o)^{r51}}$ are channel-wise-normalized mean, variance feature-map values of output-image after passing through the encoder. φ is the feature-map value of the layer of encoder denoted by i. L_{id} is identity-loss and w_{id1}, w_{id2} are the weights assigned to identity-loss. This loss function helps to preserve the structure of content-image while simultaneously maintains the style-patterns.

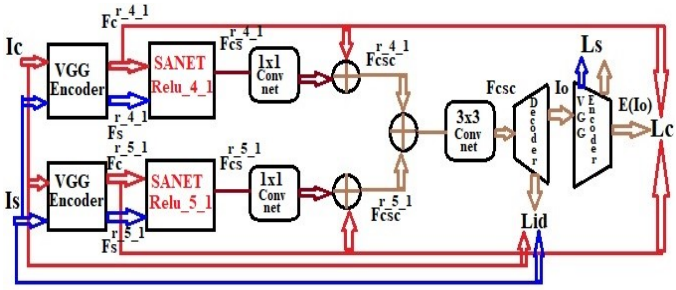


Fig.3: Network Architecture of SANET model

IV. OPTIMIZATION AND EXPANSION OF NST

Optimization algorithms are applied to optimize the output of NST by minimizing the value of loss-function. They work during training and learning of the model to update the weights and finding the relevant values for parameters to get an optimal solution. In Table 1, some popular optimization methods [24] are discussed with their merits and drawbacks.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:25:38 UTC from IEEE Xplore. Restrictions apply.

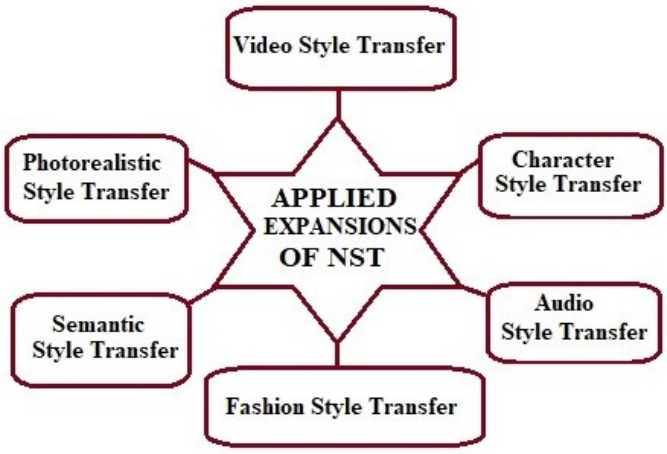


Fig.4: Expansion of NST in various applications

approach [26] was suggested that applies a flow-subnetwork to compose feature-flow and to consolidate the knowledge of optical-flow. This model is based on a network that contains a previously trained combination of encoder and decoder. The semantic transformation [27] of style is an expansion of NST, which is applied to a pair of input images holding analogous content. The method is to obtain a semantic correlation between content-image and style-image. Hence, the style of every region of style-image is applied to the corresponding semantically analogous part of the content-image. Style transformation is also operated on characters [28] for producing unique fonts and unusual text-effects. By collectively training the conditional-generative model and an ornamentation-network model, the style transformation of characters can be accomplished. Another expansion is to generate distinct sound effects by allotting the desired style from the target-audios [29]. It uses two approaches based on iterative optimization of audio and optimization of the employed model. Photorealistic transformation [30] refers to

Table 1: Analysis and comparison of optimization algorithms

Type	Description	Merits	Drawbacks
Gradient Descent-optimizer	It is a traditional algorithm for the optimization of neural network models. It updates weights and tunes the parameters of the model iteratively to minimize the value of loss function.	Updates weights in neural network models iteratively.	It is very slow and takes ample computation time for huge datasets. It is difficult to regulate the learning rate as the algorithm sometimes stuck in local minima.
AdaGrad-optimizer	It applies a separate learning-rate for each parameter. It modifies the learning rate of the parameter at each step based on the computation of the previous gradients of that parameter.	Very helpful for handling scattered data and it is not required to set the learning rate manually.	Its main limitation is the decrease and decay in its learning rate value.
AdaDelta-optimizer	It is an improvement of AdaGrad-optimizer. It entirely replaces the learning-rate parameter by the exponential moving-average of squared-deltas.	No need for the learning-rate parameter.	There is a necessity of two-state variables for saving the second-moments of gradients and the variation in parameters.
Adam-optimizer	It is an adaptive-method for defining separate learning-rates for every parameter using momentum and therefore termed as adaptive-moment-estimation.	It separately defines the learning-rate for every parameter.	Its gives better result in the starting of training, but performance diminishes as the time increases. Therefore, for some tasks, it does not converge to an optimum value.
L-BFGS-optimizer	It is a limited-memory-BFGS algorithm in the family of QUASI-Newton-methods because it takes a small amount of memory. It uses a line-search approach, which makes it much steady in training and more accessible to examine for convergence.	It is the most efficient, speedy, and popular approach of optimization compared to previous methods.	It does not scale well while updating and calculating gradients. Hence, there is a necessity of mini-batch-training that needs to estimate gradients on small subsets of data.

Several approaches of NST are expanded to style transformation of audio, video, and characters, etc. For altering the style of videos, the requirement is to achieve a smooth variation among contiguous frames of video. A method [25] based on iterative optimization of images uses temporal-consistency loss obtained by calculating optical-flow. It produces stylized videos by excluding temporal-artifacts but requires significant time for computing results. Another

the transfer of style of arrangements of colors. Its primary intent is to conserve the fundamental structure of the content-image by utilizing a two-stage optimization method, namely stylization, and smoothing. Hence, for decreasing the deformation of content-image, it employs photorealism-regularization. Another expansion of NST approaches is in the field of fashion [31] to create clothes containing the desired styles of fashion. In the synthesized output, the applied GAN-

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:25:38 UTC from IEEE Xplore. Restrictions apply.

model should protect the basic design of cloth while blending with the given target style.

V. IMPLEMENTATION RESULTS AND ANALYSIS

The characteristics of various distinct procedures for neural style fusion and transformation are thoroughly analyzed [32]. The charts given below in Fig.5 and Fig.6 represent the comparison of running time performance on input-images of distinct dimensions. It has been observed that the algorithm based on optimization of image proposed by Gatys et al. [1] takes a considerable value of computation time when compared with offline methods based on optimization of employed models [3] [4] [5]. However, the quality of generated results is better when compared to others. For real-time applications of arbitrary style transformation, there is a trade-off between the quality of results and the running time of existing algorithms. Therefore, a balance should be maintained between quality and speed.

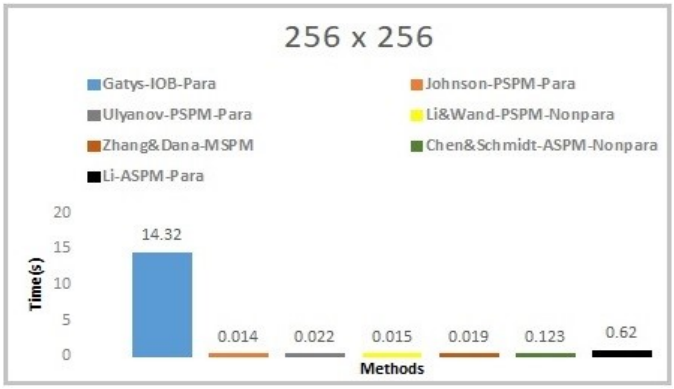


Fig.5: Comparison of the run-time of some NST approaches on 256×256 dimensional images

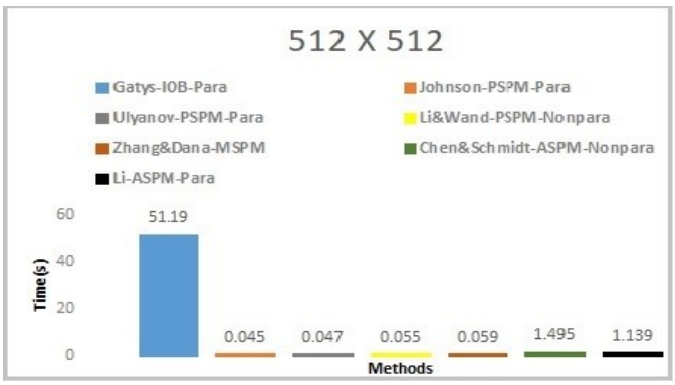


Fig.6: Comparison of the run-time of some NST approaches on 512×512 dimensional images

A network of the SANET model is deployed using Anaconda, PyTorch 0.4.1, CUDA 9.2, and open CV 3.4.2 python. During the training of the model for content-images VOC 2012 dataset is used, and for the style-images WikiArt dataset is used. A PC comprising Windows 10, Intel Core i5 eighth-generation quad-core CPU, NVIDIA Geforce GTX 1050 GPU, and a RAM of size 8 gigabytes is used. It also demands an internet of extremely high-speed. Programming Language used is python. The input is taken as two pre-processed images and resized them to 512×512. During training, the size of each batch is six, and 180000 default iterations are applied. The initial values assigned to weights while computing the loss are as follows, w_c=1.0, w_s=4.0, w_{id1}=50, and w_{id2}=1. It has been perceived that, if the weight w_c is increased in the absence of identity-loss function, then spatial-distribution and structure of content-image can be protected but with a lack of style-features. However, if the weights for identity-loss function are increased, then it's possible to preserve the semantic-arrangement of content along with sustaining the style-features. Consequently, the total loss is estimated by incorporating identity-loss along with content and style losses. Adam optimizer is applied to minimize the value of loss function. The encoder uses a previously trained VGG nineteen network. Two SANET models are jointly trained and integrated by taking the VGG feature-maps as inputs and encodes output feature-maps using distinct layers Relu_4_1 and Relu_5_1. Outputs of these SANETs are combined by using the symmetric decoder. After that, the trained network of SANET models is applied to the pictures taken by us on our college campus to examine the execution time and quality of the obtained results.

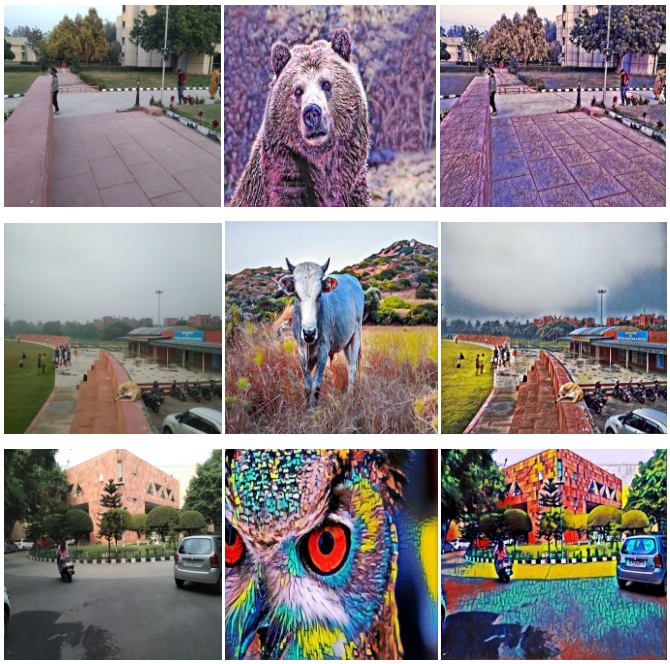


Fig.7: Experimental results of style transformation using SANET model

For determining the quality of the SANET network, its comparison is analyzed with four types of models for fusing and transforming arbitrary-styles. It is observed that AdaIN [8] is the fastest ASPM approach, but it renders a suboptimal result as it copies the channel wise mean-variance value, and the trained decoder usually attach frequently seen similar patterns and style-textures to all stylized outputs. Hence, its resulting output is less appealing and sometimes fails to retain the distribution of colors. Despite matching the second-order-statistics optimally in WCT [7], sometimes, it creates unseen and collapsed-patterns. Avatar-Net [9] blends the images based on the semantic-arrangement of content and employs multi-scale stylization. Still, because of its dependence on the patch size, it fails to realize the local and global style-patterns

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:25:38 UTC from IEEE Xplore. Restrictions apply.

simultaneously. Seldom Avatar-Net and WCT render the blurred hair-texture and distorted vision of content due to the collapsed brush strokes. In contradiction, the network containing two integrated SANET models can generate various exquisite styles efficiently.

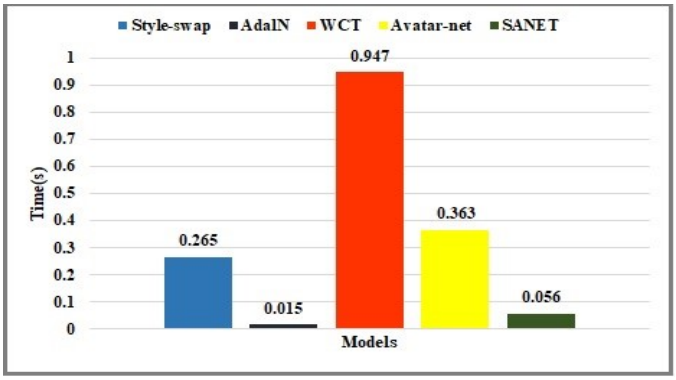


Fig.8: Comparison of execution time of SANET with other ASPM models

The graph in Fig.8 displays a comparison of the running time of SANET with other existing methods of arbitrary style transformation for images of 512×512 dimensions. SANET is much faster than style-swap, Avatar-net, and WCT models but a bit slower in comparison to AdaIN. However, the quality of the result surpasses AdaIN and others. Therefore, this method is both effective and practically efficient.

VI. CONCLUSION

This paper presents an advanced network containing a combination of two SANET Models for style fusion and transformation. This network ensures the semantic-arrangement and spatial-distribution of content-image while fusing with the characteristics of new arbitrary style-patterns by computing a new identity-loss component. The execution time of different NST algorithms are compared on input-images of distinct dimensions. It is observed that online neural models are slow but deliver better results in comparison to fast offline neural models. In all existing approaches, there is a trade-off between the quality of results and the running time of existing algorithms. Our results demonstrate that this model is much faster than other ASPM models and produces results with better quality and flexibility, which makes it much effective and practically efficient. They are planning to expand this model for the style transformation of videos in the future.

REFERENCES

[1] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2414–2423, 2016.

[2] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," in ArXiv e-prints, Aug. 2015.

[3] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for realtime style transfer and super-resolution," in European Conference on Computer Vision, pp. 694–711, 2016.

[4] D. Ulyanov, V. Lebedev, A. Vedaldi, and V. Lempitsky, "Texture networks: Feed-forward synthesis of textures and stylized images," in International Conference on Machine Learning, pp. 1349–1357, 2016.

[5] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6924–6932, 2017.

[6] T. Q. Chen and M. Schmidt, "Fast patch-based style transfer of arbitrary style," in Proceedings of the NIPS Workshop on Constructive Machine Learning, 2016.

[7] Y. Li, C. Fang, J. Yang, Z. Wang, X. Lu, and M.-H. Yang, "Universal style transfer via feature transforms," in Advances in Neural Information Processing Systems, pp. 385–395, 2017 .

[8] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in Proceedings of the IEEE International Conference on Computer Vision, pp. 1501–1510, 2017.

[9] L. Sheng, Z. Lin, J. Shao, and X. Wang. Avatar-Net: Multiscale zero-shot style transfer by feature decoration. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8242–8250, 2018.

[10] J. E. Kyprianidis, J. Collomosse, T. Wang, and T. Isenberg, "State of the 'art': A taxonomy of artistic stylization techniques for images and video," IEEE transactions on visualization and computer graphics, vol. 19, no. 5, pp. 866–885, 2013.

[11] A. Hertzmann, "Painterly rendering with curved brush strokes of multiple sizes," in Proceedings of the 25th annual conference on Computer graphics and interactive techniques. ACM, pp. 453–460, 1998.

[12] B. Gooch, G. Coombe, and P. Shirley, "Artistic vision: painterly rendering using computer vision techniques," in Proceedings of the 2nd international symposium on Non-photorealistic animation and rendering. ACM pp. 83–ff, 2002.

[13] Y.-Z. Song, P. L. Rosin, P. M. Hall, and J. Collomosse, "Arty shapes," in Proceedings of the Fourth Eurographics conference on Computational Aesthetics in Graphics, Visualization and Imaging, Eurographics Association, pp. 65–72, 2008.

[14] H. Winnemoller, S. C. Olsen, and B. Gooch, "Real-time video abstraction," in ACM Transactions On Graphics (TOG), vol. 25, no. 3. ACM, pp. 1221–1226, 2006.

[15] M. Zhao and S.-C. Zhu, "Portrait painting using active templates," in Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering. ACM, pp. 117–124, 2011.

[16] Y. Li, N. Wang, J. Liu, and X. Hou, "Demystifying neural style transfer," in Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, pp. 2230–2236, 2017.

[17] S. Li, X. Xu, L. Nie, and T.-S. Chua, "Laplacian-steered neural style transfer," in Proceedings of the 2017 ACM on Multimedia Conference. ACM, pp. 1716–1724, 2017.

[18] C. Li and M. Wand, "Combining markov random fields and convolutional neural networks for image synthesis," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2479–2486, 2016.

[19] C. Li and M. Wand, "Precomputed real-time texture synthesis with markovian generative adversarial networks," in European Conference on Computer Vision, pp. 702–716, 2016.

[20] V. Dumoulin, J. Shlens, and M. Kudlur, "A learned representation for artistic style," in International Conference on Learning Representations, 2017.

[21] D. Chen, L. Yuan, J. Liao, N. Yu, and G. Hua, "Stylebank: An explicit representation for neural image style transfer," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1897–1906, 2017.

[22] H. Zhang and K. Dana, "Multi-style generative network for real-time transfer," arXiv preprint arXiv:1703.06953, 2017.

[23] D. Y. Park and K. H. Lee., "Arbitrary style transfer with style-attentional networks," arXiv preprint arXiv:1812.02342, 2018.

[24] S. Ruder, "An overview of gradient descent optimization algorithms" arXiv preprint arXiv:1609:04747, 2017.

[25] M. Ruder, A. Dosovitskiy, and T. Brox, "Artistic style transfer for videos," in German Conference on Pattern Recognition, pp. 26–36, 2016.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:25:38 UTC from IEEE Xplore. Restrictions apply.

[26] D. Chen, J. Liao, L. Yuan, N. Yu, and G. Hua, "Coherent online video style transfer," in Proceedings of the IEEE International Conference on Computer Vision, pp. 1105–1114, 2017.

[27] A. J. Champandard, "Semantic style transfer and turning two-bit doodles into fine artworks," ArXiv e-prints, Mar. 2016.

[28] S. Azadi, M. Fisher, V. Kim, Z. Wang, E. Shechtman, and T. Darrell, "Multi-content gan for few-shot font style transfer," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[29] P. Verma and J. O. Smith, "Neural style transfer for audio spectograms," in Proceedings of the NIPS Workshop on Machine Learning for Creativity and Design, 2017.

[30] Y. Li, M.-Y. Liu, X. Li, M.-H. Yang, and J. Kautz, "A closed-form solution to photorealistic image stylization," in European Conference on Computer Vision, 2018.

[31] S. Jiang and Y. Fu, "Fashion style generator," in Proceedings of the 26th International Joint Conference on Artificial Intelligence. AAAI Press, pp. 3721–3727, 2017.

[32] Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu, M. Song. "Neural Style Transfer: A Review", IEEE Transactions on Visualization and Computer Graphics,2019.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:25:38 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Dementia Detection Using Machine Learning by Stacking Models

Ram Murti Rawat
Assistant Professor, Department of Computer Science
Delhi Technological University, Delhi, India
rammurtirawat@dtu.ac.in

Mohammad Akram
Department of Computer Science
Delhi Technological University, Delhi, India
mohammadakram077@gmail.com

Mithil
Department of Computer Science
Delhi Technological University, Delhi, India
mithilmk7@gmail.com

Setti Suneel Pradeep
Department of Computer Science
Delhi Technological University, Delhi, India
ernestsolomoninzion@gmail.com

***Abstract*— Dementia is a syndrome which weakens the memory, thinking, behavior, and the ability to perform normal day-to-day activities. Although dementia mainly affects older people, it is not a part of aging. Machine Learning (ML) is considered as one of the contemporary approaches in predicting, identifying, and making decisions without having human involvement. ML is quickly evolving in the medical industry ranging from diagnosis to visualization of diseases and the study of disease transmission. The present study aims to explore the usage of the stacking of two models, namely gradient boosting machine and artificial neural network is used in the prediction of dementia and validate its performance through statistical analysis.**

Keywords*— *Machine learning, OASIS, Artificial Neural Network, Gradient Boosting Machine, Stacking, Ensemble Learning

I. INTRODUCTION

Machine learning (ML) is considered as an integral part of Artificial Intelligence (AI), also a data analysis technique that computerizes the explanatory model structure. In most scenarios, based on the learning method, two types of ML algorithms (supervised & unsupervised) were used [1]. At present, these algorithms are engaging in all the major industries like healthcare, banking, transport, social media, etc. [1] [2]. Above all, the medical industry is advancing quickly with high volumes of information and increasing difficulties in inventory and patient outcomes. Economically developed nations such as the USA, Japan, and European countries are even facing problems with the enormous collection of medical data [3]. However, by using conventional techniques, it is not possible to analyze this significant volume of information because of time consumption and efforts. Therefore, ML techniques are coming up with various algorithms and programs to avoid these issues. Besides that, the selection of a proper algorithm is not an easy task since it depends on multiple factors such as data volume, information type, and outcomes related to industry requirements [1][4].

Nowadays, Machine Learning algorithms are progressively utilized in neuro-imaging studies like a prediction of Alzheimer's disease (AD) from auxiliary MRI. Also, many studies attempted different ML strategies in predicting AD and their causes [3] [5]. In the study of AD prediction and retrieval, a multistage classifier utilizing ML, including Naive Bayes classifier, support vector machine (SVM), and K-nearest neighbor (KNN) is used to group Alzheimer's illness in the more acceptable and effective way [6]. Similarly, a study from Ref. [7], concluding that the utilization of locally linear embedding (LLE) kind of unsupervised learning is utilized to categorize AD based on fundamental MRI data. Besides, some preliminary studies with ML techniques concluded that these methods are valid and accomplish with high precision (up to 98%) in diagnosing clinical events with analysis of patient medical records [8][4].

Despite it, AD is one of the standard types of dementia and associated mostly with older people [9]. In this paper, a brief explanation is given on the prediction of dementia and calculate performance by using a Stack of two models, namely Gradient Boosting Machine and Artificial Neural Network, i.e., Ensemble Learning. Ensemble learning is a ground-breaking ML method that has shown significant advantages in numerous applications. An ensemble with regards to ML can be extensively characterized as an ML framework that is built with a lot of individual base learners as ensemble members working in parallel and whose predicted results are joined to deliver a single output for a given issue. This single output should have better performance on average than any other ensemble member with uncorrelated error on the target datasets [10].

II. LITERATURE REVIEW

The use of AI in the domain of healthcare has seen vast popularity in recent years, albeit it has a long history. In the 1970s, an expert system was designed for aiding progress in healthcare. Continuous research and development are features of this domain. Many researchers have attempted to address

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:52:38 UTC from IEEE Xplore. Restrictions apply.

this topic and a brief overview of which is presented in the forthcoming section [11].

A short survey of the related work is included in this section. Tohka et al. [12] used Alzheimer's disease Neuroimaging Initiative dataset for comparison of various feature selection techniques by using machine learning for dementia for anatomical brain MRI. SVM and Logistic Regression are used as classifiers, and then it is applied post feature selection techniques. It shows that removing age feature improves the average accuracy with all the classifiers, although the author isn't clear about the discoveries of this investigation for other brain diseases [13].

Zhang Y et al. [14] have proposed a computer-aided diagnosis (CAD) system for MRI images of brain-based on Eigenbrains. This approach generates eigenbrain images based on EEG using key slices from the 3D volumetric data from the MRI. Subsequently, Kernel SVM is included and trained by particle swarm optimization algorithm. The proposed approach was effective in AD subject prediction and discriminant brain-region detection in MRI scanning; still, some advantages and disadvantages of eigenbrain also exist, such as it achieves very high classification accuracy, which can be improved further by combining other features. However, it is essentially two-dimensional, which does not reduce the redundancy along the slice direction; also, it needs preprocessing of spatial registration, which costs a large amount of computation power.

Moradi E et al. [15] has proposed a technique for identifying mild cognitive impairment (MCI) as a transitional stage between age-related cognitive decline and AD. A novel MRI – based biomarker that was developed using machine learning strategies. Noteworthy techniques used in this model included using semi-supervised learning on data available from AD patients and performing feature selection using regularized logistic regression, removing aging effects from MRI data before classifier training to prevent possible confounding between changes due to AD and those due to standard aging, and construction of an aggregate biomarker by learning separate MRI biomarker and subsequently combining age and cognitive measures about MCI subjects by applying random forest classifier. Their aggregate biomarker achieved a very high predictive performance. However, it needs preprocessing of each MRI image, which requires 8 min per image on a high-performance system, thus costs a large amount of computation power.

III. EXPLORATORY DATA ANALYSIS

Analysis of the relationship between the various features of OASIS data explicitly through the graph is a crucial step to understand the nature of the data and eventually to develop a machine learning model. The following figures are implemented to achieve the same.

The graph illustrated in Fig.1 suggests that gender and dementia are correlated and that men are more likely to be affected by dementia than women. It also indicates that women having dementia are diagnosed late, meaning that they show symptoms a little late as compared to men.

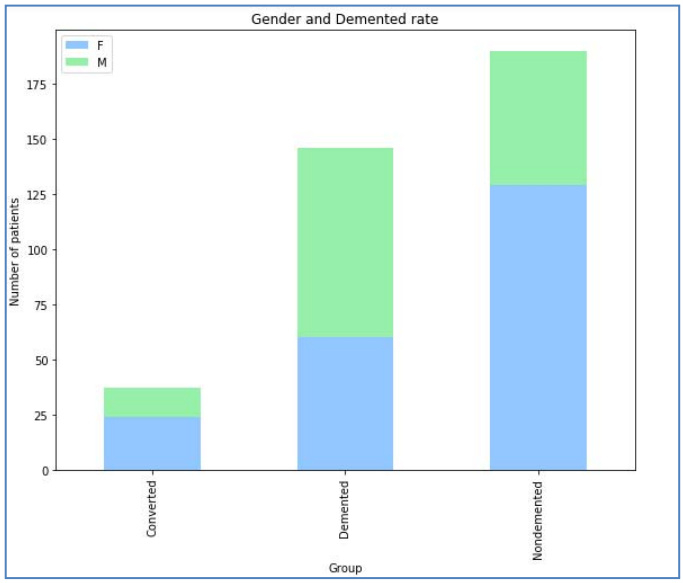


Fig. 1. Gender-wise distribution of dementia

The graph illustrated in Fig. 2 suggests that the Mini-Mental State Examination (MMSE) score of people who have dementia is lowest. In contrast, the people nondemented have high MMSE scores, and the people classified as converted have MMSE scores between those two classes.

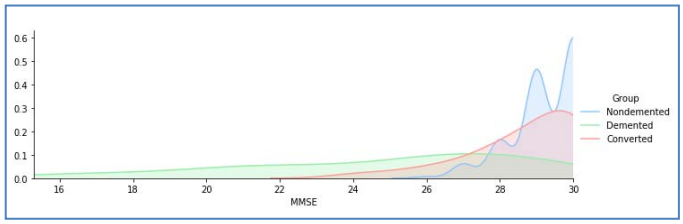


Fig. 2. Distribution of MMSE for nondemented, demented and converted Group

The graph illustrated in Fig. 3 suggests that Atlas Scaling Factor (ASF) value is higher for the converted Group than the demented Group and nondemented Group. This results in the addition of ASF in the feature set.

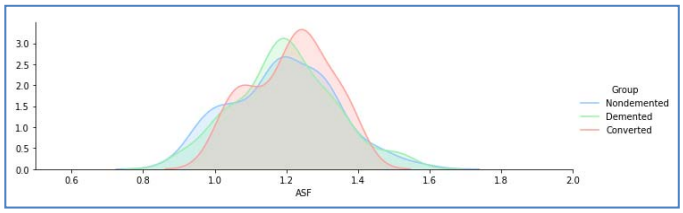


Fig. 3. Distribution of ASF for nondemented, demented and converted Group

The graph in Fig.4 proposes that the estimated Total Intracranial Volume (e-TIV) value is higher for the converted Group than the demented Group and nondemented Group. This results in the presence of e-TIV in the feature set.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:52:38 UTC from IEEE Xplore. Restrictions apply.

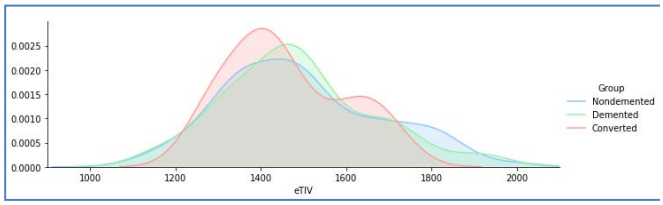


Fig. 4. Distribution of e-TIV for nondemented, demented and converted Group

The graph illustrated in Fig.5 suggests that the nondemented Group has a higher brain volume ratio than the demented and converted Group. This is assumed to be because of the dementia disease that affects the brain resulting in shrinking its tissue.

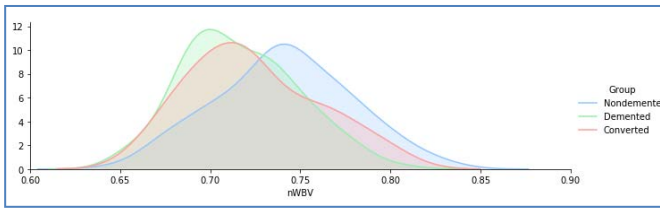


Fig. 5. Distribution of n-WBV for nondemented, demented and converted Group

The following table shows the min, max, mean, and median of the OASIS Dataset features.

TABLE I. OASIS Dataset min, max, mean, median

Feature	Min	Max	Mean	Median
MR Delay	0.0	2639.0	595.11	552.0
EDUC	6.0	23.0	14.6	15.0
SES	1.0	5.0	2.41	2.0
MMSE	4.0	30.0	27.34	29.0
CDR	0.0	2.0	0.29	0.0
eTIV	1106.0	2004.0	1488.13	1470.0
nWBV	0.64	0.84	0.73	0.73
ASF	0.88	1.587	1.19	1.19

IV. DATA COLLECTION AND PROPOSED METHODOLOGY

A. Dataset

A longitudinal collection of OASIS - MRI data set, comprising of demented and nondemented subjects with right-hand (R) type aging from 60 to 96 is considered. A sample size of 150 subjects, including men and women, have attended scanning sessions more than two visits; sessions were separated by at least one year with 373MR Sessions [16]. The sample training data included with demographic values of Subject ID, MRI ID, Group, Visit, MR delay, Sex, Age, Social Economic Status (SES), Education level (EDUC), Mini- Mental State Examination (MMSE), Clinical Dementia Ratio (CDR), estimated Total Intracranial Volume (e-TIV), normalized Whole Brain Volume (n-WBV), and Atlas Scaling Factor (ASF). In particular, some subjects treated as demented at initial visit later transformed into the nondemented managed by converted type. If CDR value is equal to zero, the subjects were considered as mostly nondemented, simultaneously if CDR ≥ 1, the subjects will face the tendency to have dementia [9].

1. *Data collection*

The trained data set collected from the Open Access Series of Imaging Studies (OASIS) included longitudinal MRI data of 150 subjects.

2. *Data Preprocessing*

Real-world data is available more likely incomplete with missing entries. Therefore, data preprocessing is one of the data mining techniques to address this issue. Missing entries were filled-up by the median of particular attribute values.

3. *Attribute Selection*

Select a specific characteristics to predict the outcome to do mapping with input correspondence values. Group column is chosen as an output variable that corresponds to the dementia status based on other input variables.

4. *Input variable matching*

The performance of any ML model largely depends on the number of input attributes taken into consideration. To maintain better performance, the selection of the corresponding attributes, instead of selecting multiple ones is important. Features such as SUBJECT_ID, MRI_ID, Visit, and Hand were removed as they were not affecting the groups in any way. Hand Feature was removed because the dataset consists of MRI of Right-handed people only.

5. *Classifier*

Three groups of dementia are considered they are, demented, nondemented, and converted.

B. Methodology

In this paper, stacking is used, which is an ensemble learning technique that involves combining multiple machine learning models to improve the machine learning results [10]. Similarly, here two machine learning models are used for the classification of Demented, Nondemented, and Converted groups, and the result of one model is added as an additional feature to the existing dataset. This new dataset is fed to the other model to train.

The two models that are used here are Gradient Boosting Machine and Artificial Neural Network. The reason for selecting Gradient Boosting Machine is because of its fast processing and iterative nature to improve from a weak classifier to a robust classifier. It is an ensemble meta-algorithm that primarily reduces bias and also variance, which is an essential criterion to analyze our models and prevent them from over-fitting [13].

Artificial neural networks are found to be effective as general nonlinear models with applications to medical diagnosis, prognosis, and survival analysis. Dementia syndrome is very complicated. The relationship between various risk factors is not dependent; thus, most of the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:52:38 UTC from IEEE Xplore. Restrictions apply.

traditional statistical models are inapplicable. The second model ANN is chosen because of its nonlinear adaptive processing system made of a large number of artificial neurons in the right way to connect each other. The ANN can solve the problems of unambiguous and uncertain medical information more effectively and since it can monitor complex relations; therefore, ANN is selected as a second model here [4].

Our problem is a multi-class classification problem as it contains three classes Demented, Nondemented and Converted thus there are two heuristic methods for multi-class classification, they are:

1. *One-vs-Rest*
2. *One-vs-One*

Here the One-vs-Rest heuristic method is used on Gradient Boosting Machine.

Fig. 6 illustrates the idea of using three different Gradient Boosting Machine for each separate Group – nondemented, demented, and converted.

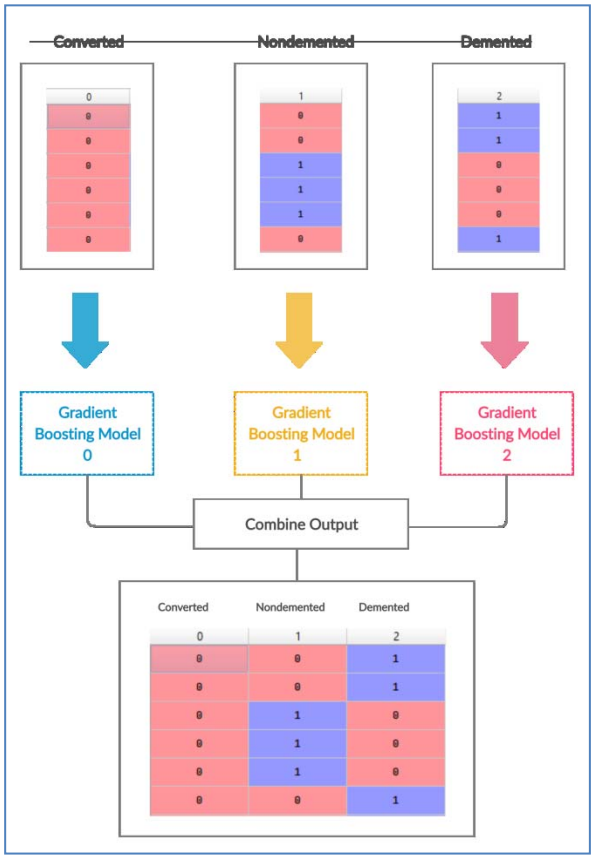


Fig. 6. Training 3 Gradient Boosting Machine

After training on the test data, these three Gradient Boosting Models predict the output for the whole dataset, and their predicted output is combined and added to the original dataset as another feature, which is later fed to the ANN Model for training and testing as illustrated in Fig.7.

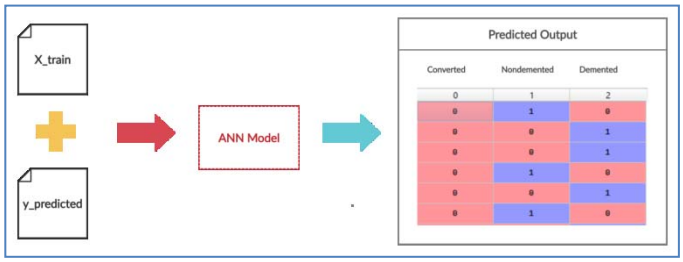


Fig. 7. Predicted Output + Original Dataset fed to ANN Model

V. RESULT

The same data is tested on the Gradient Boosting Model and ANN Model separately, and then the combined version of these two models. The following results are observed:

1. *Gradient Boosting Machine*

TABLE II. Confusion Matrix of given subjects Gradient Boosting Machine TND*: True Nondemented; TD*: True Demented; TC*: True Converted; PND*: Predict Nondemented; PD*: Predict Demented, and PC*: Predict Converted

	TC	TD	TND
PC	3	2	2
PD	1	29	0
PND	4	1	33

```
==> Classification Report -- Gradient Boosting

              precision    recall  f1-score   support

           0       0.38      0.43      0.40         7
           1       0.91      0.97      0.94        30
           2       0.94      0.87      0.90        38

    accuracy                           0.87        75
   macro avg       0.74      0.75      0.75        75
weighted avg       0.88      0.87      0.87        75
```

Fig. 8. Classification Report of Gradient Boosting Model – [0, 1, 2] stands for [Converted, Demented, Nondemented]

2. *ANN Model*

TABLE III. Confusion Matrix of given subjects ANN Model TND*: True Nondemented; TD*: True Demented; TC*: True Converted; PND*: Predict Nondemented; PD*: Predict Demented, and PC*: Predict Converted

	TC	TD	TND
PC	0	5	2
PD	0	30	0
PND	0	1	37

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:52:38 UTC from IEEE Xplore. Restrictions apply.

```
==> Classification Report -- ANN

              precision    recall  f1-score   support

           0       0.00      0.00      0.00         7
           1       0.83      1.00      0.91        30
           2       0.95      0.97      0.96        38

    accuracy                           0.89        75
   macro avg       0.59      0.66      0.62        75
weighted avg       0.81      0.89      0.85        75
```

Fig. 9. Classification Report of ANN Model – [0, 1, 2] stands for [Converted, Demented, Nondemented]

3. *Stack Model - Gradient Boosting Model + ANN Model*

TABLE IV. Confusion Matrix of given subjects Gradient Boosting Machine + ANN TND*: True Nondemented; TD*: True Demented; TC*: True Converted; PND*: Predict Nondemented; PD*: Predict Demented, and PC*: Predict Converted

	TC	**TD**	**TND**
PC	3	2	2
PD	1	29	0
PND	4	1	35

```
==> Classification Report -- GB + ANN

              precision    recall  f1-score   support

           0       0.50      0.43      0.46         7
           1       0.91      0.97      0.94        30
           2       0.95      0.92      0.93        38

    accuracy                           0.89        75
   macro avg       0.78      0.77      0.78        75
weighted avg       0.89      0.89      0.89        75
```

Fig. 10. Classification Report of Stack Model – [0, 1, 2] stands for [Converted, Demented, Nondemented]

From the above results, it is clearly observed that the ANN produced excellent results as it detects the highest number of demented and nondemented groups correctly. Still, the problem with that is its precision and recall for detecting converted Group is zero, as seen in the Classification Report of ANN Model in Fig. 9 and in the confusion matrix in TABLE III; hence this model is not appropriate to choose alone for dementia analysis. The solo Gradient Boosting Model also provides excellent results giving weighted average accuracy of 87% as seen in the Classification Report in Fig. 8 and confusion matrix in TABLE II, but it is observed that the stack model provides the highest weighted average accuracy of 89%, the best precision and recall for all the groups, as seen in its Classification Report in Fig. 10, and the confusion matrix in TABLE IV.

VI. CONCLUSION

This paper introduces a method of detecting dementia from auxiliary MRI reports and a few other features. Gradient Boosting Machine and ANN are exploited as well as their stacking with each other to achieve these tasks. The Neural Network is trained with the new dataset, including the predicted output of the Gradient Boosting Machine, produced the best results as compared to the results produced by both the models' individually.

Dementia is one of the significant health issues that has challenged health experts worldwide. Also, it mostly happened in older people with age > 60. For more than two decades, healthcare professionals are working continuously to find a method to forecast it in the earlier stages and to find a medicine that can cure. Unfortunately, yet there are no proper medicines to cure this disease completely. Many computer scientists are working along with medicinal institutions performing research activities on this disease. Still, there is an extreme need for the identification of relevant characteristics that can forecast the detection of dementia.

It tis then approached with the stacking of models for classification and prediction of dementia and achieved optimized results with efficient performance values. The future work involves the identification of a new set of attributes that can help in detecting early cognitive impairment and dementia.

REFERENCES

[1] Huddleston SH, Brown GG., "Machine learning. Informs analytics body of knowledge", 2018.

[2] Libbrecht MW., Noble WS., "Machine learning applications in genetics and genomics.", Nat Rev Genet 2015.

[3] Ichikawa D., Saito T., Ujita W., Oyama H., "How can machine-learning methods assist in virtual screening for hyperuricemia? A healthcare machine-learning approach.", J. Biomed Inform, 2016.

[4] Wang, Naibo & Chen, Jinghua & Xiao, Hui & Wu, Lei & Jiang, Han & Zhou, Yueping. "Application of artificial neural network model in diagnosis of Alzheimer's disease.", BMC Neurology, 2019.

[5] Kaur P., Sharma M., Mittal M., "Big data and machine learning based secure healthcare framework.", Procedia Comput Sci, 2018.

[6] Kruthika KR., Rajeswari, Maheshappa HD., "Multistage classifier-based approach for Alzheimer's disease prediction and retrieval.", Inf Med Unlocked 2019, Nov. 2018.

[7] Liu X., Tosun D., Weiner MW., Schuff N., "Locally linear embedding (LLE) for MRI based Alzheimer's disease classification.", Neuroimage 2013.

[8] XIE R., Khalil I., Badsha S., Atiquzzaman M., "Collaborative extreme learning machine with a confidence interval for P2P learning in healthcare.", Comput Network, 2019.

[9] Battineni, Gopi & Chintalapudi, Nalini & Amenta, Francesco, "Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM).", Informatics in Medicine Unlocked, 2019.

[10] Yun Yang, "Ensemble Learning" in *Temporal Data Mining Via Unsupervised Ensemble Learning*, Elsevier, 2017, pp. 35-56.[online]. Available: http://www.sciencedirect.com/science/article/pii/B978012811654800004X

[11] Yuvasri Raghavan, Shubham Agarwal, "Comparative Study of Machine Learning Models for Alzheimer's Detection.", IRJET, Oct. 2019.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:52:38 UTC from IEEE Xplore. Restrictions apply.

[12] Tohka, J., Moradi, E., Huttunen, H., & Alzheimer's Disease Neuroimaging Initiative, "Comparison of feature selection techniques in machine learning for anatomical brain mri in dementia.", Neuroinformatics, 2016.

[13] Deepika Bansal, Rita Chhikara, Kavita Khanna, Poonam Gupta. "Comparative Analysis of Various Machine Learning Algorithms for Detecting Dementia.", Procedia Computer Science, 2018.

[14] Yudong Zhang, Zhengchao Dong, Preetha Phillips, Shuihua Wang, Genlin Ji, Jiquan Yang, and Ti-Fei Yuan, "Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning," frontiers in Computational Neuroscience, vol 9, June 2015.

[15] Moradi E., Pepe A., Gaser C., Huttunen H., Tohka J., "Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects," NeuroImage, vol. 104, pp. 398-412, Jan. 2015.

[16] Jacob Boysen, "MRI and Alzheimers," August 16, 2017.[online]. Available: https://www.kaggle.com/jboysen/mri-and-alzheimers

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:52:38 UTC from IEEE Xplore. Restrictions apply.

JJCO Japanese Journal of Clinical Oncology

Japanese Journal of Clinical Oncology, 2020, 00(00)1–11
doi: 10.1093/jjco/hyaa130
Original Article

OXFORD

Original Article

Descriptive epidemiology of incidence and mortality of primary liver cancer in 185 countries: evidence from GLOBOCAN 2018

Rajesh Sharma *

University School of Management and Entrepreneurship, Delhi Technological University, Delhi, India

*For reprints and all correspondence: Rajesh Sharma, University School of Management and Entrepreneurship, Delhi Technological University, Vivek Vihar- Phase II, Delhi-110095, India, E-mail: rajesh.sharma@dtu.ac.in

Received 1 June 2020; Editorial Decision 2 July 2020; Accepted 3 July 2020

Abstract

Purpose: This study aims to examine the burden of primary liver cancer in 185 countries in 2018.
Methods: The estimates of incidence, mortality and prevalence of primary liver cancer were procured from GLOBOCAN 2018. The development status of a country was measured using the human development index—a composite indicator of income per capita, education and life expectancy.
Results: Globally, primary liver cancer resulted in an estimated 781 631 deaths at age-standardized mortality rate of 8.5/100 000, and 841 080 cases were estimated to be diagnosed in 2018. Males accounted for 596 574 cases and 548 375 deaths, which is more than twice the burden of primary liver cancer in females (cases: 244 506; deaths: 233 456). The global age-standardized incidence rate was 9.3/100 000 in 2018, varying from Morocco (1.1/100 000) to Mongolia (93.7/100 000). There were remarkable variations in terms of age-standardized mortality rate, too, which ranged from 1/100 000 in Nepal to 75.4/100 000 in Mongolia. East Asia was the top region contributing 55.6% of global cases and 54.7% of global deaths.
Conclusions: Since majority of the primary liver cancer burden pertains to hepatocellular carcinoma and screening approaches are yet to be fully proven, the policy focus must be on prevention approaches through the hepatitis-B vaccine, early detection of hepatitis-C infection, reduced alcohol consumption, obesity control, reduced aflatoxin exposure and containment of other modifiable risk factors.

Key words: liver cancer, incidence, mortality, mortality-to-incidence ratio, GLOBOCAN

Introduction

Primary liver cancer (PLC) is the fourth leading cause of cancer death among all cancer groups worldwide [1], and its incidence and mortality have increased in the last three decades [2]. PLC has one of the worst 5-year survival rates among 18 cancer groups examined in Concord-3 [3]. Histologically, hepatocellular carcinoma (HCC) constitutes 70–85% of PLC, while 10–25% is intrahepatic cholangiocarcinoma (ICC) [4,5]. Hepatitis-B virus (HBV) and hepatitis-C virus (HCV) contributes around 80% of HCC across the globe [6]. Aflatoxin exposure is another significant risk factor of HCC in Africa and East Asia [7]. The main source of Aflatoxin exposure is fungi, such as Aspergillus Flavas and Aspergillus Parasiticus [8]. Apart from HBV and HCV infections, other risk factors of HCC are alcohol consumption, obesity and non-alcoholic fatty liver disease (NAFLD) [9].

Liver cancer not only exerts substantial burden in adults, it also exerts significant burden in children and adolescents; around 67–80% of pediatric PLC is hepatoblastoma and the rest 20–33% is HCC [10]. The incidence of HCC is greater than hepatoblastoma in East Asia, while the converse is true in the Western countries [11,12]. In hyperendemic countries such as China, HCC in younger children

© The Author(s) 2020. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permission@oup.com.

Downloaded from https://academic.oup.com/jjco/article-abstract/doi/10.1093/jjco/hyaa130/5876941 by Delhi Technological University user on 03 August 2020

is mostly acquired through perinatal exposures [13–15]; in contrast, in low-endemic areas, the common causes of HCC are tyrosinemia and other metabolic disorders [10].

For cancer control policy formulation and implementation at the global, regional and national levels, it is pertinent to understand the epidemiology of PLC. This study envisages to systematically examine the epidemiology of PLC in 185 countries drawing upon estimates from GLOBOCAN 2018 [1,16,17]. This study provides a comprehensive and comparable assessment of PLC burden measured in terms of all-age and age-standardized incidence, mortality and prevalence at the global, regional and national levels. Following previous literature [18–21], the mortality-to-incidence ratio (MIR) is employed as a proxy for 5-year survival rates. I also studied the PLC epidemiology, both age- and sex-wise, in the light of the socio-economic status of countries measured using the human development index (HDI, [22]). While previous studies on PLC epidemiology either focussed on select countries or regions, this is the first study providing a comprehensive assessment of PLC burden in 185 countries for one of the most recent year (2018).

Materials and methods

GLOBOCAN 2018 estimates

The estimates of incidence and mortality of PLC for 185 countries were drawn from GLOBOCAN 2018 [1,16,17]. The estimates of PLC pertained to code C22 (liver and intrahepatic bile ducts cancer) of the international classification of disease (ICD) tenth revision (ICD-10). The GLOBOCAN estimates were available in terms of age-wise and sex-wise incidence, mortality and prevalence (1-year, 3-year and 5-year) for the year 2018. All-age and age-specific incidence and death counts are reported along with 95% uncertainty intervals (UIs) inside square brackets; these UIs reflect uncertainties associated with availability and quality of data from the cancer registry, death registration and timeliness of data reporting [1,16,17].

The age-wise estimates were examined for the 16 age groups: below-15, 15 to 19, 20 to 24, 25 to 29, 30 to 34, 35 to 39, 40 to 44, 45 to 49, 50 to 54, 55 to 59, 60 to 64, 65 to 69, 70 to 74, 75 to 79, 80 to 84 and 85-plus. The PLC estimates were provided by GLOBOCAN for 185 countries in 21 world regions as defined by the United Nations Population Division (UNPD). In addition to incidence, mortality and 5-year prevalence, I also examined the global, regional and national MIR, which is employed as a proxy for 5-year survival rates. The MIRs were calculated as the ratio of all-age death counts and all-age incidence counts of the mean estimates.

Human development index

The comparative performance of countries in terms of various indicators of PLC burden was evaluated vis-à-vis a country's socio-economic status measured by its HDI [22,23]. HDI, an indicator of socio-economic status, is proposed and developed by the United Nations Development Program (UNDP). HDI has been used as an indicator of socio-economic status of countries in previous epidemiological studies too [20,21]; the advantage with this indicator is that it is a composite indicator of three indicators of development, namely, gross national income (GNI) per capita, education (composite construct of mean years of schooling and expected years of schooling) and longevity (life expectancy at birth). The UNDP rescales each of these indicators on a 0 (lowest) to 1 (highest) scale, and the geometric mean of three indicators provides the final value of the HDI as per the following formula [22].

$$\mathrm{HDI} = \sqrt[3]{\mathrm{INCOME} * \mathrm{EDUCATION} * \mathrm{LE}}$$

INCOME denotes income index; EDUCATION denotes education index; LE denotes life expectancy index. HDI values were drawn only for the year 2018, which is consistent with GLOBOCAN estimates. Based on HDI values, countries were categorized by UNDP into four groupings: low HDI (HDI < 0.550), medium HDI (0.550–0.669), high HDI (0.700–0.799) and very high HDI (>0.800).

Results

Global and regional burden

Globally, 841 080 [95% UI, 817 635–865 198] cases of PLC were estimated to be diagnosed at an age-standardized rate of 9.3/100 000 in 2018 (Table 1). The PLC resulted in an estimated 781 631 [737 605–828 285] deaths at an age-standardized rate of 8.5/100 000. Globally, the 5-year prevalence of PLC was 675 210 in 2018. East Asia was the leading region in terms of both incidence and death counts, with 467 327 [459 452–475 337] cases and 427 932 [421 450–434 514] deaths, respectively, in 2018 (Table 1). South-east Asia was the second leading region with 89 010 [83 523–94 858] cases and 88 429 [82 662–94 599] deaths in 2018. The 5-year prevalence was again the highest in East Asia (383 221), followed by South-east Asia (69 441) and South-central Asia (35 246). East Asia also had the highest age-standardized incidence rate (ASIR) of 17.7/100 000, with the lowest ASIR estimated in South-central Asia (2.5/100 000). The age-standardized mortality rate (ASMR) was again the highest in East Asia at 16/100 000, and the lowest ASMR was recorded by South-central Asia at 2.3/100 000 (Table 1).

Country-wise burden

Country-wise, China, USA and Japan were the three leading countries in terms of new cases with 392 868 [386 346–399 500], 37 948 [37 032–38 887] and 35 535 [34 731–36 358] cases, respectively, in 2018 (Supplementary Table S1). In terms of death counts too, China was the leading country with 273 014 [268 132–277 985] deaths, followed by the USA with an estimated 20 564 [19 973–21 172] deaths in 2018. Vietnam was ranked third in terms of death counts with 19 708 [17 974–21 609] deaths, which is higher than Japan (18 539 [17 987–19 108]) even though Vietnam had lesser number of new cases (25 335 [23 628–27 166]) than Japan (35 535 [34 731–36 358]). In terms of 5-year prevalence, China (296 780), Japan (43 875), USA (31 023) and South Korea (25 015) were the leading countries (Supplementary Table S1).

The top three countries in terms of ASIR were Mongolia (93.7), Egypt (32.2) and the Gambia (23.9) with Algeria (1.4), Morocco (1.1) and Nepal (1.1) recording the lowest ASIR (Fig. 1a). The ASMR mostly followed the patterns of ASIR with Mongolia (75.4), Egypt (31.8) and the Gambia (26.2) as the three leading countries and Nepal (1/100 000) had the least ASMR (Fig. 1b). At the global level, the MIR was 0.93, varying from 0.44 in Samoa to 1.6 in Iceland (Fig. 1c).

Among the countries with >100 estimates cases, I also plotted the ASIRs, ASMRs and MIRs as per HDI categories (Fig. 2a–c); this figure illustrates that where ASIR and ASMRs displayed mixed patterns as per HDI levels, in terms of MIR, however, high HDI and very high HDI countries had comparatively lower MIR than low

Downloaded from https://academic.oup.com/jjco/article-abstract/doi/10.1093/jjco/hyaa130/5876941 by Delhi Technological University user on 03 August 2020

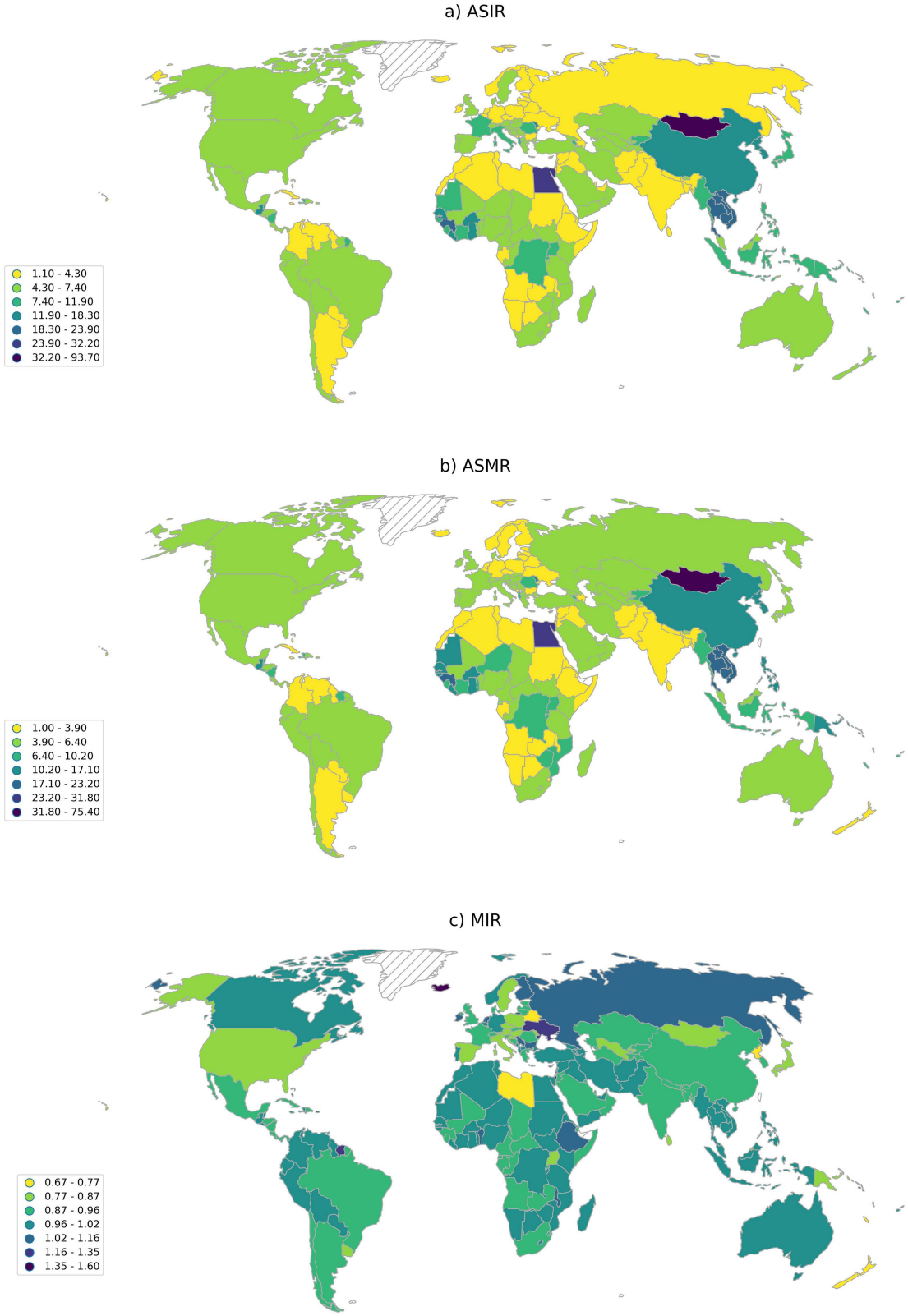


Figure 1. Geographical distribution of primary liver cancer statistics in 2018. (1a) ASIR (1b) ASMR (1c) MIR; ASIR: Age-standardized incidence rate (cases per 100 000); ASMR: Age-standardized mortality rate (deaths per 100 000); MIR: mortality-to-incidence ratio. The MIR was calculated by dividing all-age death counts with all-age incidence. The estimates of liver cancer burden were drawn from GLOBOCAN 2018.

Downloaded from https://academic.oup.com/jjco/article-abstract/doi/10.1093/jjco/hyaa130/5876941 by Delhi Technological University user on 03 August 2020

Table 1. Region-wise burden of primary liver cancer in 2018

Population	Incidence	ASIR	Deaths	ASMR	Prevalence	MIR
Australia and New Zealand	2921 [2796–3051]	5.7	2709 [2591–2833]	4.7	2354	0.93
Caribbean	2923 [2613–3270]	5	2791 [2513–3099]	4.7	2117	0.95
Central America	11 229 [10 870–11 600]	6.3	10 672 [10 376–10 976]	5.9	8453	0.95
Central and Eastern Europe	22 784 [22 150–23 437]	4	22 745 [22 181–23 323]	3.9	16 039	1.00
Eastern Africa	11 550 [8507–15 682]	4.8	11 251 [8067–15 692]	4.7	11 705	0.97
Eastern Asia	467 327 [459 452–475 337]	17.7	427 932 [421 450–434 514]	16	383 221	0.92
Melanesia	915 [789–1061]	11.4	819 [718.7–933.3]	10.6	894	0.90
Micronesia	85 [70–104]	15.2	68 [54–86]	12	24	0.80
Middle Africa	6010 [4325–8351]	6.5	5853 [4248–8064]	6.5	6115	0.97
North America	41 851 [41 406–42 301]	6.6	34 339 [33 925–34 759]	4.8	34 107	0.82
Northern Africa	27 935 [25 427–30 691]	14.1	27 505 [24 873–30 415]	13.9	21 466	0.98
Northern Europe	10 997 [10 693–11 309]	4.7	9997 [9701–10 302]	3.8	7878	0.91
Polynesia	67 [49–92]	9.2	54 [33–89]	7.4	40	0.81
South America	24 248 [21 806–26 964]	4.6	22 973 [22 393–23 569]	4.3	17 225	0.95
South-Central Asia	44 010 [41 162–47 055]	2.5	40 812 [38 208–43 594]	2.3	35 246	0.93
South-Eastern Asia	89 010 [83 523–94 858]	13.3	88 429 [82 662–94 599]	13.2	69 441	0.99
Southern Africa	2710 [2438–3012]	4.9	2597 [2476–2724]	4.7	2344	0.96
Southern Europe	25 026 [23 934–26 168]	6.8	21 996 [21 420–22 587]	5.3	17 262	0.88
Western Africa	16 574 13 547.8–20 276]	8.3	16 356 [13 212–20 248]	8.2	15 106	0.99
Western Asia	9249 [8248–10 372]	4	9096 [8404–9845]	4	6875	0.98
Western Europe	23 659 [22 949–24 391]	5.3	22 637 [22 095–23 193]	4.5	17 298	0.96
Global	841 080 [817 635–865 198]	9.3	781 631 [737 605–828 285]	8.5	675 210	0.93

Incidence: all-age new cases; deaths: all-age death counts; ASIR: age-standardized incidence rate (cases per 100 000); ASMR: age-standardized mortality rate (counts per 100 000); prevalence: 5-year prevalence; MIR: mortality-to-incidence ratio. The MIR was calculated by dividing all-age death counts with all-age incidence. Data source: GLOBOCAN 2018 (IARC).

and medium HDI countries (Fig. 2c).[1] For instance, Belarus (0.67), New Zealand (0.73) and South Korea (0.73), the very high HDI countries, recorded one of the lowest MIRs. Uganda (0.86) in the low HDI category, Zambia (0.89) in medium HDI category and Libya (0.77) and Mongolia (0.79) in the high HDI category recorded the lowest MIR in their respective categories.

Sex-wise burden

Globally, males accounted for 596 574 [576 976–616 838] cases and females accounted for 244 506 [231 861–257840] cases in 2018. Male ASIR (15.5/100 000) was again higher than the female ASIR (6.5/100 000). In terms of death counts too, males accounted for nearly two-third of fatalities with 548 375 [511 583–587 813] deaths

[1] I also plotted ASIR, ASMR and MIR as per world bank country classification in low, lower-middle, upper-middle and high income countries (Supplementary Figure S1). The results were almost similar as to that of HDI (Fig. 2) classifications with few exceptions. Country classification was based on GNI per capita available at https://datahelpdesk.worldbank.org/knowledgebase/articles/906519 (accessed on 13 May, 2020).

Downloaded from https://academic.oup.com/jjco/article-abstract/doi/10.1093/jjco/hyaa130/5876941 by Delhi Technological University user on 03 August 2020

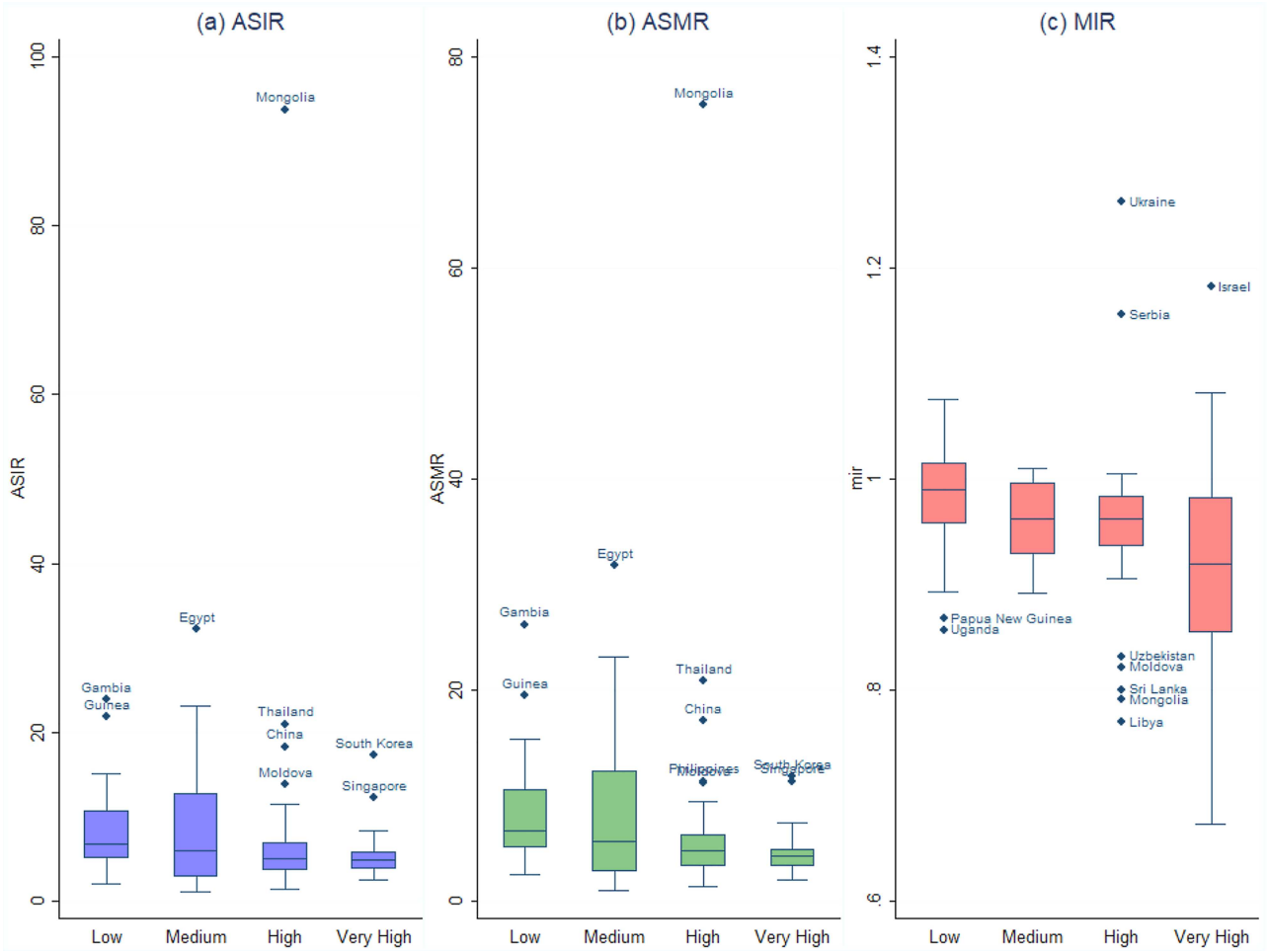


Figure 2. Boxplot of HDI category wise primary liver cancer statistics. (a) ASIR (b) ASMR (c) MIR. The MIR was calculated by dividing all-age death counts with all-age incidence. The MIR was calculated by dividing all-age death counts with all-age incidence. The countries were categorized by United Nations Development Program (UNDP) as low (HDI < 0.550), medium (0.550–0.669), high (0.700–0.799) and very high (>0.800). The estimates of liver cancer burden were drawn from GLOBOCAN 2018.

and the rest of 233 256 [209 922–259 184] deaths due to PLC occurred in females. Male ASMR (14.2/100 000) was also more than double than that of female ASMR (6.2/100 000). The male–female difference in incident cases and death counts was also preserved in regions (Fig. 3a–d). Few regions such as East Asia and South-east Asia had greater disparities in males and females, registering 3-fold difference or more between males and females (Fig. 3a–d). Similar differences were observed across regions in terms of ASIR and ASMR. Few regions in America (e.g. Central America and South America) had lesser male–female disparities in terms of counts as well as age-standardized rates (Fig. 3a–d).

Age-wise distribution

Globally, PLC incidence and deaths peaked in the 64–69 age group with a worsening ratio of death counts and cases in older age groups (Fig. 4). The age-wise distribution of incidence and mortality differed remarkably between different HDI groupings with higher incidence and mortality in older age groups for countries higher in the development spectrum (Supplementary Figure S2). For instance, in low and medium HDI categories, cases and deaths plateaued in 54–59 age groups and declined thereafter, whereas very high HDI category had maximum estimated deaths in the older-most age group (85-plus). Moreover, the low HDI group of countries had a non-negligible portion of cases and deaths in child and adolescent age groups (below the age 20); in contrast, the proportion of PLC cases and deaths in other HDI groupings was much lower in those aged below 20 (Supplementary Figure S2).[2]

Discussion

The PLC cases and deaths increased from 748 000 and 695 000, respectively, in 2008 (GLOBOCAN 2008, [24]) to 841 080 and 781 631, respectively, in 2018. The age-standardized rates underwent a reduction with ASIR falling from 10.8/100 000 in 2008 [24] to 9.8/100 000 in 2018 and the ASMR falling from 10.0/100 000 in

2 Similar results were observed when countries were categorised geographically in 21 world regions. Supplementary Figure S3 illustrates percentage of cases and deaths in four non-overlapping age groups (below-20, 20–29, 50–69 and 70-plus) in 21 world regions. It shows that nearly 30–40% of PLC cases and deaths occurred in the below-50 age groups in all regions of Africa (except North Africa); in contrast, in nearly all European regions, >90% cases and deaths occurred in above-50 age groups (Supplementary Figure S3).

Downloaded from https://academic.oup.com/jjco/article-abstract/doi/10.1093/jjco/hyaa130/5876941 by Delhi Technological University user on 03 August 2020

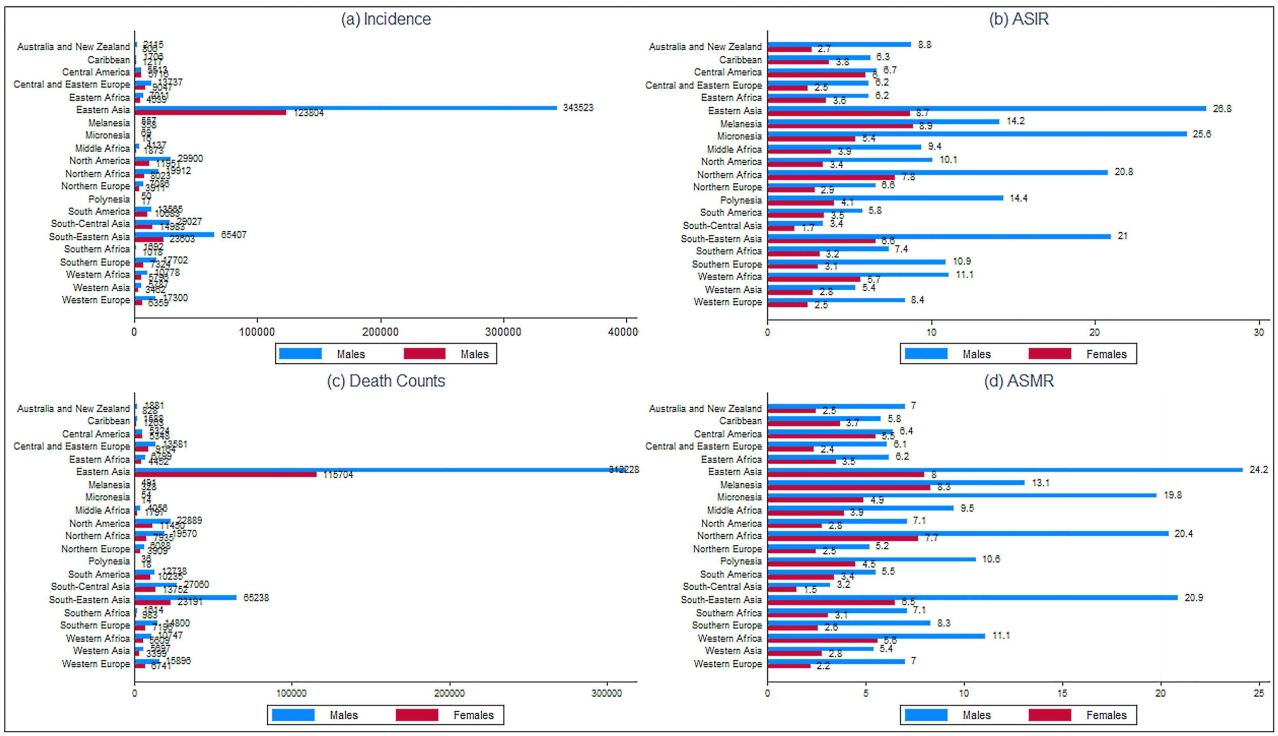


Figure 3. Sex-wise burden of primary liver cancer in 21 regions. Incidence: all-age new cases of liver cancer; deaths: all-age deaths due to liver cancer. The estimates of liver cancer burden were drawn from GLOBOCAN 2018.

2008 to 8.5/100 000 in 2018. PLC has one of the worst survival rates [3], which is also evident from very high MIR of PLC in comparison with other cancer groups. The MIR of PLC was 0.93 in 2018 as per GLOBOCAN 2018 estimates [1], which is only better than that of pancreatic cancer (0.94). In comparison, the MIR of PLC was worse than that of other frequent cancer sites: breast cancer (0.30), prostate cancer (0.28), colorectum cancer (0.48), stomach cancer (0.76) and lung cancer (0.84).[3] East Asia was the top region contributing 55.6% of global cases and 54.7% of global deaths. The age-standardized rates were again the highest in East Asian countries led by Mongolia with ASIR of 93.7/100 000 and ASMR of 75.4/100 000 in 2018.

Concord-3 [3] demonstrated that 5-year survival rates of liver cancer ranged 5–30% across countries, and the highest survival rates ranging between 20 and 29% were observed in South Korea, Singapore and Taiwan. Following previous studies [18–21], MIR was employed as the proxy for 5-year survival rates, and there were marked differences in MIR at the country level. I found that low HDI countries generally recorded high MIRs (i.e. lower 5-year survival rates) and lower MIRs (or higher 5-year survival rates) were recorded by developed countries such as South Korea, Singapore and Japan, which corroborates previous findings of higher 5-year survival rates in developed countries than the developing ones [3].

Males accounted for nearly two-third PLC cases and deaths in 2018, with several regions recording male age-standardized rates triple or more than females (Fig. 3). Such gender-wise disparities were observed in previous epidemiological studies [2] as well as in rodent HCC models, too [25,26]. The male–female differences in PLC incidence are commonly associated with anticarcinogenic effects of female sex hormones [27,28] and lower prevalence of alcohol consumption in females than males [29].

A vast majority of PLC incidence pertains to HCC arising from HBV or HCV infections with HBV infection as the dominant risk factor of HCC in South-east Asia and Africa and HCV infection identified as the leading risk factor of HCC in developed countries such as Japan [30]. HCV infection, for instance, contributes to 70% of HCC cases in Japan [31], which has decreased to 60% recently [32]. World Health Assembly held in May 2016 endorsed Global Health sector strategy 2016–21, which calls for the elimination of viral hepatitis, 96% of which comprises of HBV and HCV, as public health target by 2030 through reduction of new infections by 90% and associated mortality by 65% [33]. Due to increased HBV vaccination coverage, HBV prevalence has decreased across the world in percentage terms over 1990–2005, yet the number of HBsAg positive persons has increased from 223 million to 240 million with high endemicity in Sub-Saharan Africa [34].

Regarding the HCC burden due to HCV infection, the development of direct-acting antiviral drugs (DAAs) has significantly boosted the treatment paradigm of HCV infections [35]. The DAAs, taken orally, target specific proteins, and improves sustained virologic response rate (SVR) to >90% and ~100% in a few instances [35,36]. These DAAs hold enormous potential in developing countries such as India, where HCV treatment rates are low [37]. Moreover, the low-cost generic versions of DAAs make these cost-effective in developing countries such as India [38].

Diabetes mellitus [39,40] and obesity [41] are other identified risk factors of PLC. Obesity and diabetes are not only associated with HCC but are associated with increased risk of ICC, too [42]. The

[3] The MIR for different cancer groups is calculated by the author using GLOBOCAN 2018 estimates [1,16,17].

Downloaded from https://academic.oup.com/jjco/article-abstract/doi/10.1093/jjco/hyaa130/5876941 by Delhi Technological University user on 03 August 2020

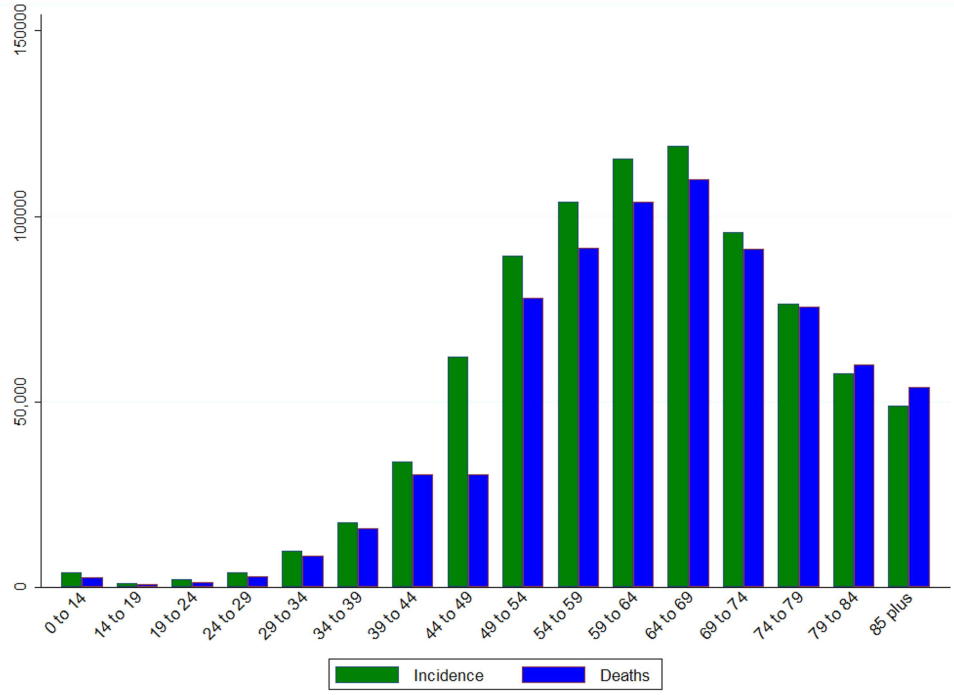


Figure 4. Age-group wise distribution of incidence and mortality of primary liver cancer in 2018. Incidence: new cases; deaths: deaths due to liver cancer. The estimates of liver cancer were drawn from GLOBOCAN 2018.

unabated incidence and prevalence of diabetes mellitus is expected to further result in an increasing burden of chronic liver diseases, including PLC worldwide [43]. In addition to HBV and HCV, the incidence of NAFLD related to HCC has been increasing worldwide, with prevalence varying from 6 to 35% in different countries [44,45].

Moreover, NAFLD has been identified as the fastest-growing cause of cancer-related deaths [46]. The prevalence of NAFLD has increased in parts of East Asia, USA, Europe and Australia [47]. Although the focus on HBV infection and vaccination is expected to result in the downward trend of HCC in Asia, the increasing trend of other risk factors such as obesity, diabetes mellitus and NAFLD is expected to negate the gains from HBV vaccination [31].

China is the leading country in terms of both incidence and death counts. Although China has employed tremendous efforts to shed its leadership position in liver diseases through disease prevention, diagnosis, management and therapeutics [48], still, China was ranked first as per GLOBOCAN 2018 estimates contributing 46.7% cases and 34.9% deaths in 2018. HBV vaccination is fully embedded into the neonatal immunization program in China since 2002 [49]. Despite expanded access to HBV vaccine in China, 20% of children aged below 5 years, and 40% of those aged 5–19 remain unprotected from HBV infection [50]. As described earlier, China may have worked tremendously for expanded access to HBV vaccination; the growth of other risk factors such as obesity may contribute to a greater burden in the future. For instance, China experienced a 414% increase in overweight and obesity between 1982 and 2002 [51].

In several low and medium HDI countries, PLC exerted a significant burden in child and adolescent age groups (Supplementary Figure S2). Pediatric HCC is different from adult HCC in a few key aspects [52,53]. For instance, in adults, HCC arises along with prolonged chronic hepatitis and cirrhosis [54]; in children and adolescents, however, HBV infection is acquired perinatally or due to inherited metabolic disorders [55,56]. Moreover, HCC in children and adolescents is more advanced than in adults [57]. Thus, screening of children whose mothers are HBV carriers is essential in the early detection of pediatric HCC [58].

As far as early screening and detection of liver cancer in adults is concerned, few approaches using ultrasonography, α-fetoprotein (AFP) levels in serum, and surveillance of cirrhosis patients have been considered. The American Association for liver diseases (AASLD) in 2005 issued level 1 recommendation for screening of liver cancer using ultrasonography at six-month intervals for high-risk patients [59]. The AASLD's recommendation, however, was based on a randomized control trial (RCT) conducted in China that demonstrated mortality-reduction benefits of liver cancer screening in the treatment group [60]. The reported benefits of screening, however, were put into question [61]. Currently, major US agencies, which provide recommendations on cancer screening, do not endorse standard or routine screening for liver cancer, as is the case with mammographic screening for breast cancer and colonoscopy for colorectal cancer.[4]

Although, screening for liver cancer hasn't been recommended, United States Preventive Services Task Force (USPSTF) recommends screening for HCV infection in adults aged 20–79 years [64]. It was done due to two reasons. First, there is a sharp spurt in HCV infection in young adults aged 20–39 years in the USA, which coincides with

[4] United States Preventive Services Task Force (USPSTF) has not recommended for liver cancer screening; the American cancer society found no convincing evidence for screening recommendation of PLC [62]; the national cancer institute (NCI) found no convincing beneficial effects of PLC screening [63].

Downloaded from https://academic.oup.com/jjco/article-abstract/doi/10.1093/jjco/hyaa130/5876941 by Delhi Technological University user on 03 August 2020

nation-wide surge in the rate of injection drug use [65]; second, nearly half of HCV patients are unaware of the disease; thereby leading to late diagnosis, which results in life-threatening complications such as cirrhosis and HCC [66].

In summary, apart from worse survival rates, the PLC epidemiology is profoundly different from other frequent cancer groups in a few key aspects. First, geographically other cancer groups such as breast cancer [21], colorectal cancer [67] and stomach cancer [68] are quite dispersed across the world; the PLC, however, is highly concentrated in Asia. Second, while screening modalities are available for breast cancer, colorectal cancer and prostate cancer [20,21,67], there are less-than-ideal or proven screening modalities available for early screening and detection of PLC. Third, a non-negligible number of cases and deaths occurred due to liver cancer in children and adolescents, which is quite different from other frequent cancer groups such as prostate cancer [20], breast cancer [21] and colorectal cancer [67], which predominantly affect comparatively older age groups.

Limitations

There are two limitations of the study. First, the reliability of GLOBOCAN estimates crucially depends upon data availability from population-based cancer registries (PBCRs). In several low- and medium-income countries, however, either PBCRs are not available or lack sufficient coverage. In the absence of reliable and high-quality data from PBCRs, the GLOBOCAN estimates draw strength from data available either from regional registries or from neighbouring countries; the estimates reported and examined in this paper, therefore, are expected to be biased. Second, I employed HDI as an indicator of socio-economic status of a country, which is one of the widely employed indicators of socio-economic status of a country. For a set of countries, with data unavailability, the incidence to mortality ratios (method 3a of incidence) and MIRs (method 3 of mortality) were scaled as per HDI of countries and HDI of respective regions [16].[5] The general patterns of results, however, were similar if GNI per capita was considered as the indicator of development.

Conclusions

In the absence of proven screening devices for PLC, primary prevention through containment of risk factors seems to be the apt choice as a large percentage of PLC cases can be prevented through HBV vaccination, containment of alcohol consumption and obesity control. A large portion of PLC is HCC, and there exists proven vaccination against HBV; it is time to identify high-risk areas and scale up the vaccination in high-risk adults who would have missed the compulsory doses of the HBV vaccine. Moreover, as effective treatment of HCV infection is available through DAAs, early detection of HCV infection must be ensured to prevent HCV-induced sequelae such as cirrhosis and HCC. Lastly, policy makers must keep a strict vigil on the increasing prevalence of risk factors such as NAFLD, obesity and diabetes, which can reverse the gains achieved through HBV vaccination and DAA treatment regimens.

[5] As per Ferlay et al [16], the formula of estimation of such countries is modified as below.Method 3b: $I_{\text{National}} = M_{\text{National}} * \left(\frac{I_{\text{Regional}}}{M_{\text{Regional}}}\right) * \left(\frac{\text{HDI}_{\text{National}}}{\text{HDI}_{\text{Regional}}}\right)$
Method 3: $M_{\text{National}} = I_{\text{National}} * \left(\frac{M_{\text{Regional}}}{I_{\text{Regional}}}\right) * \left(\frac{\text{HDI}_{\text{Regional}}}{\text{HDI}_{\text{National}}}\right)$

Compliance with ethical standards

Disclosure of potential conflicts of interest

The author declare that they have no conflict of interest.

Research involving human participants and/or animals

This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent

The research was conducted using data available in the public domain and did not include any human participants or animals.

Supplementary material

Supplementary material can be found at *Japanese Journal of Clinical Oncology* online.

Acknowledgement

I thank International Agency for Research on Cancer for providing GLOBOCAN 2018 data in the public domain.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394–424.
2. Global Burden of Disease Liver Cancer Collaboration. The burden of primary liver cancer and underlying Etiologies from 1990 to 2015 at the global, regional, and National Level: results from the global burden of disease study 2015. *JAMA Oncol*. 2017;3:1683–91.
3. Allemani C, Matsuda T, Di Carlo V, *et al*. Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet* 2018;391:1023–75.
4. Sia D, Villanueva A, Friedman SL, Llovet JM. Liver cancer cell of origin, molecular class, and effects on patient prognosis. *Gastroenterol* 2017;152:745–61.
5. Massarweh NN, El-Serag HB. Epidemiology of hepatocellular carcinoma and intrahepatic cholangiocarcinoma. *Cancer Control* 2017;24:1073274817729245.
6. de Martel C, Maucort-Boulch D, Plummer M, Franceschi S. Worldwide relative contribution of hepatitis B and C viruses in hepatocellular carcinoma. *Hepatology* 2015;62:1190–200.
7. Bosetti C, Turati F, La Vecchia C. Hepatocellular carcinoma epidemiology. *Best Pract Res Clin Gastroenterol* 2014;28:753–70.
8. Kurtzman CP, Horn BW, Hesseltine CW. Aspergillus nomius, a new aflatoxin-producing species related to Aspergillus flavus and Aspergillus tamarii. *Antonie van leeuwenhoek* 1987;53:147–58.
9. Fattovich G, Stroffolini T, Zagni I, Donato F. Hepatocellular carcinoma in cirrhosis: incidence and risk factors. *Gastroenterol* 2004;127:S35–50.
10. Khanna R, Verma SK. Pediatric hepatocellular carcinoma. *World J Gastroenterol*. 2018;24:3980–99.
11. Zhang XF, Liu XM, Wei T, Liu C, Li MX, Long ZD, Lv Y. Clinical characteristics and outcome of hepatocellular carcinoma in children and adolescents. *Pediatr Surg Int* 2013;29:763–70.
12. Hung GY, Horng JL, Yen HJ, Lee CY, Lin LY. Changing incidence patterns of hepatocellular carcinoma among age groups in Taiwan. *J Hepatol* 2015;63:1390–6.

Downloaded from https://academic.oup.com/jjco/article-abstract/doi/10.1093/jjco/hyaa130/5876941 by Delhi Technological University user on 03 August 2020

13. Chang MH, Chen DS, Hsu HC, Hsu HY, Lee CY. Maternal transmission of hepatitis B virus in childhood hepatocellular carcinoma. *Cancer* 1989;64:2377–80.
14. Liang X, Bi S, Yang W, *et al*. Epidemiological serosurvey of hepatitis B in China—declining HBV prevalence due to hepatitis B vaccination. *Vaccine* 2009;27:6550–7.
15. Lok AS, McMahon BJ. Chronic hepatitis B. *Hepatol* 2007;45:507–39.
16. Ferlay J, Colombet M, Soerjomataram I, *et al*. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer* 2019;144:1941–53.
17. Ferlay J, Ervik M, Lam F, *et al*. *Global Cancer Observatory*. https://gco.iarc.fr/. (28 June 2020, date last accessed).
18. Vostakolaei F, Karim-Kos HE, Janssen-Heijnen ML, Visser O, Verbeek AL, Kiemeney LA. The validity of the mortality to incidence ratio as a proxy for site-specific cancer survival. *Eur J Public Health* 2010;21:573–7.
19. Choi E, Lee S, Nhung BC, Suh M, Park B, Jun JK, Choi KS. Cancer mortality-to-incidence ratio as an indicator of cancer management outcomes in Organization for Economic Cooperation and Development countries. *Epidemiol Health* 2017;39:e2017006.
20. Sharma R. The burden of prostate cancer is associated with human development index: evidence from 87 countries, 1990–2016. *EPMA J* 2019a;10:137–52.
21. Sharma R. Breast cancer incidence, mortality and mortality-to-incidence ratio (MIR) are associated with human development, 1990–2016: evidence from global burden of disease study 2016. *Breast Cancer* 2019b;26:428–45.
22. UNDP. *Human Development Statistical Update 2018*. http://hdr.undp.org/sites/default/files/2018_human_development_statistical_update.pdf (24 June 2019, date last accessed).
23. Human Development Database: http://hdr.undp.org/en/data# (20 June 2020, date last accessed).
24. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010;127:2893–917.
25. Ghebranious N, Sell S. Hepatitis B injury, male gender, aflatoxin, and p53 expression each contribute to hepatocarcinogenesis in transgenic mice. *Hepatology* 1998;27:383–91.
26. Maeda S, Kamata H, Luo JL, Leffert H, Karin M. IKKβ couples hepatocyte death to cytokine-driven compensatory proliferation that promotes chemical hepatocarcinogenesis. *Cell* 2005;121:977–90.
27. De Maria N, Manno M, Villa E. Sex hormones and liver cancer. *Mol Cell Endocrinol* 2002;193:59–63.
28. Yeh SH, Chen PJ. Gender disparity of hepatocellular carcinoma: the roles of sex hormones. *Oncology* 2010;78:172–9.
29. Wilsnack RW, Wilsnack SC, Kristjanson AF, Vogeltanz-Holm ND, Gmel G. Gender and alcohol consumption: patterns from the multinational GENACIS project. *Addiction* 2009;104:1487–500.
30. Mittal S, El-Serag HB. Epidemiology of hepatocellular carcinoma: consider the population. *J Clin Gastroenterol* 2013;47:S2–6.
31. Choo SP, Tan WL, Goh BK, Tai WM, Zhu AX. Comparison of hepatocellular carcinoma in eastern versus western populations. *Cancer* 2016;122:3430–46.
32. Kudo M, Izumi N, Kubo S, *et al*. Report of the 20th Nationwide follow-up survey of primary liver cancer in Japan. *Hepatol Res* 2020;50:15–46.
33. WHO. *Global Hepatitis Report*. https://apps.who.int/iris/bitstream/handle/10665/255016/9789241565455-eng.pdf (6 May 2020, date last accessed).
34. Ott JJ, Stevens GA, Groeger J, Wiersma ST. Global epidemiology of hepatitis B virus infection: new estimates of age-specific HBsAg seroprevalence and endemicity. *Vaccine* 2012;30:2212–9.
35. Centers for Disease Control and Prevention. *Viral Hepatitis Surveillance United States*. https://www.cdc.gov/hepatitis/statistics/2014surveillance/pdfs/2014hepsurveillancerpt.pdf (26 June 2020, date last accessed).
36. Asselah T, Marcellin P, Schinazi RF. Treatment of hepatitis C virus infection with direct-acting antiviral agents: 100% cure? *Liver Int* 2018;38:7–13.
37. Blach S, Zeuzem S, Manns M, *et al*. Global prevalence and genotype distribution of hepatitis C virus infection in 2015: a modelling study. *Lancet Gastroenterol Hepatol* 2017;2:161–76.
38. Aggarwal R, Chen Q, Goel A, Seguy N, Pendse R, Ayer T, Chhatwal J. Cost-effectiveness of hepatitis C treatment using generic direct-acting antivirals available in India. *PloS one* 2017;12:e0176503.
39. Adami H-O, Chow W-H, Nyrén O, *et al*. Excess risk of primary liver cancer in patients with diabetes mellitus. *J Natl Cancer Inst* 1996;88:1472–7.
40. El-serag HB, Tran T, Everhart JE. Diabetes increases the risk of chronic liver disease and hepatocellular carcinoma. *Gastroenterol* 2004;126:460–8.
41. Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. *N Engl J Med* 2003;348:1625–38.
42. Petrick JL, Thistle JE, Zeleniuch-Jacquotte A, *et al*. Body mass index, diabetes and intrahepatic cholangiocarcinoma risk: the liver cancer pooling project and meta-analysis. *Am J Gastroenterol* 2018;113:1494.
43. Wild SH, Morling JR, McAllister DA, *et al*. Type 2 diabetes and risk of hospital admission or death for chronic liver diseases. *J hepatol* 2016;64:1358–64.
44. Bellentani S. The epidemiology of non-alcoholic fatty liver disease. *Liver Int* 2017;37:81–4.
45. Said A, Ghufran A. Epidemic of non-alcoholic fatty liver disease and hepatocellular carcinoma. *World J Clin Oncol* 2017;8:429–36.
46. Zoller H, Tilg H. Nonalcoholic fatty liver disease and hepatocellular carcinoma. *Metabolism* 2016;65:1151–60.
47. Farrell GC, Wong VW, Chitturi S. NAFLD in Asia—as common and important as in the west. *Nat Rev Gastroenterol Hepatol* 2013;10:307–18.
48. Wang FS, Fan JG, Zhang Z, Gao B, Wang HY. The global burden of liver disease: the major impact of China. *Hepatology* 2014;60:2099–108.
49. Qu C, Duan Z, Chen K, Zou H. Reducing liver cancer risk beginning at birth: experiences of preventing chronic hepatitis B virus infection in China. *Hepatoma Res* 2017;3:228–40.
50. Hutton DW, So SK, Brandeau ML. Cost-effectiveness of nationwide hepatitis B catch-up vaccination among children and adolescents in China. *Hepatology* 2010;51:405–14.
51. Asia Pacific Cohort Studies Collaboration. The burden of overweight and obesity in the Asia–Pacific region. *Obes Rev* 2007;8:191–6.
52. Czauderna P. Adult type vs. childhood hepatocellular carcinoma–are they the same or different lesions? Biology, natural history, prognosis, and treatment. *Med Pediatr Oncol* 2002;39:519–23.
53. Chen JC, Chen CC, Chen WJ, Lai HS, Hung WT, Lee PH. Hepatocellular carcinoma in children: clinical review and comparison with adult cases. *J Pediatr Surg* 1998;33:1350–4.
54. El-Serag HB. Hepatocellular carcinoma. *N Engl J Med* 2011;365:1118–27. doi: 10.1056/nejm ra 1001683.
55. McAteer JP, Goldin AB, Healey PJ, Gow KW. Surgical treatment of primary liver tumors in children: outcomes analysis of resection and transplantation in the SEER database. *Pediatr Transplant* 2013;17:744–50.
56. Darbari A, Sabin KM, Shapiro CN, Schwarz KB. Epidemiology of primary hepatic malignancies in US children. *Hepatology* 2003;38:560–6.
57. Wang J, Mao Y, Liu Y, Chen Z, Chen M, Lao X, Li S. Hepatocellular carcinoma in children and adolescents: clinical characteristics and treatment. *J Gastrointest Surg* 2017;21:1128–35.
58. Zhang XF, Liu XM, Wei T, Liu C, Li MX, Long ZD, Lv Y. Clinical characteristics and outcome of hepatocellular carcinoma in children and adolescents. *Pediatr Surg Int* 2013;29:763–70.
59. Bruix J, Sherman M. Management of hepatocellular carcinoma: an update. *Hepatol* 2011;53:1020–2.
60. Zhang BH, Yang BH, Tang ZY. Randomized controlled trial of screening for hepatocellular carcinoma. *J Cancer Res Clin Oncol* 2004;130:417–22.
61. Lederle FA, Pocha C. Screening for liver cancer: the rush to judgment. *Ann Intern Med* 2012;156:387–9.

Downloaded from https://academic.oup.com/jjco/article-abstract/doi/10.1093/jjco/hyaa130/5876941 by Delhi Technological University user on 03 August 2020

62. American Cancer Society. *Can Liver Cancer Be Found Early?* https://www.cancer.org/cancer/liver-cancer/detection-diagnosis-staging/detection.html (28 June 2020, date last accessed).
63. National Cancer Institute. *Liver (Hepatocellular) Cancer Screening–Health Professional Version*. https://www.cancer.gov/types/liver/hp/liver-screening-pdq (28 June 2020, date last accessed).
64. Owens DK, Davidson KW, Krist AH, *et al*. Screening for hepatitis C virus infection in adolescents and adults: US preventive services task force recommendation statement. *JAMA* 2020;323:970–5.
65. Zibbell JE, Asher AK, Patel RC, Kupronis B, Iqbal K, Ward JW, Holtzman D. Increases in acute hepatitis C virus infection related to a growing opioid epidemic and associated injection drug use, United States, 2004 to 2014. *Am J Public Health* 2018;108:175–81.
66. Kapila N, Muir AJ. A big step forward in hepatitis C screening. *Clin Liver Dis* 2020;15:149.
67. Sharma R. An examination of colorectal cancer burden by socioeconomic status: evidence from GLOBOCAN 2018. *EPMA J* 2020;11:95–117.
68. Etemadi A, Safiri S, Sepanlou SG, *et al*. The global, regional, and national burden of stomach cancer in 195 countries, 1990–2017: a systematic analysis for the global burden of disease study 2017. *Lancet Gastroenterol Hepatol* 2020;5:42–54.

Downloaded from https://academic.oup.com/jjco/article-abstract/doi/10.1093/jjco/hyaa130/5876941 by Delhi Technological University user on 03 August 2020

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Detecting Intrusions and Attacks in the Network Traffic using Anomaly based Techniques

Vinod Kumar
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
vinod_k@dtu.ac.in

Vinay Choudhary
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
choudharyvinay92@gmail.com

Vivek Sahrawat
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
viveksahrawat7@gmail.com

Vinay Kumar
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
vinayverma9650@gmail.com

***Abstract*— Technology has become the backbone of today's Information and Communication Technology. Today a large number of transactions are carried out online and thus possess a risk of information stealing and misuse. Intrusion Detection techniques are thus developed and applied to the networks and individual host computers to detect such intrusive actions and alarm the administrators. A secure machine-controlled anomaly detection system is a more practical procedure to help in network analysis. An anomaly detection system (ADS) examine the network flow and focuses on detecting uncommon network behaviour, and classify them into attacks. The propsoed research work present a comparison of some of these intrusion detection techniques using different machine learning algorithms and ensemble learning. A subset of attributes (significant) is chosen from the primitive set of attributes using a random forest regressor technique, and then, the chosen set of significant features is used to train different classifiers. Moreover, the proposed work has been analyzed by experimenting on one of the publically available dataset CICIDS2017, available from the Canadian University of Cybersecurity.**

***Keywords*— Classification, anomaly detection system (ADS), CICIDS2017, machine learning.**

I. INTRODUCTION

Established institutions and ample folks communicate with one another over the internet every day. Within the past twenty years, where the number of individuals using the net services has been hyperbolic. Today this range has exceeded four billion and is increasing exponentially. Moreover, any business, government enterprise, or industry relies heavily on technology to conduct their day to day activities, transactions, or communicate. The data today is not restricted to have a physical presence, the e-money, photos, texts, id passwords, videos, and many more forms of virtual or online data also exists. And there is a need to preserve and safeguard this online data. There is an improvement in infrastructures and the protocols of the wireless devices and IoT applications. Online services and networks are often used for communicating and exchanging sensitive information. Attackers often peep into these networks and get access to sensitive and confidential information which suggests the requirement for some mechanism to detect such actions and then alert the authorities about the same.

The recent COVID-19 outbreak further forced the world to adopt the concept of work from home to keep their services and businesses running, and this remote working mechanism comes with a whole new and wider dimension of internet and network security, with making their networks more vulnerable to cyber-attacks. With the remote working concept, a greater amount of sensitive and confidential information has to be shared via the network, which brings dozens of opportunities for the hackers and the intruders to steal this sensitive information and data. Mobile office and work from home concept allow the employees, who are authorized to access the corporate network and have to share sensitive information over the mobile or computer networks.

Every aspect of cybersecurity reviles around the CIA triad. The C here stands for confidentiality, which means keeping the data hidden from unwanted access. I stand for integrity means keeping the data unaltered throughout the transmission process. The sent and received data over the net should be the same. Finally, the A stands for availability, meaning the data should be available to the intended use when required. And security aspects enforces us to be ready with backups whenever the data is lost or is delayed. There are some major network attacks:

- DNS cache poisoning attack, it tricks the DNS

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:59:53 UTC from IEEE Xplore. Restrictions apply.

servers and makes it accept a fake DNS address of the corresponding IP address of the requested legitimate websites. This attack is primarily used for malicious banking credential stealing.

- Another major network attack is the man-in-the-middle attack. An intruder places itself between two hosts, who think that they are directly communicating with each other but the communication is happening between the middle attacker with each of the two hosts simultaneously. This attack has the potential to access and modify the information passing to and fro the network and hence compromising the CIA triad principles.
- Denial of service attack challenges the availability principle of security goals. The hosts or the severs are overwhelmed by bombarding them with huge unnecessary requests, which results in the crashing of the systems and making legitimate users starve for accessing the required resources. This may delay the entire service of an enterprise and severely affect the time-bound services, which can cause slow network performance. The DoS attack may also be made in a distributed fashion and is known as distributed denial of service attack, where a combined network of physically separate machines are used to overwhelm the server with the request.

It is a challenge for the researchers to develop efficient algorithms to detect such types of network attacks before harm is caused to the victim organization. For this purpose, some benchmarking datasets are required to evaluate the developed algorithms and techniques.

In this paper, some tasks are implemented for detecting an anomaly in networks by making use of effective machine learning techniques. Also, for evaluating the most useful features or attributes and consider them for further study, a random forest regressor algorithm is used during pre-processing.

The rest of the research paper is structured into various sections. The related work that has been done in the field of anomaly detection is highlighted in section 2. Section 3 presents the background details regarding the dataset and algorithms used for our study. Section 4 consists of the details of the pre-processing stage and the experimental setup. Section 5 then compares the results of various algorithms and analyses them. Section 6 concludes our work and Section 7 presents the scope for future work.

II. RELATED WORK

For anomaly detection in networks by machine learning techniques, a large quantity of harmful and harmless network traffic data is required for training and testing steps, However, real-time network traffic can't be used in public due to security problems [1] and thus to develop efficient intrusion detection systems, several datasets have been simulated in almost real environments. DARPA 98, KDD 99, CAIDA (Centre of Applied Internet Data Analysis), NSL-KDD, ISXC 2012 and, CICIDS 2017 are various datasets for network anomaly detection. The initially available datasets were DARPA dataset [2], which was made publically available in 1998, and several improvements were made to it after that."Contests were organized in the initial years for introducing effective new techniques for intrusion detection. With all the advancements and new emerging attacks, there was a need for more comprehensive, balanced, up to date and perfect datasets" [3]–[4]. Recently, the Canadian University of Cybersecurity launched a new dataset, containing novel attacks. This dataset is named as CICIDS-2017. Moreover, this dataset also incorporates all the 11 criteria of a good dataset, as suggested by the authors of [8].

For anomaly detection, these datasets are used as a benchmark in performance evaluation. There is no single "Fit-For-All" algorithm for detecting various attacks differing enormously based on their source, target, and mode of application. The anomaly detection systems are often classified based on the type of environment in which they are deployed as "Host Anomaly Detection Systems" and "Network Anomaly Detection Systems" [6]. The proposed work aim to detect attacks by finding anomalies in the "Network Anomaly Detection Systems" in this paper. Another classification of ADS can be based on techniques employed in implementing them. There are mainly two categories: "Signature-based ADS" and "Anomaly-based ADS". Signature-based ADS were prevalent in real-world applications since the last decade. But due to their inability to detect zero-day attacks, they are replaced by anomaly-based ADS. Signature-based ADS were based on concepts like string matching, regular expressions, and finite state machines [7].

III. BACKGROUND

A. Network Intrusion Detection System

The network of computers was first secured by Network Hardening Principle. Network hardening refers to the process of securing networks by making configuration changes to reduce vulnerabilities. But these changes are often a burden on the administration and hence other mechanisms were evaluated. One such mechanism evolved was of firewall, where one can disable and filter the unnecessary services and restrict access to them. This was achieved but creating manual rules for traffic, similar to the white-listing process. This came with its challenge of adding or deleting new rules, every time some new service was desired to work or restrict and thus causing inconvenience for the administrators.

All these challenges were resolved with the advancement in the field of artificial intelligence. Machine learning concepts

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:59:53 UTC from IEEE Xplore. Restrictions apply.

provided an important and efficient way to monitor and analyze the traffic flowing across the networks. These techniques let us make a baseline of what our normal network traffic looks like and then any deviation from this baseline can be treated as an anomaly and users can be alarmed regarding the suspicious network traffic presence. The requirements for these anomaly detection systems are :

- an architecture to collect and extract the traffic data,
- a good dataset schema to store the data along with its attributes, and
- efficient algorithms to analyze this stored data to deduce outliers and intrusions from it.

Today most of the systems employ more efficient and effective anomaly-based systems using the machine learning algorithms which can generalize the type of attacks from real-time network traffic. Machine learning algorithms generally consists of two phases – the "learning phase" and the "inference phase" [8]. The learning phase enables us to build the network traffic profile, which generalizes the benign and normal traffic passing through the network. During the inference phase, any abnormal behaviour is treated as an anomaly and indicates the presence of attack, and the system administrators are alerted about the same.

For anomaly detection in networks by machine learning techniques, a large quantity of harmful and harmless network traffic data is required for training and testing steps. The anomaly detection-based approach focuses on detecting uncommon network behaviour by examining network flow. This anomaly detection technique has been booming in detecting attacks that's not encountered yet, thus it is effective against zero-day attacks.
There are various difficulties in the implementation of NIDS:

- First, the detection rate (DR) is needed to be enhanced and False Prediction Rate should be reduced.
- Second, extracting, recording, and selecting the important features is a very crucial process as the classifier's performance is completely dependent on the choice of input features. Hence, a dataset with good quality and quantity of features is desired.
- Third, unbalanced and skewed datasets can become a significant hurdle in the wholesome analysis of the performance of proposed algorithms, and results may be skewed. Therefore, the dataset must be well balanced and contain records for all attacks and labels in the right prevalence ratio.

B. Description of CICIDS 2017 Dataset

At the University of New Brunswick, the Canadian Institute for Cybersecurity created CICIDS2017 [9]. This dataset consists of a 5-day (3rd July- 7th July 2017) data stream on a network created by computers having up-to-date operating systems like Windows7 / 8.1 / 10, Ubuntu 12/16, Kali and Mac. DoS (Denial of Service) HULK, DDoS (Distributed DoS), FTP-Patator, DoS Slowloris, SSH-Patator, DoS SlowHTTPTest, DoS Goldeneye, Botnet, PortScan, Web Attack, Infiltration, Heartbleed are various types of attacks present in the datasets available. The 2017 CICIDS dataset has the subsequent benefits over other datasets:

- The obtained information is the real-world data; obtained from a testbed consisting of real computers.
- Data streams are collected from computer systems with the up-to-date operating system. There's operating system diversity (Windows, Linux, and Mac) between each victim computer.
- Datasets are labeled. To apply the machine learning techniques, feature extraction, which is an important step, was applied and 85 features (see Appendix A for the feature list) were obtained.
- Both raw information (pcap files - captured network packets files) and processed data (CSV files- comma-separated data files) are out there to work on.
- In the course of deciding which attack to take place, the 2016 McAfee security report was used, therefore there's a large and up-to-date assortment of attacks.
- It is more abundant than other datasets regarding protocols used. It conjointly includes the HTTPS (Hypertext Transfer Protocol Secure) protocol additionally to FTP, HTTP, SSH, and e-mail protocols.

On the other hand, CICIDS2017 dataset has subsequent disadvantages:

- Raw data files and processed data files are giant (97.9 GB and 1147.3 MB respectively).
- Unlike the KDD99 and NSL-KDD datasets, CICIDS2017 doesn't have separate files dedicated to training and testing. These sections ought to be created by users. The way to do that is handled within the creation of the training and testing section.
- The formation of DARPA98, KDD99 then NSL-KDD is analogous to an evolutionary process. At every step, faults and deficiencies within the previous dataset have been discovered, and measures were taken to forestall them. On the other hand, CICIDS 2017 is incredibly new and has not been studied much yet, therefore it's possible to contain some minor mistakes. What these mistakes are and the way to repair them is processed within the Data Pre-processing section.

Dataset is divided into a training dataset and a testing dataset. 80% of original dataset instances are selected randomly and referred to as a training dataset and the remaining instances are referred to as testing datasets.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:59:53 UTC from IEEE Xplore. Restrictions apply.

C. Random Forest As Feature Selection Technique

Random forest works on the concept of the fair election as result is based on the prediction of several independent decision trees and a group of decision trees having the same prediction will outperform individual decision trees. It uses a random subset of a given dataset and a random subset of features for creating an individual decision tree so that individual decision tree has little amount of correlation with each other.

To ensure that individual decision tree has little amount of correlation with each other random forest uses the following two methods:

- Bagging (Bootstrap Aggregation) – Decision trees are very sensitive to changes in the dataset. The random forest takes advantage of this property of decision trees and allows every individual decision tree to select a subset of the dataset from the available dataset in a random way. This results in independent and different decision trees. This process of allowing random selection of data is known as bagging.
- Randomness in feature – When a node is splitted in a decision tree, every possible feature is used to select the path to be taken in the process of making predictions. In contrast, random forest allows every individual decision tree to select a subset of features from the available feature list in a random way. This results in even more variations among the different decision trees and lower correlation across trees and tress become more independent from each other.

For feature selection, random forest technique assigns weight to different features according to how much a particular feature reduces the accuracy in the classification of the individual decision tree. Dependent variables have a low weight assigned as the Random forest technique ensures that there is a little amount of correlation between individual decision trees.

D. Ensemble Learning

The ensemble method's goal is to mix the prediction results of multiple learning algorithms. Many base classifiers/regressors designed and trained a different learning algorithm for better predictions than individual learning models. Ensemble Methods are divided into two groups

- Sequential ensemble methods – Here base learners are geerated sequentially. The dependence of base learners on each other is exploited here. AdaBoost is a suitable example.
- Parallel ensemble methods – Here base learners are generated parallely. The independence of the base learners is used here. Random Forest is an example.

The concept behind the ensemble classifier is to mix different machine learning algorithms and based on majority vote also called hard voting or the average prediction chances also called soft voting to make predictions. Such a classifier is often useful for a group of equally well-performing classifier/regressor to balance out their weaknesses. In the majority vote, the output for a selected sample is that category label that represents the bulk of the category labels classified by each regressor/classifier. e.g. Different predictions for a given sample (To classify whether the given image is of a hot dog or not).

- model 1 -> yes
- model 2 -> yes
- model 3 -> no

then using the ensemble learning model will predict "yes" based on the majority class prediction. Soft vote predicts the category label as an average of the sum of predicted probabilities. Specific weights are often assigned to each classifier/regressor through the weight parameter. Once weights are assigned, the expected class label probabilities for each classifier/regressor are collected, multiplied by the allotted classifier's weight, and averaged. The final class label is then predicted from the individual category label with the best average likelihood.

IV. EXPERIMENTS

A. Data pre-processing

Various techniques for data pre-processing are performed on the dataset downloaded from the official CICIDS2017 website [10]. Dataset is pre-processed to get rid of the mistakes and defects present. It's necessary to pre-process the dataset before training and testing. For this, some defects within the CICIDS2017 dataset are processed and corrected. One major is the class imbalance problem [12], which is corrected by following the pre-processing stage.

The dataset contains 3119345 records, collected, and analyzed using CICFlowMeter [11]. Once the records are examined, it is seen that around 2.5 lacks records are incomplete or incorrect. The primary step is to delete the surplus records. The next error within the dataset is within the columns that form up the features. The dataset contains 86 columns that define the properties e.g. Flow ID, Source IP, Source Port, etc. The Fwd Header Length feature (which tells about the forward direction data flow for total bytes used) was repeated two times one in the 41st column and second time in the 62nd column. This error is removed by deleting the repeating columns. Table 1 contains the description of the dataset.

Another editing that needs to be done in the dataset is to convert the value of the features including string and categorical values (Flow ID, Source IP, Timestamp, External IP, Destination IP) into numerical data to be easily understandable by different machine learning algorithms. This

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:59:53 UTC from IEEE Xplore. Restrictions apply.

is often be done using LabelEncoder() function provided by the Sklearn package. In this way, all string values that don't seem to be understandable by machine learning operations can get integer values and become more efficient for further processing. The "Label" tagged column is also modified, if there is an attack then it's value will be 1 and if there is no attack then it's value will be 0. Finally, some structural changes are done to the dataset, including:

- In the Label feature, the character "–" (Unicode Decimal Code &#8211) used to identify the web attack subtypes (Brute Force and SQL Injection) should get replaced with the character "-" (Unicode Decimal Code &#45), as utf-8, the default Pandas library, won't recognize it. Otherwise, the Pandas library won't recognize this character and can fail.
- "Flow Bytes/s", "Flow Packets/s" features have valued to like "Infinity" and "NaN" with numerical values, that are changed to -1 and 0 respectively to make them efficient for various machine learning algorithms.

Table 1. Description of CICIDS2017 Dataset

Recording Day	pcap file size (GB)	Recording Hours	CSV file size (MB)	Attack
Mon	10.0	Full Day	257	-
Tues	10.0	Full Day	166	SSH-Patator, FTP-Patator
Wed	12.0	Full Day	272	DoS Hulk, GoldenEye, Slowloris, Slowhttptest, Heartbleed
Thurs	7.7	Morning	88	Web Attacks
		Afternoon	103	Infiltration
Fri	8.2	Morning	72	Bot
		Afternoon	190	PortScan, DDos

B. Experimental Setup

The technical specifications of the computer used in the implementation:

- CPU: "Intel(R) Core (TM) i5-7200U CPU @ 2.70GHz"
- RAM: 8 GB
- OS: "Windows 10 Pro 64-bit"
- GPU: "Nvidia 940 mx 8gb ddr4"

Jupyter-Notebook 5.7.8 on Anaconda 2019.03 with Python 3.7.3 is used for implementing data pre-processing steps and various machine learning algorithms. Various python libraries are used: pandas, matplotlib, and sklearn. The preprocessed training data set having optimal features yielded by the pre-processing stage are used for the comparison of various machine learning classifiers, with the ensemble learning approach. Machine learning classifiers and ensemble learning classifiers are trained for 1000 epochs.

C. Performance Measures

The results of the algorithms are evaluated and compared according to four measures i.e. accuracy, recall, precision, and F1-Score. These measures have a value between 0 and 1. If the value is close to 1, then the performance of the algorithm used increases, and if the value is close to 0, then it decreases.

- Accuracy – The ratio of correctly predicted instances to a total number of instances in the dataset.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

- Recall (Sensitivity) – The ratio of instances predicted as an attack to all the instances corresponding to attacks.

$$Recall(Sensitivity) = \frac{TP}{TP + FN} \quad (2)$$

- Precision – The ratio of correctly predicted data instances as an attack to all data instances predicted as an attack.

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

- F-measure (F-score/F1-score) – The harmonic-mean of recall/sensitivity and precision. This measure is the most important. Hence, more focus is given on F-measure.

$$F1 - Score = \frac{2 * (Sensitivity) * (Recall)}{Sensitivity + Recall} \quad (4)$$

Where,
TP: Actual attack data predicted as an attack.
FP: Actual no-attack data predicted as an attack.
FN: Actual attack data predicted as no-attack.
TN: Actual no-attack data predicted as no-attack.

V. RESULTS AND DISCUSSION

In this section, the results of the study carried out are presented. The performance evaluation criterion is based on the values of the F1-measure. Histograms are created to visualize the comparison between different features and different machine learning algorithms. A random forest regressor [13] is used for necessary feature selection. First, a simple principal component analysis (PCA) is used for feature selection, but PCA is not efficient when there is an enormous amount of feature in the dataset. The below figure shows the importance of various features.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:59:53 UTC from IEEE Xplore. Restrictions apply.

Table 2. Important features with importance weights computed using random forest [9]

FEATURES	IMPORTANCE WEIGHT ($*10^{(-5)}$)
Standard Backward Packet Length	24656
Flow Byte/s	17876
Forward Packets Total Length	11721
Standard Forward Packet Length	6389
Standard IAT	1017
Minimum IAT	658
Total IAT	499
Flow Duration	413
Maximum Flow IAT	354
Maximum Backward Packet Length	339
Mean IAT	329
Backward Packets Total Length	135
Minimum Forward Packet Length	63
Flow Packet/s	58
Mean Backward Packet Length	53
Mean Forward Packet Length	51
Total Backward Packet/s	23
Total Forward Packet/s	13
Maximum Forward Packet Length	12
Minimum Backward Packet Length	8

Table 2 exhibits the importance weight of the top 20 important features computed using random forest after data pre-processing.

Four machine learning algorithms, ID3 (Decision Tree), MLP, ensemble learning, and Naïve Bayes, are used and compare them on the premise of various performance evaluation criterion i.e. precision, recall, accuracy, and F1-Score. Table 3 exhibits the value of precision, recall, accuracy, and F1-Score of various algorithms. The resulted performance based on F1-measure is as follows (F1-measure takes a value between 0 and 1): Ensemble Learning (NB, ID3 and, MLP): 0.89, ID3: 0.95, MLP: 0.84 and Naive Bayes: 0.82.

Table 3. Performance evaluation data for different algorithms

"ML Algorithm"	Precision	Recall	Accuracy	F1-Score
Naïve Bayes	0.82	0.82	0.8165	0.82
ID3	**0.95**	**0.95**	**0.9515**	**0.95**
Ensemble Learning	0.89	0.89	0.891	0.89
MLP	0.84	0.84	0.8387	0.84

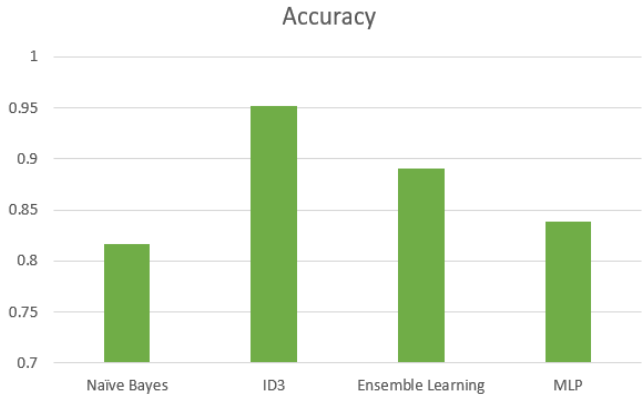


Fig. 1. Comparing Accuracy Parameter of Algorithms

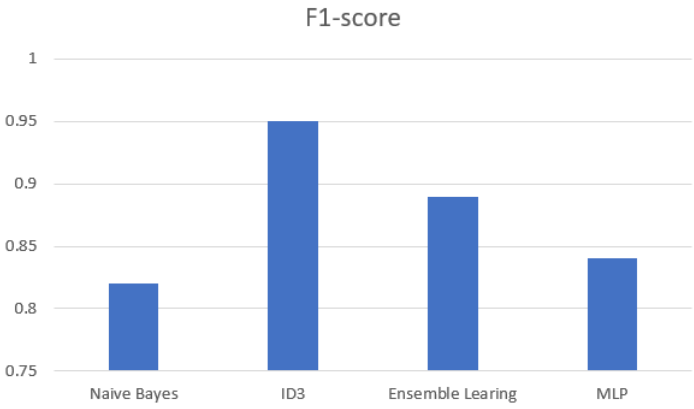


Fig. 2. Comparing F1 score of Algorithms

Fig. 1 and fig. 2 exhibit the accuracy comparison and F1-Score comparison respectively that are obtained after applying various machine learning algorithms. It is evident from fig. 1 and fig. 2 that ID3 and Ensemble Learning perform best out of the various algorithms applied after data pre-processing and feature selection.

Why ID3 come out with the best results?

- ID3 (Iterative Dichotomiser 3) is very intelligible, versatile, and easy to debug. It'll work with classification and regression problems. If you are attempting to predict a categorical value or to predict a continuous value ID3 will handle both problems, but it can handle categorical data efficiently than Naïve Bayes and MLP.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:59:53 UTC from IEEE Xplore. Restrictions apply.

- All Decision Trees is that they solely want a part of data, and they can build a classifier directly from that data. Whereas in other algorithms, choosing features is up to you. There's no way to just toss a part of data, and it picks the most important features that it'll use to classify.
- ID3 is quick regarding computation than other algorithms. ID3 does not produce classifiers using all datasets. It's an iterative method in which a rule is computed and every data record that satisfies the rule is going to be removed and additional rules are generated until all data records aren't covered.

VI. CONCLUSION

The study aims to implement the task for detecting an anomaly in networks by making use of effective machine learning techniques and compare their efficiencies. CICIDS2017 dataset has been used because of its wide attack diversity and up-to-datedness. A network flow is described by the features (over eighty) contained within the dataset. A subset of attributes is chosen from the primitive set of attributes using the Random Forest Regressor technique to make our mind up that which feature is used in machine learning algorithms. Finally, four machine learning algorithms i.e. Naive Bayes, ID3, MLP, and ensemble learning, is practiced on the dataset. The resulted performance based on F1-measure is as follows (F1-measure takes a value between 0 and 1): Ensemble Learning (NB, ID3 and, MLP): 0.89, ID3: 0.95, MLP: 0.84 and Naive Bayes: 0.82. The lower accuracy and huge false alarm rates were also tackled to a great extent. Also, the class imbalance problem was rectified to get better results.

VII. FUTURE SCOPE

In this paper, the primary aim is to implement numerous machine learning algorithms independently of one another for anomaly detection in networks. However, this technique has weak practical applicability in real life. To overcome this drawback, a multi-layered /hierarchical machine learning structure is often designed. Moreover, because of such a structure, it's attainable to save time, CPU power, and memory. There is no "fit-for-all" solution for intrusion detection task [14]. For example, in a two-tiered structure, the primary layer is often created from quick and computationally low-cost algorithms such as Naïve Bayes, therefore network traffic can be observed continuously and at a minimal cost. The primary step, when detecting an anomaly, is to transmit it to an upper layer composed of algorithms with higher performance such as ID3, ensemble learning, etc. The ultimate layer that creates up the determination mechanism takes the preventive decision to protect the network against various attacks. Also, the dataset used, consist of CSV (comma-separated values) files, was obtained from a network of computers created by the Canadian Institute for Cybersecurity. This work is open to future enhancements. Unluckily, in real-time systems, this technique isn't much viable. However, this drawback can be resolved by adding a module that catches real network data and makes it practicable with various machine learning algorithms.

References

[1] R. Sommer, V. Paxson, Outside the Closed World: On Using Machine Learning For Network Intrusion Detection, in: IEEE Symposium on Security and Privacy, IEEE, 2010, pp. 305–316. doi:10.1109/SP.2010.25.

[2] R. Lippmann, J. Haines, D. Fried, J. Korba and K. Das. "The 1999 DARPA off-line intrusion detection evaluation". Computer networks, vol. 34, no. 4, pp. 579 595, 2000. DOI http://dx.doi.org/10.1016/S1389- 1286(00)00139-0.

[3] Nehinbe, J. O. (2011), "A critical evaluation of datasets for investigating idss and ipss researches", In IEEE 10th International Conference on CIS, pages 92–97.

[4] Ali Shiravi, HadiShiravi, M. T. and Ghorbani, A. A. (2012), "Toward developing a systematic approach to generate benchmark datasets for intrusion detection. Computers and Security", 31(3):357 – 374.

[5] Gharib, A., Sharafaldin, I., Habibi Lashkari, A., and Ghorbani, A. A. (2016, "An evaluation framework for intrusion detection dataset", In 2016 International Conference on Information Science and Security (ICISS), pages 1-6.

[6] Abhishek Pharate , Harsha Bhat , Vaibhav Shilimkar "Classification of Intrusion Detection System" IJCS Volume 118 – No. 7, May 2015.

[7] Subaira AS, Anitha P., "Efficient classification mechanism for network intrusion detection system based on data mining techniques: a survey. In international conference on intelligent systems and control", 2014 (pp. 274-80).

[8] S. Ustebay, Z. Turgut, and M. A. Aydin, "Intrusion detection system with recursive feature elimination by using random Forest and deep learning classifier," in 2018 international congress on big data, deep learning and fighting cyber terrorism (IBIGDELFT), 2018, pp.71-76

[9] Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorba-ni, "Toward Generating a New Intrusion Detection Dataset and Intru-sion Traffic Characterization", 4th International Conference on In-formation Systems Security and Privacy (ICISSP), Purtogal, Janu-ary 2018

[10] Search UNB [Internet]. University of New Brunswick est.1785. [cited 2019May26]. Available from: https://www.unb.ca/cic/datasets/ids-2017.html

[11] CICFlowMeter (2017). Canadian institute for cybersecurity (cic).

[12] Panigrahi R, Borah S., "A detailed analysis of CICIDS2017 dataset for designing Intrusion Detection Systems", International Journal of Engineering & Technology 7 (3.24), 2018, 479-482.

[13] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python.

[14] Zhang, C., Pan, X., Li, H., Gardiner, A., Sargent, I., Hare, J., Atkinson, P.M., "A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification", ISPRS Journal of Photogrammetry and Remote Sensing 140 (2018) 133–144,doi:http://dx.doi.org/10.1016/j.isprsjprs.2017.07.014

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:59:53 UTC from IEEE Xplore. Restrictions apply.

The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/1741-0401.htm

Developing a framework for evaluating sustainability index for logistics service providers: graph theory matrix approach

Anchal Gupta
Delhi Technological University, Delhi, India, and
Rajesh Kumar Singh
Department of Operations Management, Management Development Institute, Gurgaon, India

Received 30 December 2019
Revised 30 March 2020
8 April 2020
14 May 2020
17 June 2020
Accepted 17 June 2020

Abstract

Purpose – The purpose of this study is to identify sustainability practices and to propose a framework for evaluating the sustainability index of logistics service providers (LSPs).

Design/methodology/approach – Sustainable practices followed by LSPs are identified through literature review and analysis of a case study. Thirteen such sustainable practices are identified. Thereafter, with expert inputs, nine sustainable practices are shortlisted and considered for the evaluation of the sustainability index in the proposed framework. Graph Theory Matrix Approach has been applied to evaluate the sustainability index of an LSP.

Findings – Major practices identified for evaluating sustainability index include the use of recyclable packaging, use of renewable energy sources, green procurement, reduction in carbon emissions, use of CNG/electric vehicles, rainwater harvesting and so on. The sustainability index of an LSP is evaluated by using the proposed framework.

Practical implications – LSPs can benchmark their sustainability index with respect to the best in the industry. Based on it, LSPs can also identify potential areas for improvement.

Originality/value – Novelty of the study lies in the proposed framework for evaluation of sustainability index which can be used to develop strategies for green logistics. LSPs can also improve their performance in terms of sustainability measures by adopting green logistics.

Keywords Logistics service providers, Sustainable practices, Sustainability index, Graph theory matrix approach, case study

Paper type Research paper

1. Introduction

"Sustainability" – a term first coined by the World Commission on Environment and Development (WCED) in the year 1987 (WCED, 1987) – was adopted gradually by most of the organizations around the world. In the year 2015, the United Nations has set up 17 sustainable development goals (SDGs) for improving sustainable development in business organizations as well as world economies, with a target to achieve these goals by 2030 (Szabo *et al.*, 2016; Salvia *et al.*, 2019; Modgil *et al.*, 2020). Due to this, organizations worldwide are seeking to implement sustainable solutions. With the increasing scarcity of resources as well as increasing needs of society and organizations, there is an urgent need to ensure sufficient availability for current use and conserve for meeting the needs of future generations. Moreover, the environment is continuously deteriorating (Seidel *et al.*, 2013). Therefore, many organizations including logistics service providers (LSPs) are adopting and implementing sustainable practices for optimum use of scarce resources and taking steps for a better global future.

Compared to other industry segments, LSPs contribute more toward environmental degradation in terms of noise and air pollution, carbon emissions, fuel consumption, global warming, resource depletion, and so on (Jazairy and Haartman, 2018). The logistics sector is



International Journal of Productivity and Performance Management
© Emerald Publishing Limited
1741-0401
DOI 10.1108/IJPPM-12-2019-0593

expected to grow exponentially, which will lead to even worse environmental conditions (Wang *et al.*, 2016). Therefore, LSPs are required to initiate sustainable logistics solutions for handling current and upcoming environmental challenges. Previous studies have explored the benefits of the adoption of green practices by various organizations (Perotti *et al.*, 2012; Dubey *et al.*, 2017; Majumdar and Sinha, 2018). Many authors have also discussed the impact of the implementation of sustainable practices on firms' performance (Perotti *et al.*, 2012; Green *et al.*, 2019). Ansari and Kant (2017) explored the positive impact of the sustainable practices of logistics service providers on their efficiency, productivity, and profitability. Many authors have also highlighted the importance of the triple bottom line approach in logistics for improving sustainability (Agrawal *et al.*, 2016; Agrawal and Singh, 2019; Halldorsson, 2019; Kumar and Goswami, 2019).

The existing literature focuses more on green logistics practices in developed countries compared to developing countries (Gandhi *et al.*, 2015; Mangla *et al.*, 2015; Agrawal and Singh, 2019). Gupta and Singh (2020) indicate the need for greater emphasis on developing countries to create a balance between environmental sustainability and the operational excellence of LSPs. Presently, due to environmental consciousness, even consumers in developing nations have started expecting green products. LSPs in developing economies have initiated sustainable practices but only on a small scale. In India, LSPs have started transforming their processes into green processes by changing their conventional fleet into CNG (Compressed Natural Gas) or hybrid fleets for reducing carbon emissions. LSPs have also initiated various activities such as planting trees in the surroundings of their facilities, storing rainwater for washing vehicles and installing solar panels for saving electricity (Gupta *et al.*, 2018). LSPs are also working toward a reduction in waste, which is created in different forms such as underutilized resources, replication of tasks, and inadequate planning of resource allocation (Qian *et al.*, 2019). Moreover, LSPs are exploring alternative ways of reducing carbon emissions, fuel consumption, and noise levels for long-term sustainability. Despite the measures being adopted by LSPs, the actual adoption rate is very low due to lack of continuous evaluation measures. Owing to the lack of appropriate frameworks in the literature, most LSPs struggle in evaluating their sustainability index (Chhabra and Singh, 2016). There is, therefore, a need to perform an in-depth analysis of sustainable practices followed by LSPs and to evaluate the sustainability index of such providers in a developing country. This study analyzes and evaluates the sustainability performance of LSPs in the Indian context. The main objectives of this study are as follows:

(1) To identify the sustainable practices followed by LSPs

(2) To analyze the sustainable practices specifically for LSPs in the Indian context

(3) To evaluate the sustainability index using a Graph Theory Matrix Approach (GTMA)

(4) To propose a framework for evaluating the sustainability index of LSPs

This study is further elaborated under the following sections. Section 2 comprises the literature review on green initiatives and sustainable practices followed by LSPs. In Section 3, the case study method is used to identify and analyze sustainable practices in developing economies. In Section 4, the sustainability index has been evaluated by using GTMA. The implications of the study along with future scope are discussed in Section 5. The conclusions of the study are concluded in Section 6.

2. Literature review

There is a significant body of literature available on evolution, adoption, and implementation of sustainable practices in both the manufacturing and service sectors (Perotti *et al.*, 2012;

Anwer *et al.*, 2018; Majumdar and Sinha, 2018; Mensah and Casadevall, 2019). Sustainability means the conservation of natural resources and judicious use of resources for long term business continuity. Due to continuous environmental degradation and scarcity of resources, resource sustainability has become an important concern for all organizations including LSPs (Coaffee, 2008; Zhang *et al.*, 2017). LSPs have started redefining their conventional processes into green processes as a contribution towards society and the environment (Ferreira *et al.*, 2015). Previous studies discussed the adoption and importance of sustainability in the field of logistics (Gandhi *et al.*, 2015; Mangla *et al.*, 2017). Centobelli *et al.* (2017) conducted a systematic literature review on green and sustainable practices followed by LSPs by exploring more than 415 articles from 1960 to 2014. Evangelista *et al.* (2018) conducted a comprehensive literature review on the environmental sustainability of LSPs for the period of 2000–2016.

Researchers have not only conducted comprehensive literature reviews on environmental sustainability but also discussed the benefits that organizations can obtain by implementing sustainable practices. Various authors captured different aspects of logistics and linked them with sustainability measures. For example, Wilding *et al.* (2012) highlighted the importance of supplier assessment and collaboration in improving sustainability. Martinez-Jurado *et al.* (2014) explored the relationship between lean management, supply chain management, and sustainability. Marques (2019) conducted a detailed and systematic literature review on sustainable supply network management from a knowledge perspective. Many practitioners have already proved the strong relationship among the implementation of sustainable practices, an increase in profitability, and improvement in organizational performance (Perotti *et al.*, 2012; Singh *et al.*, 2016; Ansari and Kant, 2017). Several researchers have established the relationship among implementation of green sustainable practices and organizational culture, service quality, and operational performance (Zhu *et al.*, 2008; Bask *et al.*, 2018; Garcia-Dastugue; Eroglu, 2019; Green, 2019; San-Ong *et al.*, 2019).

Despite the extensive literature on logistics and sustainability, LSPs, specifically those in developing countries, are yet to implement green practices in all their processes (Agrawal and Singh, 2019). As a result, it is imperative for LSPs in emerging economies to practice resource conservation and ensure long-term business continuity (Singh and Agrawal, 2018). Moreover, organizations, including LSPs, are expected to follow strict guidelines for environmental protection, issued by the government and the United Nations (UN). The UN has set sustainable development goals (SDGs-7, 12 and 13) for the implementation of sustainable practices in the organizations (The Sustainable Development Goals Report, 2019). The UN has also emphasized the development of metrics and indicators for measuring sustainable development at the environmental, social, and economic levels. On the other end, consumers are also becoming proactive and demanding eco-friendly products (Green *et al.*, 2015). These requirements by the government, the UN, and customers are building pressure on LSPs.

Many authors have discussed the hurdles in the adoption of sustainable practices and highlighted methods for measuring the performance of sustainable supply chains (Taticchi *et al.*, 2013; Baird, 2017; Thanki; Thakkar, 2018). Pazirandeh and Jafari (2013) critically evaluated the impact of sustainability strategies on logistics efficiency using structural equation modeling (SEM). Tajbakhsh and Hassini (2015) proposed a performance measurement framework for measuring the sustainability of various industries based on seven sustainability dimensions (economic, environmental, social, reputable, valuable, equitable, and sustainable). Trianni *et al.* (2019) analyzed 26 small and medium manufacturing organizations of Germany and Italy and observed that the organizations accorded more importance to economic indicators over environmental and social indicators. They developed a novel framework for measuring sustainability indicators. Gardas *et al.* (2019) analyzed sustainable supply chain determinants for measuring the business performance of the oil and gas supply chain. Chowdhury and Paul (2020) reviewed the

applications of multi-criteria decision-making methods (MCDM) for measuring performance, including enablers of, and barriers to, corporate sustainability.

Most LSPs struggle in evaluating their sustainability index due to the lack of appropriate frameworks (Chhabra and Singh, 2016). It is crucial to measure the performance of LSPs in terms of their sustainability. This study proposes a framework for measuring the sustainability index of LSPs. This will help other LSPs understand their current position and take corrective measures if required. This provides a significant opportunity to conduct an in-depth analysis of the sustainable practices followed by LSPs in developing countries and to evaluate their sustainability index to understand the current situation of LSPs in the Indian market. Therefore, the following research questions have been developed:

RQ1. What are the sustainable practices followed by LSPs in developing countries to fulfill the UN's sustainable development goals?

RQ2. How can the sustainability of Indian LSPs be measured?

RQ3. What suggestions can be made to other LSPs for better achievement of SDG goals?

The sustainable practices followed by LSPs are identified through a literature review as discussed in the next section and a case study is also conducted on Indian LSP to understand their sustainable practices closely as discussed in Section 3.

2.1 Identification of sustainable practices from literature review

Previous studies have explored sustainable practices adopted by LSPs specifically in context to developing countries. It has been observed that LSPs have emphasized on the reduction of carbon emissions and fuel consumption as an initiative towards the environmental protection (Yao *et al.*, 2015; Aldakhil *et al.*, 2018; Liu *et al.*, 2018). Agrawal *et al.* (2016) have explored the importance of reverse logistics to assist return management and to make optimum use of resources. Renewable resources like solar and hydro products are more preferred nowadays over expensive single time use products for long term sustainability of products (Dubey *et al.*, 2017; Sureeyatanapas *et al.*, 2018). Reuse and recycling of packaging material and pallets are also done through plastic totes and heavy containers in order to save resources (Boubeta *et al.*, 2018). The adoption of green practices in the logistics sector has also become evident by converting manual processes (paper-work) into e-processes (paper-less) through digitalization (Kayikci, 2018). The following sustainable practices are identified through the literature review:

(1) *Use of reusable or recyclable packaging:* To promote green practices, several LSPs have started using plastic totes and heavy-duty containers which are cost-efficient and can be used for multiple trips (Centobelli *et al.*, 2018; Regattieri *et al.*, 2019). Paper, cardboard, corn starch, and bubble wrap, etc. are biodegradable and preferred for packaging (Gupta *et al.*, 2020). In India, Transport Corporation of India (TCI) and Agarwal Packers and Movers Ltd. are actively reusing and recycling packaging materials for making cost-effective and sustainable logistics system (Bhushan and Malviya, 2018)

(2) *Use of renewable energy sources:* The best way to conserve limited resources is to maximize the use of renewable energy sources. LSPs have initiated the use of solar energy and hydropower for many logistical activities instead of single-time use electric ones. Installation of solar panels for producing electricity (Fontes, and Freires, 2018; Gupta *et al.*, 2018), transparent roofs for natural light to enter during daytime (Tascioglu and Keser, 2019) and storage of rainwater for gardening and cleaning vehicles (Piecyk and Björklund, 2015; Gupta *et al.*, 2018) indicate the prominent use of natural resources by LSPs. All Cargo Logistics sets a wonderful example by providing renewable energy transportation from end to end logistics (Choudhary, 2019).

(3) *Green procurement:* Green procurement is the foremost step towards green logistics (Teixeira *et al.*, 2018). LSPs prefer to purchase products or services, which will cause a minimum adverse effect on the environment. The use of alternative fuel, proper planning for unutilized material, waste reduction, and use of recyclable and reusable products can support green procurement initiatives (Meehan and Bryde, 2011; Centobelli *et al.*, 2018). DHL has adopted green logistics to make our planet more sustainable by reducing emissions and moving towards the circular economy.

(4) *Reduction in carbon emissions:* LSPs have taken initiatives for low carbon supply chain integrations. This initiative has a positive impact on LSPs' environmental and financial performance (Qian *et al.*, 2019). LSPs are required to take appropriate solutions for carbon emission reduction and to develop strategies for a sustainable environment (Guiffrida *et al.*, 2011; Tang *et al.*, 2015; Aldakhil *et al.*, 2018; Liu *et al.*, 2018). Damco, DHL and TNT are offering green services and reducing carbon footprints through various services (Das and Kulkarni, 2018).

(5) *Use of CNG/Electric fleet:* Many LSPs are making use of CNG vehicles to reduce cost, pollution, and fuel consumption (Gupta *et al.*, 2018). The use of CNG/electric vehicles (fuel-free) helps in reducing the noise level and making the air clean. LSPs are also looking forward to adopting e-vehicles in the coming years (McKinnon, 2010; Liu *et al.*, 2018). Almost all Indian LSPs including Safe Express, CTC freight carriers Ltd. and Sal Logistics Pvt. Ltd., etc. are using more of the CNG vehicles to save cost and fuel.

(6) *Resource optimization:* Due to the scarcity of resources, LSPs are looking forward to the alternative options to use limited resources in an optimum way (Centobelli *et al.*, 2017). Conservation of resources will help in the smooth functioning of logistics operations in the future. LSPs are optimizing resources by making maximum use of natural resources over man-made products (Choudhury *et al.*, 2018; He *et al.*, 2018; Winkler and Zinsmeister, 2019). Om Logistics Ltd. conserves its natural resources by making maximum use of solar light and rainwater harvesting in their warehouses.

(7) *Reduction in fuel consumption:* Less fuel consumption by LSPs is considered as economic, social and environmental initiative (Blanco and Sheffi, 2017). Automobile companies are adding fuel-saving features in new vehicles by using new engine technologies. Moreover, the transition from fuel to electric vehicles is a great initiative for reducing fuel consumption and environmental pollution (Tang *et al.*, 2015; Yao *et al.*, 2015). Although, in India, there are many infrastructural bottlenecks to implement e-vehicles entirely, but the government is working in this direction. Jayem Logistics has significantly contributed towards eco-friendly transportation, reduction in fuel consumption, and pollution.

3. Study of sustainable practices in context to an Indian LSP

Most of the identified sustainable practices from the literature review are in context to both developing and developed countries. There is, therefore, a need to understand and identify the sustainable practices of Indian LSPs more closely to make this study more relevant in the Indian context. A real case illustration on Indian LSP has been discussed to have a thorough understanding of sustainable practices followed by them. The information is collected based on multiple visits and discussions with middle and senior management of the case organization.

3.1 Profile of case organization (ABC ltd.)

ABC Ltd. (given name) started its business as a multi-modal transportation organization in the mid-1980s in India. Its objective is to provide logistics support to the manufacturing organizations in the nation as well as across boundaries. With an annual turnover of around INR 1,500 crores (the US$ 225.47 million), it serves all the significant businesses sectors such as automobile, hardware, telecom, retail and pharmaceuticals.

The organization has a wide distribution network of more than 3,500 client base of domestic, transnational, and international presence majorly in the corporate B2B sphere. It provides service to most of the automobile manufacturers and more than 2000 auto segment part producers in India. Their customer base includes all significant brand names like Mahindra and Mahindra, Maruti Suzuki India Ltd., SUZUKI Motors, India Yamaha Motors Pvt. Ltd., Toyota, Daimler, Honda Motors India, Heromoto Corp, Tata Motors, Samsung Electronics India Limited, VIVO Mobiles India, Bharati Airtel, Cadila Pharmaceuticals, Arvind Fashions, etc. ABC Ltd. is conferred with many awards, recognitions, and excellence awards for its outstanding contribution to the logistics sector both in domestic and international services.

3.2 Organizational assets

ABC Ltd. ensures timely delivery and assured services to the clients through its technologically advanced assets. The organization offers modernized multi-client and dedicated warehousing spaces over more than 100 locations with its captive warehousing strength of more than 20 lac square meters in PAN India. It manages hub and spoke distribution through more than 40 Cross-Dock centers and 700 branches for Part Truck Load (PTL) distribution. This Indian LSP delivers more than 14,000 pin codes pan India and is serving over 85 countries through its own offices, joint ventures, exclusive partners, etc.

The case organization offers door to door customized deliveries through more than 4,500 modern technologically advanced, all-weather containerized vehicles of varied capacities ranging from 0.5 metric tons to 40 metric tons. It has more than 5,000 trained workforce, who work sincerely 24 by 7 for committed services, and customer query resolution. ABC Ltd. is equipped with advanced in-house created applications and SAP software along with ORACLE based databases for providing timely and accurate information to all stakeholders. It also uses modern advanced information technology tools and real-time applications to track and trace consignments through owned servers.

3.3 Organizational processes

The organization's order processing is made effective with invoice creation on behalf of their customers, while also creating and managing receipts for storage and dispatch of shipments, and sales tax or excise duties. It also incorporates key parameters like time and motion studies for optimization of resources, designing of warehouse layout, 5S support, and adaptable stock piling arrangements by using the latest and coordinated IT tools for the smooth functioning of warehouses. For managing inventory, the organization prefers to arrange their stocks by using concepts of LIFO (Last in First Out), FIFO (First in First Out), etc. This LSP also encourages the usage of mathematical models along with IT tools for analyzing the inventory costs, transportation costs, and stock management.

An ideal mix of roadways, railways, and the airways are offered by the organization to handle customer consignments and to serve them as per commitments. It offers flexibility to the customers in terms of selection of cargo, selection of location, the type of product and the size of shipments. This LSP also provides value-added services which include services like labeling and kitting of goods, packaging with bubble wrap or directly in carton boxes, and

sorting of goods. Document maintenance, packaging redesigning, consultancy on supply chain optimization, and return logistics are also provided.

3.4 Sustainability measures

In order to conserve natural resources, ABC Ltd. is carrying out numerous activities to manage resources at the ecological, societal and economic levels. The organization is also certified with ISO 14001:2015 certifications for maintaining green practices for sustainability.

Some of the sustainable practices that it follows are as under:

(1) *Reduction in carbon emissions:* All the trucks are outfitted with GPS technology which helps in tracking the exact location, while also enabling reduction in carbon emissions. The organization takes exceptional consideration for vehicles which have valid pollution check and carbon emission certificates.

(2) *Green practices:* ABC Ltd. plants more than 5,000 trees annually to enhance green cover around their warehouses. There is a culture of planting tree saplings on various occasions as a step towards environmental conservation. Air sanitizing plants are also grown around warehouses or distribution centers to purify the air and make surroundings clean.

(3) *Digitalization of processes:* The case organization is adopting process automation, in order to change its manual transactions to digital processes. This initiative not just reduces the use of paper and makes document handling easier but also brings transparency to the system.

(4) *Use of solar energy:* To optimize resources, ABC Ltd. has adopted proper conservation of natural resources. Additionally, the organization has installed solar panels on the roofs of most of the warehouses to generate power and to give back to society by saving electricity.

(5) *Use of natural/Renewable resources:* With preference to use natural light, the organization has built transparent roofs at their facilities. This reduces carbon emission and helps in saving electricity.

(6) *Rainwater harvesting:* Committed to green practices, ABC Ltd. utilizes rainwater stored in its underground tanks for further use in washing vehicles, cleaning the floors, or gardening purposes.

(7) *Education initiatives:* In line with PMKVY (Prime Minister Kaushal Vikas Yojna, Government of India scheme), the organization is running its non-profit education and training institute to increase employability of individuals in the logistics industry. Under this initiative, it offers various short term courses related to logistics and supply chain along with practical training sessions. The meritorious candidates are offered to work with the organization according to their abilities and domain knowledge.

(8) *Healthy working environment:* Offering employee friendly policies and environment, this LSP provides its employees with flexible working hours. In order to provide a comfortable and sound environment, residential and mess services are provided to the employees.

3.5 Selection of sustainable practices

After identifying thirteen sustainable practices from the literature review and case analysis, a dedicated team of three experts with over ten years of experience and working with leading

Indian LSPs, helped in finalizing the sustainable practices in Indian context. In order to achieve this, a list of all identified sustainable practices was shared with experts and their inputs were solicited on application of these practices in their respective organizations and ensuing impact on performance. The level of application of these sustainable practices in their organizations helped in understanding the importance of sustainability in Indian context. After expert inputs, nine sustainable practices were finalized. Out of these nine, five were identified from literature, one was from the case study and three were found to be common in both case study and literature review. The final list of sustainable practices after discussing with Indian experts is given in Table 1. As the objective of this study is to evaluate the sustainability index, therefore, the GTMA is applied for evaluating the sustainability index. The detailed research methodology will be discussed in the next section.

4. Evaluation of sustainability index- Graph Theory Matrix Approach (GTMA)

After identifying sustainability practices, a framework for the evaluation of the sustainability index of an LSP is proposed in this study. The Graph Theory Matrix Approach has been applied for the evaluation of the sustainability index, which considers any number of quantitative and qualitative sustainable practices simultaneously. The conventional methods like cause and effect diagram, process flow charts and block diagrams are not

Table 1. Sustainable practices followed by logistics service providers

S. No	Sustainable practices	Description	References	Source
1	Use of recyclable packaging (*S*1)	Using recyclable and reusable raw material for packaging to promote green practices	Gupta *et al.* (2020), Regattieri *et al.* (2019), Centobelli *et al.* (2018), Jilani *et al.* (2018), Lew *et al.* (2018), Sureeyatanapas *et al.* (2018), Chhabra *et al.* (2017)	Literature review
2	Use of renewable energy sources (*S*2)	Making use of renewable resources for long term sustainability	Taşçıoglu and Keser (2019), Fontes, and Freires (2018), Gupta *et al.* (2018), Piecyk and Björklund (2015), McKinnon (2010)	Literature review and Case study
3	Green procurement (*S*3)	Using green practices for purchasing	Centobelli *et al.* (2018), Teixeira *et al.* (2018), Meehan and Bryde (2011)	Literature review
4	Reduction in carbon emissions (*S*4)	Reducing carbon emission to protect the environment	Aldakhil *et al.* (2018), Centobelli *et al.* (2018), Gupta *et al.* (2018), Lew *et al.* (2018), Tang *et al.* (2015), Guiffrida *et al.* (2011)	Literature review and Case study
5	Use of CNG/ Electric/ hybrid fleet (*S*5)	Using Eco-friendly fleet to reduce the pollution and consumption of fuel	Liu *et al.* (2018), Gupta *et al.* (2018), McKinnon (2010)	Literature review
6	Resource optimization (*S*6)	Optimum use of available resources for long term business continuity	Winkler and Zinsmeister (2019), He *et al.* (2018), Choudhury *et al.* (2018), Centobelli *et al.* (2017)	Literature review
7	Digitization of processes (*S*7)	Adoption of paperless processes and use of latest technologies	Singh *et al.* (2019)	Literature review and Case study
8	Rainwater harvesting (*S*8)	Using rainwater for gardening or washing vehicles	Gupta *et al.* (2018)	Case study
9	Reduction in fuel consumption (*S*9)	Reducing the consumption of fuel for conserving resources	Blanco and Sheffi (2017), Yao *et al.* (2015), Tang *et al.* (2015)	Literature review

suitable for further mathematical analysis whereas this method establishes the directional relationship among variables and also take care of their interdependencies by mathematical analysis (Raj *et al.*, 2010; Jain and Raj, 2016; Hudnurkar; Ambekar, 2019).

It is a more preferred approach over conventional methods for representation and quantification. Digraph assists in the graphical representation of variables, which supports better visual analysis and provides valuable inputs for mathematical modeling. Many researchers have applied the GTMA in different fields. Franceschini *et al.* (2006) have used this approach for finding the quality index. Darvish *et al.* (2009) applied GTMA for the ranking of contractors. Attri *et al.* (2014) used this method to compute the intensity of TPM (Total Productive Maintenance) barriers. Apart from manufacturing, many authors have applied this approach in diversified areas such as to model the various mechanisms required for supply chain coordination (Kaur *et al.*, 2006), to evaluate the technology-based self-service banking service quality (Sindhwani and Goel, 2015), to analyze disposition decisions in reverse logistics (Agrawal *et al.*, 2016), to find the agility index of supply chains (Singh *et al.*, 2017), for quantifying the barriers in the implementation of technologies (Bhandari *et al.*, 2019) and to measure the flexibility index for supply chains (Singh and Kumar, 2019), etc.

It is found to be a very important decision-making tool that helps in creating associations among variables and evaluating by forming an index. Based on the literature review and case study, nine sustainable practices have been identified in this study as shown in Table 1. This approach includes various steps of developing a digraph between variables depending upon their mutual correlations. The research methodology framework is shown in Figure 1. Further, a permanent function is computed by developing a permanent matrix by using Equation (1).

4.1 Variables digraph

The sustainable practices adopted by LSPs are represented in the form of nodes and edges by developing a variable digraph as shown in Figure 2. The nodes represent the sustainable practices and the edges represent their linkages. S_i shows the inheritance of practices and s_{ij} shows the degree of dependence of jth practice on ith practice. It also indicates the directed edge from node i to node j. The linkages among all sustainable practices taken in this study are shown in Figure 2. The digraph for detailed analysis has been discussed in Figure 3.

The permanent function of sustainable practices (Perm S) shown in Figure 3 is written as:

$$\text{Perm } S = \begin{vmatrix} S1 & s_{12} & s_{13} & s_{14} & s_{15} & s_{16} & s_{17} & s_{18} & s_{19} \\ s_{21} & S2 & s_{23} & s_{24} & s_{25} & s_{26} & s_{27} & s_{28} & s_{29} \\ s_{31} & s_{32} & S3 & s_{34} & s_{35} & s_{36} & s_{37} & s_{38} & s_{39} \\ s_{41} & s_{42} & s_{43} & S4 & s_{45} & s_{46} & s_{47} & s_{48} & s_{49} \\ s_{51} & s_{52} & s_{53} & s_{54} & S5 & s_{56} & s_{57} & s_{58} & s_{59} \\ s_{61} & s_{62} & s_{63} & s_{64} & s_{65} & S6 & s_{67} & s_{68} & s_{69} \\ s_{71} & s_{72} & s_{73} & s_{74} & s_{75} & s_{76} & S7 & s_{78} & s_{79} \\ s_{81} & s_{82} & s_{83} & s_{84} & s_{85} & s_{86} & s_{87} & S8 & s_{89} \\ s_{91} & s_{92} & s_{93} & s_{94} & s_{95} & s_{96} & s_{97} & s_{98} & S9 \end{vmatrix}$$

4.2 Quantification of S_is and s_{ij}s

To find out the value of the permanent function (Perm S), there is a need to quantify S_is and s_{ij}s. The qualitative measure of inheritance of practices (S_i) is rated on a scale of 1–5 by experts by considering 1 to be the lowest and 5 to be the highest. The qualitative measures of interdependencies among sustainable practices (s_{ij}) have been quantitatively measured by finding the relative dependence on each other based on the value assigned by experts and

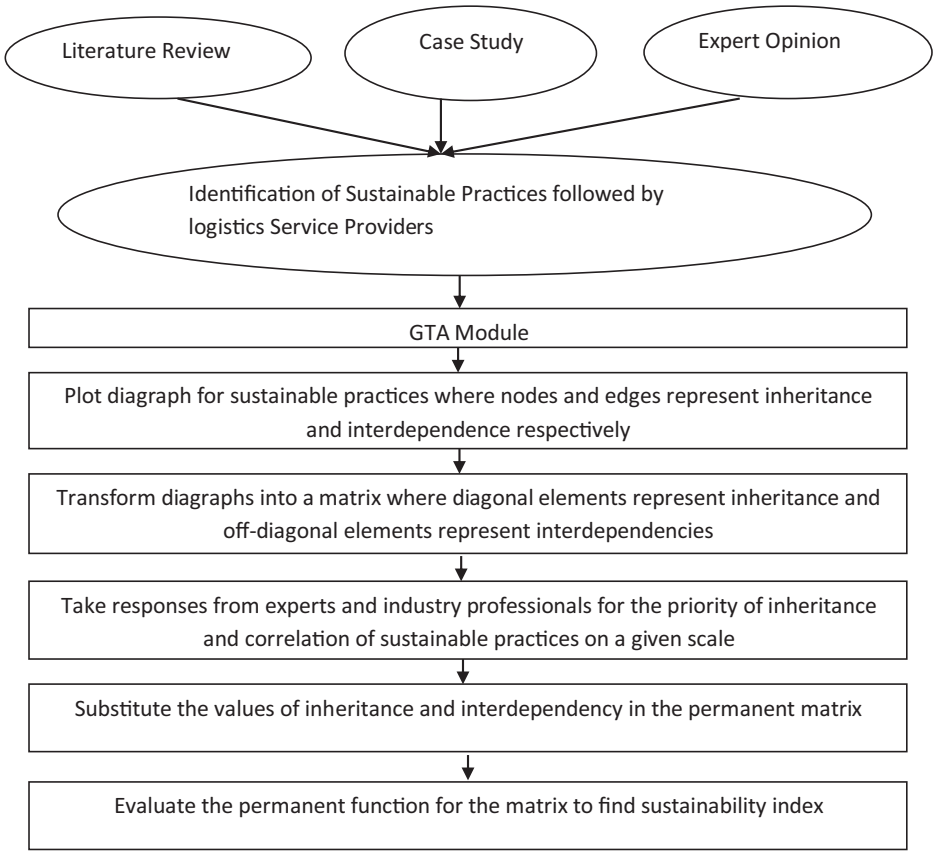


Figure 1.
Research methodology framework

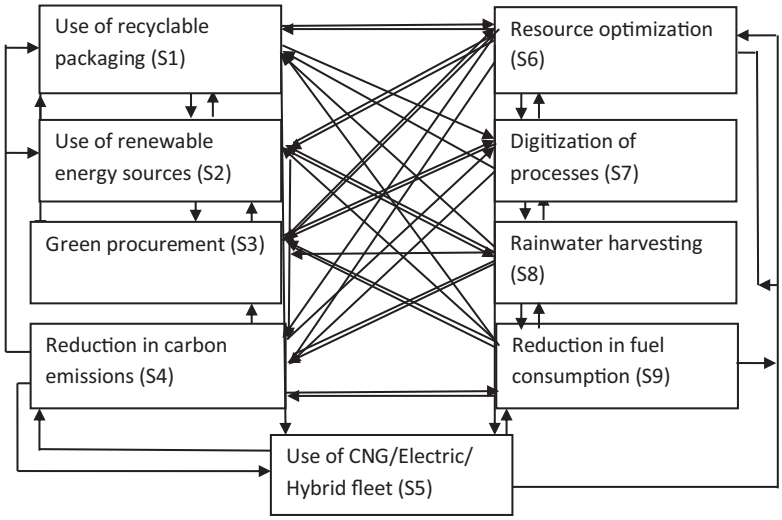


Figure 2.
Interaction among sustainable practices followed by LSPs

discussed in Table 2. A structured questionnaire was designed and sent to all concerned. The experts have been asked to respond on the importance given to sustainable practices (*S*1 to *S*9) and also to rate the relative dependence of one sustainable practice on another according to criteria given in Table 2. For instance, s_{12}- dependence of sustainable practice 2 on sustainable practice 1 which means the degree of dependence of use of energy renewable resources (*S*2) on use of recyclable packaging (*S*1) and similarly, the relative dependence will find for all other possible combinations.

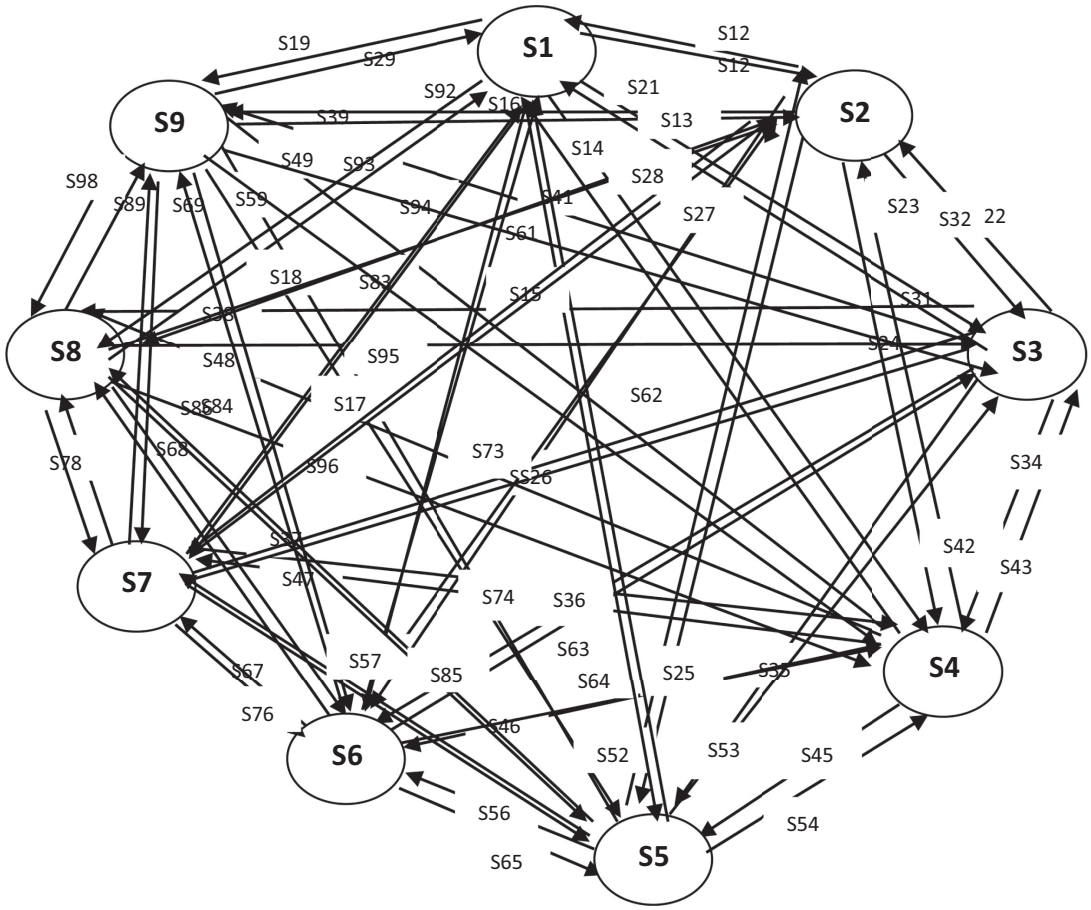


Figure 3.
Digraph representation of elements of permanent matrix

Qualitative measures of interdependencies	Relative dependence of sustainable practices s_{ij}	s_{ji} = 10-sij
Two attributes are of equal importance	5	5
One attribute is slightly dependent on the other	6	4
One attribute is very dependent on the other	7	3
One attribute is most dependent on the other	8	2
One attribute is extremely dependent on the other	9	1
One attribute is exceptionally dependent on the other	10	0

Source(s): Muduli *et al.* (2013)

Table 2.
Relative dependence of sustainable practices (s_{ij}s)

4.3 Variables' permanent matrix

The permanent function is a mathematical expression which helps to determine the sustainability index of LSPs. The calculation for the permanent function is almost the same as finding the determinant of the matrix but considers only positive terms. The value of the permanent function of sustainable practices (Perm *S*) is computed by using Equation (1) which is given as follows.

$$
\begin{aligned}
perm S = {} & \prod_{i=1}^{9} Si + \sum_{i=1}^{8}\sum_{j=i+1}^{9}\sum_{k=j+1}^{3}\sum_{l=k+1}^{4}\sum_{m=l+1}^{5}\sum_{n=m+1}^{6}\sum_{o=n+1}^{7}\sum_{p=o+1}^{8}\sum_{q=p+1}^{9}(s_{ij}s_{ji})S_kS_lS_mS_nS_nS_pS_q \\
& + \sum_{i=1}^{7}\sum_{j=1+1}^{8}\sum_{k=j+1}^{9}\sum_{l=k+1}^{4}\sum_{m=l+1}^{5}\sum_{n=m+1}^{6}\sum_{o=n+1}^{7}\sum_{p=o+1}^{8}\sum_{q=p+1}^{9}(s_{ij}s_{jk}s_{ki} + s_{ik}s_{kj}s_{ji})S_lS_mS_nS_oS_pS_q + \\
& \left[\begin{array}{l}
\sum_{i=1}^{6}\sum_{j=1+1}^{9}\sum_{k=j+1}^{8}\sum_{l=i+2}^{9}\sum_{m=1}^{8}\sum_{n=m+1}^{9}\sum_{o=n+1}^{9}\sum_{p=o+1}^{9}\sum_{q=p+1}^{9}(s_{ij}s_{jk})(s_{kl}s_{lk})S_mS_nS_oS_pS_q \\
+\sum_{i=1}^{6}\sum_{j=i+1}^{8}\sum_{k=j+1}^{9}\sum_{l=j+1}^{9}\sum_{m=1}^{8}\sum_{n=m+1}^{9}\sum_{n=m+1}^{9}\sum_{o=n+1}^{9}\sum_{p=o+1}^{9}\sum_{q=p+1}^{9}(s_{ij}s_{jk}s_{kj}s_{li} + s_{il}s_{lk}s_{kj}s_{ij})S_mS_nS_oS_pS_q
\end{array}\right] \\
& + \left[\begin{array}{l}
\sum_{i=1}^{7}\sum_{j=1+1}^{8}\sum_{k=j+1}^{9}\sum_{l=1}^{8}\sum_{m=l+1}^{9}\sum_{n=m+1}^{9}\sum_{o=n+1}^{9}\sum_{p=o+1}^{9}\sum_{q=p+1}^{9}(s_{ij}s_{jk}s_{ki} + s_{ik}s_{kj}s_{ji})(s_{lm}s_{ml})S_nS_oS_pS_q \\
+\sum_{i=1}^{5}\sum_{j=1+1}^{8}\sum_{k=i+1}^{9}\sum_{l=j+1}^{9}\sum_{m=1}^{8}\sum_{n=m+1}^{9}\sum_{o=n+1}^{9}\sum_{p=o+1}^{9}\sum_{q=p+1}^{9}(s_{ij}s_{jk}s_{kl}s_{lm}s_{mi} + s_{im}s_{ml}s_{lk}s_{kj}s_{ij})S_nS_oS_pS_q
\end{array}\right] \\
& + \left[\begin{array}{l}
\sum_{i=1}^{6}\sum_{j=1}^{8}\sum_{k=i+1}^{9}\sum_{l=j+1}^{9}\sum_{m=1}^{8}\sum_{n=m+1}^{9}\sum_{o=n+1}^{9}\sum_{p=o+1}^{9}\sum_{q=p+1}^{9}(s_{ij}s_{jk}s_{kl}s_{li} + s_{ik}s_{lk}s_{lk}s_{kj}s_{ji})(s_{mn}s_{nm})S_oS_pS_q \\
+\sum_{i=1}^{4}\sum_{j=i+1}^{8}\sum_{k=i+1}^{9}\sum_{l=j+1}^{9}\sum_{m=1}^{8}\sum_{n=m+1}^{9}\sum_{o=n+1}^{9}\sum_{p=o+1}^{9}\sum_{q=p+1}^{9}(s_{ij}s_{jk}s_{ki} + s_{ik}s_{kj}s_{ji})(s_{lm}s_{mn}s_{nl} + s_{ln}s_{nm}s_{ml})S_oS_pS_q \\
+\sum_{i=1}^{4}\sum_{j=i+1}^{9}\sum_{k=j+1}^{6}\sum_{l=i+1}^{9}\sum_{m=k+1}^{8}\sum_{n=k+2}^{9}\sum_{o=n+1}^{9}\sum_{p=n+2}^{9}\sum_{q=p+1}^{9}(s_{ij}s_{ji})(s_{kl}s_{lk})(s_{mn}s_{nm})S_oS_pS_q \\
+\sum_{i=1}^{4}\sum_{j=i+1}^{8}\sum_{k=j+1}^{9}\sum_{l=i+1}^{9}\sum_{m=i+1}^{9}\sum_{n=j+1}^{9}\sum_{o=j+1}^{9}\sum_{p=j+1}^{9}\sum_{q=j+1}^{9}(s_{ij}s_{ji})(s_{ij}s_{jk}s_{kl}s_{lm}s_{mn}s_{ni} + s_{in}s_{nm}s_{lk}s_{kj}s_{ji})(s_{mn}s_{nm})S_oS_pS_q + \ldots
\end{array}\right]
\end{aligned}
\tag{1}
$$

Substituting the values of the variables in the permanent matrix and compute the value of the permanent matrix (Perm S) by putting values in Equation (1):

$$
\text{Perm } S = \begin{vmatrix}
S1 & s_{12} & s_{13} & s_{14} & s_{15} & s_{16} & s_{17} & s_{18} & s_{19} \\
s_{21} & S2 & s_{23} & s_{24} & s_{25} & s_{26} & s_{27} & s_{28} & s_{29} \\
s_{31} & s_{32} & S3 & s_{34} & s_{35} & s_{36} & s_{37} & s_{38} & s_{39} \\
s_{41} & s_{42} & s_{43} & S4 & s_{45} & s_{46} & s_{47} & s_{48} & s_{49} \\
s_{51} & s_{52} & s_{53} & s_{54} & S5 & s_{56} & s_{57} & s_{58} & s_{59} \\
s_{61} & s_{62} & s_{63} & s_{64} & s_{65} & S6 & s_{67} & s_{68} & s_{69} \\
s_{71} & s_{72} & s_{73} & s_{74} & s_{75} & s_{76} & S7 & s_{78} & s_{79} \\
s_{81} & s_{82} & s_{83} & s_{84} & s_{85} & s_{86} & s_{87} & S8 & s_{89} \\
s_{91} & s_{92} & s_{93} & s_{94} & s_{95} & s_{96} & s_{97} & s_{98} & S9
\end{vmatrix}
$$

$$
= \begin{vmatrix}
4 & 5 & 9 & 8 & 5 & 9 & 5 & 5 & 9 \\
5 & 3 & 7 & 9 & 6 & 9 & 5 & 8 & 7 \\
1 & 3 & 2 & 9 & 9 & 7 & 9 & 6 & 8 \\
2 & 1 & 1 & 2 & 9 & 8 & 8 & 7 & 9 \\
5 & 4 & 1 & 1 & 4 & 7 & 5 & 5 & 9 \\
1 & 1 & 3 & 2 & 3 & 4 & 9 & 9 & 9 \\
5 & 5 & 1 & 2 & 5 & 1 & 2 & 5 & 5 \\
5 & 2 & 4 & 3 & 5 & 1 & 5 & 2 & 6 \\
1 & 3 & 2 & 1 & 1 & 1 & 5 & 4 & 3
\end{vmatrix}
$$

Perm $S^* = 2,30,595,214,680$

The best possible value of the index can be computed by taking the highest possible value of all variables. The value of inheritance of all sustainable practices is kept maximum, that is, 5 and the diagonal cells of the permanent matrix are kept 5 as maximum. This maximum index can be used as a reference for benchmarking of sustainability performance. Substituting all values in Equation (1) and the maximum theoretical index can be:

$$\text{Perm } S = \begin{vmatrix} 5 & 5 & 9 & 8 & 5 & 9 & 5 & 5 & 9 \\ 5 & 5 & 7 & 9 & 6 & 9 & 5 & 8 & 7 \\ 1 & 3 & 5 & 9 & 9 & 7 & 9 & 6 & 8 \\ 2 & 1 & 1 & 5 & 9 & 8 & 8 & 7 & 9 \\ 5 & 4 & 1 & 1 & 4 & 7 & 5 & 5 & 9 \\ 1 & 1 & 3 & 2 & 3 & 5 & 9 & 9 & 9 \\ 5 & 5 & 1 & 2 & 5 & 1 & 5 & 5 & 5 \\ 5 & 2 & 4 & 3 & 5 & 1 & 5 & 5 & 6 \\ 1 & 3 & 2 & 1 & 1 & 1 & 5 & 4 & 5 \end{vmatrix}$$

Perm $S^* = S^*_{\max} = 3,63,706,047,040$

S^* represents the sustainability index of LSPs. This value of sustainability index can be also used for comparison with other LSPs. It can help them in positioning themselves in terms of the adoption and implementation of sustainable practices. LSPs can analyze the importance and relationships among sustainable practices and can be implemented in their respective organizations for improving sustainability. The maximum value $S^*_{\max}$ represents the maximum index of sustainability that LSPs can achieve by adopting and implementing maximum sustainable measures.

5. Discussion

In this study, the GTMA is used to compute the sustainability index of LSPs. When the resources are limited and the environment is degrading at a very fast pace in the current market situation, the evaluation of sustainability index can help LSPs in finding suitable solutions to adopt and implement sustainable practices in their inbound and outbound processes. It can help LSPs to devise some strategies for improving the adoption and implementation of sustainable practices, in turn, positively impacting the performance of LSPs. The implementation of sustainable practices by LSPs will also help in reducing wastage of resources which further encourages the circular economy.

From the above case illustration, it has been observed that the sustainability index (S) of the given LSP is comparatively lesser than the best possible value of the sustainability index (S^*). This reflects that there is a scope of improvement for an LSP by implementing more sustainable practices in their logistical operations. LSPs can develop strategies for revision in their processes and can plan for improvement in terms of sustainability. On similar lines, Mangla *et al.* (2019) explored the agro-food supply chain by analyzing four Indian dairy-based organizations and developed strategies for improving the supply chain and logistics performance in emerging economies using the GTMA approach.

The sustainability index can also assist LSPs in positioning themselves and comparing them with their peer service providers. LSPs can develop their strategies for improving their performance through increased use of green practices. Several articles on the implementation of sustainable practices have discussed their importance in developing countries. For the successful implementation of sustainable practices, LSPs should come forward to identify the issues related to sustainability implementation and should formulate a proper strategy for handling them effectively. Similarly, Agrawal *et al.* (2016) evaluated the outsourcing index, which is very useful for managers and practitioners in formulating strategies for outsourcing reverse logistics decisions. In Bangladesh, Moktadir *et al.* (2018) applied graph theory for

identifying barriers for implementing sustainable manufacturing practices in the leather industry. They have also suggested that the managers should formulate strategies for optimum utilization of limited resources to reduce wastage. It also validates our findings in context to Indian LSP.

5.1 Managerial implications

The major implication of this study is that LSPs should select factors for evaluating the sustainability index depending upon the prevailing business environment. These factors may vary as per region and sector. Based on the sustainability index, LSPs may choose to find potential areas for improvement and a green logistics strategy can be developed accordingly. Organizations with low sustainability index can identify their weak areas and can work upon them to increase their sustainability index. Organizations with high sustainability index can compare their values with maximum possible value and can further work in the direction to achieve that level. This framework will be helpful for LSPs in benchmarking their performance with global standards in the era of the circular economy.

5.2 Limitations and future scope

In this study, a case analysis has been done for understanding the sustainable practices of LSPs. Findings from a case may vary with change in sector and location. Therefore, for the generalization of these findings, this study can be further validated by taking more case studies on LSPs in different contexts. This work can be extended for small LSPs as the future scope of the existing work. Apart from considered criteria, other sustainability criteria in context to an emerging circular economy can be also considered for further modification of framework. Further, the results can be compared after evaluating the sustainability index by different methods to check the robustness of the framework. The study may be further extended to comparing the sustainability index of logistics companies from developing and developed countries. It will help in the identification and adoption of advanced sustainability practices by LSPs of developing countries for continuous improvement. The study will not only encourage the organizations but also the society to conserve resources and make proper utilization of available resources.

6. Conclusion

In today's business environment of the circular economy, almost every organization is taking initiatives for reducing environmental degradation and increasing the sustainable use of resources. LSPs are continuously exploring innovative and sustainable ways to contribute to society and the environment. Sustainable practices adopted and implemented by various LSPs have been identified from previous studies. The case study method is also adopted to make the current study more relevant and realistic for LSPs. A real case study on an Indian LSP has been presented to clarify their processes and adoption of sustainable practices in their logistical activities. The information is based on observations and discussions with middle and senior management during multiple visits to the case organization.

This study makes two major contributions. The first is the identification of major sustainable practices in the logistics sector. The second is a framework for the evaluation of the sustainability index. The use of recyclable packaging, renewable energy sources, green procurement, carbon emissions reduction, use of CNG fleets, resource optimization, digitalization of processes, rainwater harvesting and reduction in fuel consumption are the major sustainable practices. Although previous studies have explored many performance measurement frameworks, these frameworks are not so relevant in the era of the circular economy. This study included prominent sustainability practices while evaluating the

sustainability index. We have used GTMA approach for evaluation of the sustainability index of LSPs. The application of this framework is illustrated using the real case of Indian LSP. The sustainability index can also help LSPs in formulating strategies to achieve the UN's sustainability goals. It will motivate LSPs to implement green logistics practices in their operations. This framework will also be helpful for LSPs in comparing their performance against global standards.

References

Agrawal, S. and Singh, R.K. (2019), "Analyzing disposition decisions for sustainable reverse logistics: triple Bottom Line approach", *Resources, Conservation and Recycling*, Vol. 150, p. 104448.

Agrawal, S., Singh, R.K. and Murtaza, Q. (2016), "Outsourcing decisions in reverse logistics: sustainable balanced scorecard and graph theoretic approach", *Resources, Conservation and Recycling*, Vol. 108, pp. 41-53.

Aldakhil, A.M., Nassani, A.A., Awan, U., Abro, M.M.Q. and Zaman, K. (2018), "Determinants of green logistics in BRICS countries: an integrated supply chain model for green business", *Journal of Cleaner Production*, Vol. 195, pp. 861-868.

Ansari, Z.N. and Kant, R. (2017), "Exploring the framework development status for sustainability in supply chain management: a systematic literature synthesis and future research directions", *Business Strategy and the Environment*, Vol. 26 No. 7, pp. 873-892.

Anwer, A.S., Garza-Reyes, J.A. and Kumar, V. (2018), "Best supply chain management practices and high-performance firms", *International Journal of Productivity and Performance Management*, Vol. 67 No. 9, pp. 1482-1509.

Attri, R., Grover, S. and Dev, N. (2014), "A graph theoretic approach to evaluate the intensity of barriers in the implementation of total productive maintenance (TPM)", *International Journal of Production Research*, Vol. 52 No. 10, pp. 3032-3051.

Baird, K. (2017), "The effectiveness of strategic performance measurement systems", *International Journal of Productivity and Performance Management*, Vol. 66 No. 1, pp. 3-21.

Bask, A., Rajahonka, M., Laari, S., Solakivi, T., Töyli, J. and Ojala, L. (2018), "Environmental sustainability in shipper-LSP relationships", *Journal of Cleaner Production*, Vol. 172, pp. 2986-2998.

Bhandari, D., Singh, R.K. and Garg, S.K. (2019), "Prioritization and evaluation of barriers intensity for implementation of cleaner technologies: framework for sustainable production", *Resources, Conservation and Recycling*, Vol. 146, pp. 156-167.

Bhushan, R. and Malviya, S. (2018), "Packaging solutions: reduce, recycle & don't reuse", available at: https://economictimes.indiatimes.com/industry/indlgoods/svs/packaging/reduce-recycle-don treuse-future-as-clear-as-glass-on-packaging-solutions/articleshow/64848893.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst (accessed on March 12 2020).

Blanco, E.E. and Sheffi, Y. (2017), *Green Logistics in Sustainable Supply Chains*, Springer, Cham, pp. 147-187.

Boubeta, I.G., Vázquez, M.F., Caamaño, P.D. and Prado, J.C.P. (2018), "Economic and environmental packaging sustainability: a case study", *Journal of Industrial Engineering and Management*, Vol. 11 No. 2, pp. 229-238.

Centobelli, P., Cerchione, R. and Esposito, E. (2017), "Developing the WH2 framework for environmental sustainability in logistics service providers: a taxonomy of green initiatives", *Journal of Cleaner Production*, Vol. 165, pp. 1063-1077.

Centobelli, P., Cerchione, R. and Esposito, E. (2018), "Environmental sustainability and energy-efficient supply chain management: a review of research trends and proposed guidelines", *Energies*, Vol. 11 No. 2, p. 275.

Chhabra, D. and Singh, R.K. (2016), "Justification of green logistic practices in Indian enterprises using analytic hierarchy process", *International Journal of Logistics Systems and Management*, Vol. 25 No. 3, pp. 295-312.

Chhabra, D., Garg, S.K. and Singh, R.K. (2017), "Analyzing alternatives for green logistics in an Indian automotive organization: a case study", *Journal of Cleaner Production*, Vol. 167, pp. 962-969.

Choudhary, S. (2019), "Renewable energy equipment transportation: an opportunity for end-to- end logistics", available at: https://energy.economictimes.indiatimes.com/energy-speak/renewable-energy-equipment-transportation-an-opportunity-for-end-to-end-logistics/3758 (accessed March 15 2020).

Choudhury, N., Raut, R.D., Gardas, B.B., Kharat, M.G. and Ichake, S. (2018), "Evaluation and selection of third party logistics services providers using data envelopment analysis: a sustainable approach", *International Journal of Business Excellence*, Vol. 14 No. 4, pp. 427-453.

Chowdhury, P. and Paul, S.K. (2020), "Applications of MCDM methods in research on corporate sustainability", *Management of Environmental Quality: An International Journal*, Vol. 31 No. 2, pp. 385-405.

Coaffee, J. (2008), "Risk, resilience, and environmentally sustainable cities", *Energy Policy*, Vol. 36 No. 12, pp. 4633-4638.

Darvish, M., Yasaei, M. and Saeedi, A. (2009), "Application of the graph theory *and matrix m*ethods to contractor ranking", *International Journal of Project Management*, Vol. 27 No. 6, pp. 610-619.

Das, M. and Kulkarni, V. (2018), "How green is your supply chain?", available at: https://www.thehindubusinessline.com/economy/logistics/how-green-is-your-supply-chain/article20318230.ece1 (accessed March 13 2020).

Dubey, R., Gunasekaran, A., Papadopoulos, T., Childe, S.J., Shibin, K.T. and Wamba, S.F. (2017), "Sustainable supply chain management: framework and further research directions", *Journal of Cleaner Production*, Vol. 142, pp. 1119-1130.

Evangelista, P., Santoro, L. and Thomas, A. (2018), "Environmental sustainability in third-party logistics service providers: a systematic literature review from 2000–2016", *Sustainability*, Vol. 10 No. 5, p. 1627.

Ferreira, L., Putnik, G.D., Lopes, N., Lopes, A. and Cruz-Cunha, M.M. (2015), "A cloud and ubiquitous architecture for effective environmental sensing and monitoring", *Procedia Computer Science*, Vol. 64, pp. 1256-1262.

Fontes, C.H.D.O. and Freires, F.G.M. (2018), "Sustainable and renewable energy supply chain: a system dynamics overview", *Renewable and Sustainable Energy Reviews*, Vol. 82, pp. 247-259.

Franceschini, F., Galetto, M., Grover, S., Agrawal, V.P. and Khan, I.A. (2006), "Role of human factors in TQM: a graph theoretic approach", *Benchmarking: An International Journal*, Vol. 13 No. 4, pp. 447-468.

Gandhi, S., Mangla, S.K., Kumar, P. and Kumar, D. (2015), "Evaluating factors in implementation of successful green supply chain management using DEMATEL: a case study", *International Strategic Management Review*, Vol. 3 No. 1, pp. 96-109.

García-Dastugue, S. and Eroglu, C. (2019), "Operating performance effects of service quality and environmental sustainability capabilities in logistics", *Journal of Supply Chain Management*, Vol. 55 No. 3, pp. 68-87.

Gardas, B.B., Raut, R.D. and Narkhede, B. (2019), "Determinants of sustainable supply chain management: a case study from the oil and gas supply chain", *Sustainable Production and Consumption*, Vol. 17, pp. 241-253.

Green, K.W., Toms, L.C. and Clark, J. (2015), "Impact of market orientation on environmental sustainability strategy", *Management Research Review*, Vol. 38 No. 2, pp. 217-238.

Green, K.W., Inman, R.A., Sower, V.E. and Zelbst, P.J. (2019), "Impact of JIT, TQM and green supply chain practices on environmental sustainability", *Journal of Manufacturing Technology Management*, Vol. 30 No. 1, pp. 26-47.

Guiffrida, A.L., Datta, P., Dey, A., LaGuardia, P. and Srinivasan, M. (2011), "Building sustainability in logistics operations: a research agenda", *Management Research Review*, Vol. 34 No. 11, pp. 1237-1259.

Gupta, A. and Singh, R.K. (2020), "Managing operations by a logistics company for sustainable service quality: Indian perspective", *Management of Environmental Quality: An International Journal*, Vol. ahead-of-print. No. ahead-of-print, doi: 10.1108/MEQ-11-2019-0246.

Gupta, A., Singh, R.K. and Suri, P.K. (2018), "Prioritizing critical success factors for sustainable service quality management by logistics service providers", *Vision*, Vol. 22 No. 3, pp. 295-305.

Gupta, A., Singh, R.K. and Suri, P.K. (2020), "Prioritizing best practices for logistics service providers", in Suri, P. and Yadav, R. (Eds), *Transforming Organizations through Flexible Systems Management*, Springer, Singapore, pp. 257-275.

Halldorsson, A. (2019), "Actionable sustainability in supply chains", in Zijm, H., Klumpp, M., Regattieri, A. and Heragu, S. (Eds), *Operations, Logistics and Supply Chain Management*, Springer, Cham, pp. 191-203.

He, M., Shen, J., Wu, X. and Luo, J. (2018), "Logistics space: a literature review from the sustainability perspective", *Sustainability*, Vol. 10 No. 8, p. 2815.

Hudnurkar, M. and Ambekar, S.S. (2019), "Framework for measurement of supplier satisfaction", *International Journal of Productivity and Performance Management*, Vol. 68 No. 8, pp. 1475-1492.

Jain, V. and Raj, T. (2016), "Modeling and analysis of FMS performance variables by ISM, SEM and GTMA approach", *International Journal of Production Economics*, Vol. 171, pp. 84-96.

Jazairy, A. and Haartman, R. (2018), "Analysing the institutional pressures on shippers and logistics service providers to implement green logistics practices", *International Journal of Logistics Research and Applications*, pp. 1-41.

Jilani, A., Ali, Y. and Khan, M.W. (2018), "Greening of humanitarian supply chain with focus on logistics", *International Journal of Business Performance and Supply Chain Modelling*, Vol. 10 No. 1, pp. 49-66.

Kaur, A., Kanda, A. and Deshmukh, S.G. (2006), "A graph theoretic approach for supply chain coordination", *International Journal of Logistics Systems and Management*, Vol. 2 No. 4, pp. 321-341.

Kayikci, Y. (2018), "Sustainability impact of digitization in logistics", *Procedia Manufacturing*, Vol. 21, pp. 782-789.

Kumar, G. and Goswami, M. (2019), "Sustainable supply chain performance, its practice and impact on barriers to collaboration", *International Journal of Productivity and Performance Management*, Vol. 68 No. 8, pp. 1434-1456.

Lew, A.F.R., Chew, B.C. and Hamid, S.R. (2018), "Green logistics implementation factors: a study on a global logistics provider", *Journal of Advanced Manufacturing Technology*, Vol. 12 No. 1, pp. 115-128.

Liu, X., McKinnon, A.C. and Wei, N. (2018), "An analysis of energy-related CO 2 emissions from China's logistics industry", in Liu, X. (Ed.), *Environmental Sustainability in Asian Logistics and Supply Chains*, Springer, Singapore, pp. 3-19.

Majumdar, A. and Sinha, S. (2018), "Modeling the barriers of green supply chain management in small and medium enterprises", *Management of Environmental Quality: An International Journal*, Vol. 29 No. 6, pp. 1110-1122.

Mangla, S.K., Kumar, P. and Barua, M.K. (2015), "Risk analysis in green supply chain using fuzzy AHP approach: a case study", *Resources, Conservation and Recycling*, Vol. 104, pp. 375-390.

Mangla, S.K., Govindan, K. and Luthra, S. (2017), "Prioritizing the barriers to achieve sustainable consumption and production trends in supply chains using fuzzy Analytical Hierarchy Process", *Journal of Cleaner Production*, Vol. 151, pp. 509-525.

Mangla, S.K., Sharma, Y.K., Patil, P.P., Yadav, G. and Xu, J. (2019), "Logistics and distribution challenges to managing operations for corporate sustainability: study on leading Indian diary organizations", *Journal of Cleaner Production*, Vol. 238, p. 117620.

Marques, L. (2019), "Sustainable supply network management: a systematic literature review from a knowledge perspective", *International Journal of Productivity and Performance Management*, Vol. 68 No. 6, pp. 1164-1190.

Martinez-Jurado, P.J. and Moyano-Fuentes, J. (2014), "Lean management, supply chain management and sustainability: a literature review", *Journal of Cleaner Production*, Vol. 85, pp. 134-150.

McKinnon, A. (2010), "Green logistics: the carbon agenda", *Electronic Scientific Journal of Logistics*, Vol. 6 No. 3.

Meehan, J. and Bryde, D. (2011), "Sustainable procurement practice", *Business Strategy and the Environment*, Vol. 20 No. 2, pp. 94-106.

Mensah, J. and Casadevall, S. (2019), "Sustainable development: meaning, history, principles, pillars, and implications for human action: literature review", *Cogent Social Sciences*, Vol. 5 No. 1, p. 1653531.

Modgil, S., Gupta, S. and Bhushan, B. (2020), "Building a living economy through modern information decision support systems and UN sustainable development goals", *Production Planning and Control*, pp. 1-21.

Moktadir, M.A., Ali, S.M., Rajesh, R. and Paul, S.K. (2018), "Modeling the interrelationships among barriers to sustainable supply chain management in leather industry", *Journal of Cleaner Production*, Vol. 181, pp. 631-651.

Muduli, K., Govindan, K., Barve, A. and Geng, Y. (2013), "Barriers to green supply chain management in Indian mining industries: a graph theoretic approach", *Journal of Cleaner Production*, Vol. 47, pp. 335-344.

Pazirandeh, A. and Jafari, H. (2013), "Making sense of green logistics", *International Journal of Productivity and Performance Management*, Vol. 62 No. 8, pp. 889-904.

Perotti, S., Zorzini, M., Cagno, E. and Micheli, G.J. (2012), "Green supply chain practices and company performance: the case of 3PLs in Italy", *International Journal of Physical Distribution and Logistics Management*, Vol. 42 No. 7, pp. 640-672.

Piecyk, M.I. and Björklund, M. (2015), "Logistics service providers and corporate social responsibility: sustainability reporting in the logistics industry", *International Journal of Physical Distribution and Logistics Management*, Vol. 45 No. 5, pp. 459-485.

Qian, C., Wang, S., Liu, X. and Zhang, X. (2019), "Low-carbon initiatives of logistics service providers: the perspective of supply chain integration", *Sustainability*, Vol. 11 No. 12, p. 3233.

Raj, T., Shankar, R., Suhaib, M. and Khan, R.A. (2010), "A graph-theoretic approach to evaluate the intensity of barriers in the implementation of FMSs", *International Journal of Services and Operations Management*, Vol. 7 No. 1, pp. 24-52.

Regattieri, A., Santarelli, G. and Piana, F. (2019), "Packaging logistics", in Zijm, H., Klumpp, M., Regattieri, A. and Heragu, S. (Eds), *Operations, Logistics and Supply Chain Management*, Springer, Cham, pp. 273-303.

Salvia, A.L., Leal Filho, W., Brandli, L.L. and Griebeler, J.S. (2019), "Assessing research trends related to Sustainable DevelopmentGoals: local and global issues", *Journal of Cleaner Production*, Vol. 208, pp. 841-849.

San Ong, T., Magsi, H.B. and Burgess, T.F. (2019), "Organizational culture, environmental management control systems, environmental performance of Pakistani manufacturing industry", *International Journal of Productivity and Performance Management*, Vol. 68 No. 7, pp. 1293-1322.

Seidel, S., Recker, J. and VomBrocke, J. (2013), "Sense making and sustainable practicing: functional affordances of information systems in green transformations", *MIS Quarterly*, pp. 1275-1299.

Singh, R.K. and Agrawal, S. (2018), "Analyzing disposition strategies in reverse supply chains: fuzzy TOPSIS approach", *Management of Environmental Quality: An International Journal*, Vol. 29 No. 3, pp. 427-443.

Singh, R.K. and Kumar, P. (2019), "Measuring the flexibility index for a supply chain using graph theory matrix approach", *Journal of Global Operations and Strategic Sourcing*, Vol. 13 No. 1, pp. 56-69.

Singh, R.K., Rastogi, S. and Aggarwal, M. (2016), "Analyzing the factors for implementation of green supply chain management", *Competitiveness Review*, Vol. 26 No. 3, pp. 246-264.

Singh, R.K., Koul, S. and Kumar, P. (2017), "Analyzing the interaction of factors for flexibility in supply chains", *Journal of Modeling in Management*, Vol. 12 No. 4, pp. 671-689.

Singh, R.K., Luthra, S., Mangla, S.K. and Uniyal, S. (2019), "Applications of information and communication technology for sustainable growth of SMEs in India food industry", *Resources, Conservation and Recycling*, Vol. 147, pp. 10-18.

Sindwani, R. and Goel, M. (2015), "Technology-based self-service banking service quality evaluation: a Graph Theoretic Approach", *International Journal of Advanced Science and Technology*, Vol. 80, pp. 1-18.

Sureeyatanapas, P., Poophiukhok, P. and Pathumnakul, S. (2018), "Green initiatives for logistics service providers: an investigation of antecedent factors and the contributions to corporate goals", *Journal of Cleaner Production*, Vol. 191, pp. 1-14.

Szabo, S., Nicholls, R.J., Neumann, B., Renaud, F.G., Matthews, Z., Sebesvari, Z., AghaKouchak, A., Bales, R., Ruktanonchai, C.W., Kloos, J. and Foufoula-Georgiou, E. (2016), "Making SDGs work for climate change hotspots", *Environment: Science and Policy for Sustainable Development*, Vol. 58 No. 6, pp. 24-33.

Tajbakhsh, A. and Hassini, E. (2015), "Performance measurement of sustainable supply chains: a review and research questions", *International Journal of Productivity and Performance Management*, Vol. 64 No. 6, pp. 744-783.

Tang, S., Wang, W., Yan, H. and Hao, G. (2015), "Low carbon logistics: reducing shipment frequency to cut carbon emissions", *International Journal of Production Economics*, Vol. 164, pp. 339-350.

Tascioglu, A. and Keser, H.Y. (2019), "Solar energy in the logistics sector: assessments on Turkey", *Journal of Business and Social Review in Emerging Economies*, Vol. 5 No. 2, pp. 225-236.

Taticchi, P., Tonelli, F. and Pasqualino, R. (2013), "Performance measurement of sustainable supply chains: a literature review and a research agenda", *International Journal of Productivity and Performance Management*, Vol. 62 No. 8, pp. 782-804.

Teixeira, C.R.B., Assumpcao, A.L., Correa, A.L., Savi, A.F. and Prates, G.A. (2018), "The contribution of green logistics and sustainable purchasing for green supply chain management", *Independent Journal of Management and Production*, Vol. 9 No. 3, pp. 1002-1026.

Thanki, S. and Thakkar, J. (2018), "A quantitative framework for lean and green assessment of supply chain performance", *International Journal of Productivity and Performance Management*, Vol. 67 No. 2, pp. 366-400.

The Sustainable Development Goals Report (2019), available at: https://www.un.org/sustainabledevelopment/sustainable-development-goals/ (accessed 30 December 2019).

Trianni, A., Cagno, E., Neri, A. and Howard, M. (2019), "Measuring industrial sustainability performance: empirical evidence from Italian and German manufacturing small and medium enterprises", *Journal of Cleaner Production*, Vol. 229, pp. 1355-1376.

Wang, X., Persson, G. and Huemer, L. (2016), "Logistics service providers and value creation through collaboration: a case study", *Long Range Planning*, Vol. 49 No. 1, pp. 117-128.

WCED, S.W.S. (1987), "World commission on environment and development", *Our common future*, Vol. 17, pp. 1-91.

Wilding, R., Wagner, B., Gimenez, C. and Tachizawa, E.M. (2012), "Extending sustainability to suppliers: a systematic literature review", *Supply Chain Management: An International Journal*, Vol. 17 No. 5, pp. 531-543.

Winkler, H. and Zinsmeister, L. (2019), "Trends in digitalization of intralogistics and the critical success factors of its implementation", *Brazilian Journal of Operations and Production Management*, Vol. 16 No. 3, pp. 537-549.

Yao, E., Lang, Z., Yang, Y. and Zhang, Y. (2015), "Vehicle routing problem solution considering minimising the fuel consumption", *IET Intelligent Transport Systems*, Vol. 9 No. 5, pp. 523-529.

Zhang, W., Shi, Y., Yang, M., Gu, X., Tang, R. and Pan, X. (2017), "Ecosystem evolution mechanism of manufacturing service system driven by service providers", *International Journal of Production Research*, Vol. 55 No. 12, pp. 3542-3558.

Zhu, Q., Sarkis, J. and Lai, K.H. (2008), "Confirmation of a measurement model for green supply chain management practices implementation", *International Journal of Production Economics*, Vol. 111 No. 2, pp. 261-273.

Further reading

Ansari, Z.N., Kant, R. and Shankar, R. (2019), "Prioritizing the performance outcomes due to adoption of critical success factors of supply chain remanufacturing", *Journal of Cleaner Production*, Vol. 212, pp. 779-799.

Liu, X., Ma, S., Tian, J., Jia, N. and Li, G. (2015), "A system dynamics approach to scenario analysis for urban passenger transport energy consumption and CO2 emissions: a case study of Beijing", *Energy Policy*, Vol. 85, pp. 253-270.

Sarkis, J., Helms, M.M. and Hervani, A.A. (2010), "Reverse logistics and social sustainability", *Corporate Social Responsibility and Environmental Management*, Vol. 17 No. 6, pp. 337-354.

About the authors

Anchal Gupta is doing PhD in Management from Delhi School of Management, Delhi Technological University, Delhi, India. Her interest is in Supply Chain Management, Logistics and Service Quality Management. Anchal Gupta is the corresponding author and can be contacted at: anchal.gupta1503@gmail.com

Rajesh Kumar Singh is Professor in Operations Management area at Management Development Institute, Gurugram, India. His areas of interest include Supply Chain Management, Small Business Management and Quality Management. He has published around 125 research papers in reputed international/national journals and conferences. His papers have appeared in journals such as *Production Planning and Control*, *Journal of Industrial and Production Engg*, *Industrial management and Data Systems*, *Singapore Management Review*, *International Journal of Productivity and Performance Management*, *Benchmarking: An International Journal*, *Journal of Modelling in Management*, *Competitiveness Review*, *Journal of Advances in Management Research*, *International Journals of Productivity and Quality Management* etc.

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Disease Detection in Coffee Plants Using Convolutional Neural Network

Manoj Kumar
Computer Engineering Dept.
Delhi Technological University
New Delhi, India
Email: mkumarg@dce.ac.in

Pranav Gupta
Computer Engineering Dept.
Delhi Technological University
New Delhi, India
Email: 7pranavgupta@gmail.com

Puneet Madhav
Computer Engineering Dept.
Delhi Technological University
New Delhi, India
Email: puneet.m98@gmail.com

Sachin
Computer Engineering Dept.
Delhi Technological University
New Delhi, India
Email: sachinrathore454@gmail.com

Abstract— Every year, the crop output of India is significantly affected due to delayed detection of diseases in crops. This research is a contribution to the farmers in their battle against coffee plants diseases. It will help in timely detection of diseases, resulting in increased coffee production output of India. Many coffee plant diseases like Leaf Rust, Cercospora Spots have clear visual symptoms and thus can be extracted and their classification can be done. Convolutional Neural Networks (CNNs) has proved its efficiency and accuracy in the field of image classification and pattern recognition. Hence it can act as a powerful tool in the diagnosis of coffee leaves diseases since these symptoms have clearly distinguishable patterns. Thus, a Convolution Neural Network model is proposed which utilizes the technique of Transfer Learning, reducing the training time taken by the model significantly. Further, to achieve a higher success rate, Data Augmentation technique is applied to enlarge the dataset used to train the network. The proposed model has achieved a high accuracy of 97.61%.

Keywords— Convolutional Neural Network, Coffee Plant Disease, Data Augmentation, Transfer Learning.

I. INTRODUCTION

Coffee is one of the most popular beverages and an important commercial product which is cultivated in more than 50 countries across the globe. Every day an estimated 2.25 billion coffee cups are consumed throughout the world.[1] India is among the top producers of coffee globally and has the finest quality of coffee grown in the shade rather than direct sunlight, anywhere in the world[2]. With an estimated 319,500 metric ton of coffee produced in the year 2018-19[3], of which almost 80% is exported throughout the globe. The amount of revenue generated from the coffee industry is projected to be at USD 987m in the year 2020 and its market is projected to increase annually at 7.4%. [4]

Coffee plants are highly vulnerable to various pests and diseases. Crop loss due to such diseases in coffee plants is estimated at 20-30% of the country's total production[5]. Chemical control's success on any disease depends on the disease pathogenesis stage at which the fungicide is applied [1]. Inexperienced and careless usage of pesticide may result in increased resistance of the pathogens in the long run, severely affecting the capability of plants to fight. Thus, the detection of plant diseases in an accurate and timely manner is an essential aspect of precision agriculture [2].

Soft computing can play a crucial role in the accurate and timely detection of such biotic stresses that attack the leaf of a coffee plant. Various techniques such as genetic algorithms [3], machine learning, use of computer vision and image processing [4], expert systems have been an area of active research to detect such diseases [5]. These methods have not been widely adopted due to low efficiency and success rate. In [6], the authors have used both image processing and artificial neural network to achieve a success rate of 90%. Some plant diseases do not have any conspicuous visual symptoms, however most common diseases found in coffee leaves have clear visual symptoms as shown in the Fig. 1. For example, the symptoms of coffee leaf rust (CLR) disease are oily spots on the upper surface of coffee leaf which are minute and yellowish in color, and expands further into a large circular spot that with time, turn to bright orange to red and at last, to brown with a yellow border[6].

Deep Learning have completely dominated the field of image classification for the last few years, thus the proposed Convolution Neural Network is used to detect coffee leaf diseases and categorize them into 5 classes – Healthy, Diseased Leaves with Cercospora spots, Phoma, Coffee Leaf Rust(CLR), and Leaf Miner with a high success rate.

[1] https://en.wikipedia.org/wiki/Economics_of_coffee
[2] https://en.wikipedia.org/wiki/Coffee_production_in_India
[3] https://www.indiacoffee.org/coffee-statistics.html
[4] https://www.statista.com/outlook/30010000/119/coffee/india#market-globalrevenue
[5] https://economictimes.indiatimes.com/news/economy/agriculture/india-to-get-594349-to-control-leaf-rust-in-coffee/articleshow/2466942.cms
[6] https://www.britannica.com/science/coffee-rust

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:48:41 UTC from IEEE Xplore. Restrictions apply.

The subsequent sections in the paper are organized in the following order: Section II describes the dataset used for training the model in detail. Section III explains the architectural components like Inception V3 model, Data augmentation technique and working of the proposed model. Section IV describes the various tests and experiments performed. In section V, the result is presented. Section VI contains a conclusion along with the future work.

II. DATASET

The pre-processed dataset [7] contains a wide variety of images of coffee leaves for disease detection. The pre-processing of data has been done to eliminate unwanted details and enhance the symptomatic features.

The images contain both disease-free and diseased leaves, affected by different types of biotic stresses. Dataset is divided in two parts – leaf dataset consisting of original images of the leaves and other one containing cropped images of symptoms. Fig. 1 shows the division of the dataset- disease categories with the corresponding number of leaf images.

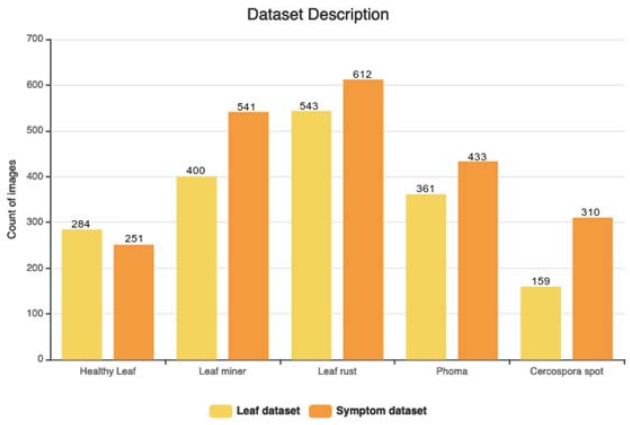


Fig. 1. Dataset Description

Details of each dataset is described below:

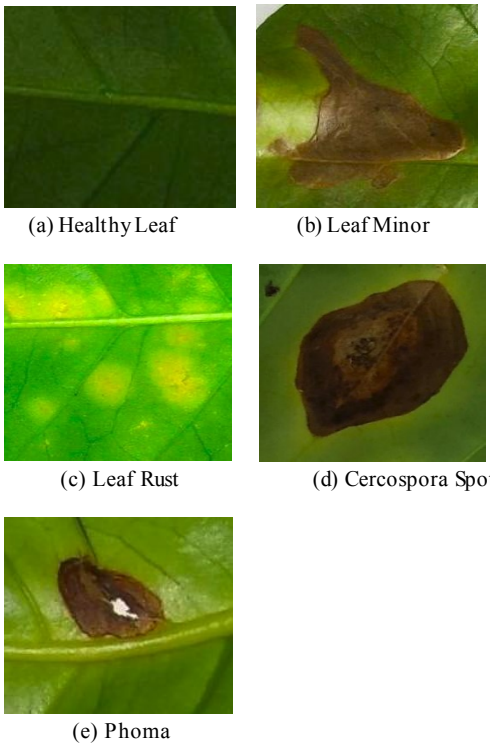
Leaf dataset: It contains images of coffee leaves and a CSV file with labels as different biotic stresses. All original images have the dimensions of 2048x1024 pixels.



Fig. 2. Sample of coffee leaves in Leaf Dataset

It consists of a total of 1747 images divided into 5 different classes - Healthy leaves, Infected leaves with Phoma, Leaf Miner, Coffee Leaf Rust and Cercospora Spots. Fig. 2 shows a sample of images found in this dataset.

Symptom dataset: A total of 2147 cropped images of the symptomatic parts of leaves are present. It is divided as per the named labelled classes - healthy leaves, leaf miner, leaf rust, phoma and cercospora. These are the 5 categories into which have classified our coffee leaves. Fig. 3 shows a sample of images available in this dataset. It is further divided into training, validation and test folders.



(a) Healthy Leaf (b) Leaf Minor

(c) Leaf Rust (d) Cercospora Spot

(e) Phoma

Fig. 3. Sample of cropped images showing only symptomatic leaf parts from Symptom Dataset

III. NETWORK ARCHITECTURE AND WORKING

The proposed network uses a Sequential model for classification. This Sequential model has been implemented using the Keras library in python language. The model has approximately 21 million parameters. Mini-Batch Gradient Descent (MBGD) is the optimization algorithm used to update the parameters. It performs much faster compared to conventional gradient descent algorithms. Fig. 4 represents the proposed neural network architecture. The images provided for training are input into the model in batches. The batch size value is kept as 32 with epoch value as 20. 1488 images belonging to 5 different classes is used for training the model and the image size has dimensions of 299x299 pixels.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:48:41 UTC from IEEE Xplore. Restrictions apply.

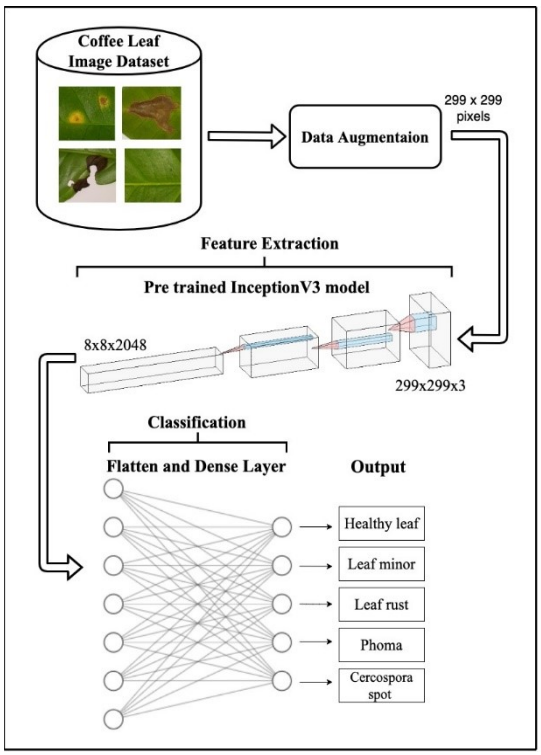


Fig. 4. Network Architecture

In the following sub-sections, explained the techniques of data augmentation and transfer learning and how it has enabled the proposed model to provide highly accurate testing results.

A. Data Augmentation

Data Augmentation is used to enhance the generalizability of the model. It applies augmentation techniques to produce a set of new images that are modified versions of our input data by application of image transformations, contrast changes, blurring images without changing their labels. It creates a varied and enlarged dataset and so the model is able to learn more features in a robust manner [8] and also prevents the problem of overfitting.

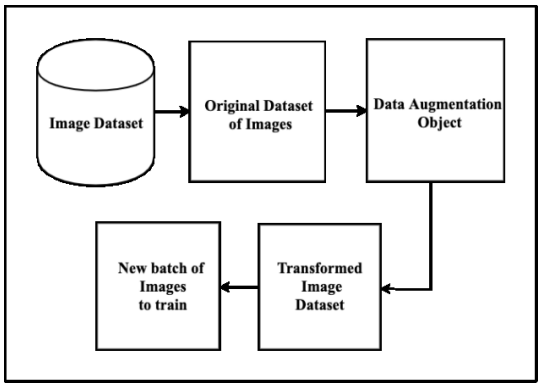


Fig. 5. Data Augmentation on the batch of images

Even though the technique enlarges the dataset but still cannot replicate the practical diversity found in nature. This is why only the image of symptomatic parts of leaves are used rather than considering the entire leaf [9]. In the proposed model and have used the in-place data augmentation technique using the Keras library in python. Implementation is called in-place because the transformation is applied on runtime i.e. it ensures that at each and every epoch new transformed image is input to the network which ultimately ensures high accuracy.

B. Transfer Learning

Transfer learning is a methodology that uses a pre-trained neural network. It makes use of knowledge gained while training on one dataset to be used on other datasets. It uses parametric values like weights and biases obtained while working on a large dataset and uses this knowledge to train on the new dataset. Image classification involves two processes, feature extraction and classification. With the help of transfer learning, one can reuse the part of the model involved in the feature extraction process and train the model for the classification part on a new dataset.

The model that used here for transfer learning is Inception v3 model. It is a 48-layer deep neural network. This model has been trained over a million images from ImageNet Database belonging to around 1000 different classes and have used TensorFlow hub library to perform transfer learning in our model and further added a flatten and a dense layer to our network. It changes the dimensionality of data and prevents the problem of overfitting of data. It has been used in the classification part.

C. Optimization Algorithm

Optimization algorithms help in the minimization (or maximization) of a given mathematical function F(x) which depends on the values of model's parameters and is used in carrying out computations for the output values. One such algorithm is Gradient Descent which falls under the umbrella of first-order optimization algorithms.

$$\theta = \theta - \alpha \nabla_\theta C(\theta; x_i, y_i) \qquad (1)$$

where θ is optimization parameter, C (θ; xi, yi) is the cost, (xi, yi) represents the data-point from the training set.

Mini-Batch Gradient Descent algorithm is used and the dataset is divided into batches of 32 images each. It combines the advantages of both Stochastic Gradient Descent and Batch Gradient Descent algorithms. The disadvantage of BGD is that it runs over the whole dataset to take a single step in the direction of minima. In contrast, SGD computes gradient for each randomly chosen datapoint which is faster but loses the advantages of vectorization and is error-prone also.

D. Categorical Cross-Entropy

Loss functions are used to evaluate the performance in the model. If our model predicts data that is highly deviated from the actual results, then loss function scales up to a large value.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:48:41 UTC from IEEE Xplore. Restrictions apply.

Thus, better predictions mean minimum value of loss function. There are many loss functions like Mean squared Error, Cosine Similarity, Binary Cross-entropy etc. One such loss function is Categorical Cross-Entropy. It is used for multi-class classification. It compares the distribution of our predictions with the actual labels. The probability of the true class is kept as 1 and for all other classes, the probability is kept 0. The equation is given as:

$$L(y, \hat{y}) = -\sum_{j=0}^{m} \sum_{i=0}^{n} (y_{ij} * log(\hat{y}_{ij})) \qquad (2)$$

where $\hat{y}$ is the predicted value and y representing the actual value, m and n represent the total number of data points and the total number of classes respectively.

IV. EXPERIMENTS

To implement this model, divided the dataset for training, validation and testing. 70% of images are used for training the model, 15% each for validation and testing. To create a varied dataset, data augmentation is implemented. The new output images are of dimensions 299x299 pixels. To decrease the training time and increase the performance, used InceptionV3 structure which is a pre-trained network. The convolutional layers that are present in it have pre-trained weights and biases. After each epoch, a new set of images are produced that are passed to the inception model. It forms the feature extraction part of the network.

Along with it, a dense and flatten layer are added to our sequential model. These layers serve in the classification part of the model. The flattening layer is used to convert a multidimensional output from a convolution layer to a linear output such that it can be directly fed into the neural network layer or to regular dense layer. Then, the dense layer performs the calculations and applies Softmax activation function.

Categorical cross-entropy is the loss function used and also tested the model for different values of learning rate as presented in Table I.

TABLE I. ACCURACY AND LOSS VALUE VARIATION WITH DIFFERENT LEARNING RATES

Learning Rate	Accuracy (%)	Loss Value
0.0001	95.82	0.48
0.001	95.22	0.42
0.005	97.61	0.35
0.01	97.31	0.36
0.1	74.03	1.73
1	32.01	10.1

Accuracy of the model increased with increase in learning rate at first, reaching a maximum at 0.005 with an accuracy of 97.61%, and started to decrease further. Hence, final value of the learning rate in the model is 0.005. Mini-batch gradient descent is used as the optimizer algorithm. Small batch size results in reduced accuracy [10]. Thus, batch size of 32 is used for maximum accuracy as shown in Table II.

TABLE II. BATCH SIZE VS ACCURACY

Batch Size	Accuracy (%)	Loss Value
4	94.03	0.49
8	95.22	0.49
16	97.31	0.36
32	97.61	0.35
64	97.01	0.37

V. RESULT

Using data-augmentation, transfer learning and MBGD as an optimizer, the proposed convolution neural network, has achieved an overall testing accuracy of 97.61% with a loss value of 0.35.

Table III represents how accuracy and loss improve with successive iterations. With every epoch, the accuracy increases and loss decrease significantly. After iteration 10, there is no significant improvement in the accuracy of our model although loss value continues to decrease slightly.

TABLE III. VARIATION IN ACCURACY AND LOSS VALUE WITH EPOCH

Epoch Number	Accuracy (%)	Loss Value
1	61.52	1.5752
2	73.43	0.7284
5	94.93	0.3632
8	97.01	0.3560
10	97.61	0.3535
12	97.61	0.3532

Further, Fig. 6 also clearly show that the value of loss and accuracy with varying epochs becomes nearly constant as more iterations are made.

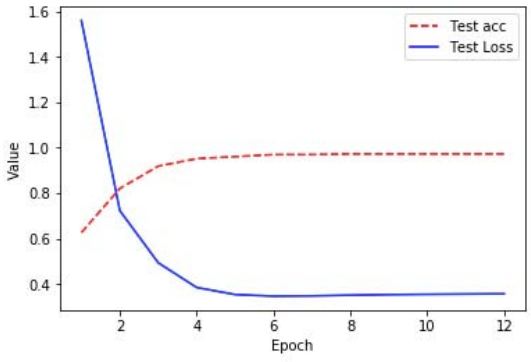


Fig. 6. Graph representing variation in accuracy and loss value with epoch

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:48:41 UTC from IEEE Xplore. Restrictions apply.

The coffee leaf images dataset are not used directly but rather used Data Augmentation technique to enhance the variety of training data, resulting into a much more reliable and accurate model and also avoids the problem of overfitting. The Transfer Learning method is applied which makes the training process faster and efficient. Moreover, flatten and dense layers are added to the network which helps in the classification process. Fig. 8 shows the confusion matrix based on the predicted results.

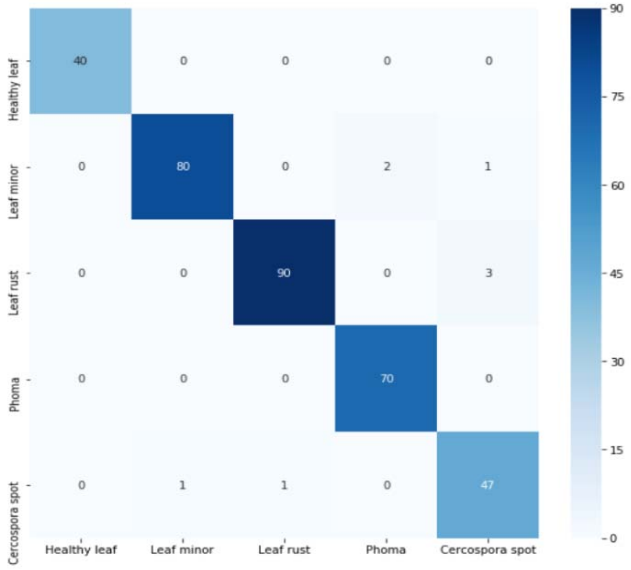


Fig. 8. Confusion matrix for proposed model

The average sensitivity and specificity computed for the proposed model is 0.974 and 0.992 respectively.

Table IV shows a comparison between various other techniques.

TABLE IV. COMPARISON OF DIFFERENT TECHNIQUES

S. No.	Technique Used	Accuracy (%)	Loss Value
1	Fuzzy Logic Based Expert System (using Decision Tree) [5]	85%	Not Available
2	Radial Basis Function Neural Network [11]	95%	Not Available
3	CNN with Data augmentation [12]	95%	0.10
4	CNN with data augmentation and transfer learning (using Inception V3 architecture) *	97.61%	0.35

*proposed model

The values of different hyper-parameters used in the proposed neural network are in Table V.

TABLE V. VALUES OF VARIOUS HYPER-PARAMETERS

Hyper-parameter	Value
Activation function	Softmax
Loss function	Categorical cross-entropy
Momentum	0.9
Batch size	32
Learning rate	0.005
Optimization function	Mini-batch gradient descent

VI. CONCLUSION AND FUTURE WORK

This research paper presents the use of deep convolutional neural networks in the detection and classification of healthy and unhealthy coffee plants. The unhealthy plant leaves are classified into 4 classes – Coffee Leaf Rust, Cercospora Spots, Leaf Miner, Phoma which are commonly occurring biotic stresses in coffee plants. One of the main contributions in our work is to propose a neural network for coffee plant disease detection with a very high level of accuracy.

In summary -

1. Identification and classification of coffee plant diseases by using the proposed neural network achieve a record-breaking accuracy of 97.61%.
2. With the help of this methodology, coffee plants farmer could be helped in timely identification of the diseases which will help in increasing the coffee production output of India.

As future work, aim to explore the requirements and necessary changes to our neural network to be able to accommodate various other new diseases in coffee plants and also account for variable environmental factors.

To make an even better prediction and achieve higher accuracy for a wide variety of diseases in coffee plants, additional variables such as temperature, air moisture, time of plantation and location weather are required.

REFERENCES

[1] Ricardo Balardin. 2010. *Factors affecting fungicide efficacy in the tropics, fungicides*, odile carisse, intechopen, doi: 10.5772/13583.

[2] S. A. Miller, F. D. Beed, and C. L. Harmon, "Plant disease diagnostic capabilities and networks," Annual Review of Phytopathology, vol. 47, pp. 15–38, 2009.

[3] P. Marcos, n. L. Silva rodovalho and a. R. Backes, "*coffee leaf rust detection using genetic algorithm,*" 2019 xv workshop de visão computacional (wvc), são bernardo do campo, brazil, 2019, pp. 16-20.

[4] L. Han, m. S. Haleem and m. Taylor, "*a novel computer vision-based approach to automatic detection and severity assessment of crop diseases,*" 2015 science and information conference (sai), london, 2015, pp. 638-644.

[5] Mettleq, A. S. A., & Abu-Naser, S. S. (2019). *A Rule Based System for the Diagnosis of Coffee Diseases.* International

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:48:41 UTC from IEEE Xplore. Restrictions apply.

Journal of Academic Information Systems Research (IJAISR), 3(3), 1-8.

[6] Suhartono, Derwin & Aditya, Wahyu & Lestari, Miranty & Yasin, Muhammad. (2013). *Expert System in Detecting Coffee Plant Diseases*. International Journal of Electrical Energy. 156-162. 10.12720/ijoee.1.3.156-162

[7] Krohling, Renato A.; Esgario , Guilherme J. M.; Ventura, José A. (2019), "BRACOL - A Brazilian Arabica Coffee Leaf images dataset to identification and quantification of coffee diseases and pests", Mendeley Data, v1. Available from: http://dx.doi.org/10.17632/yy2k5y8mxg.1

[8] . j. Shijie, w. Ping, j. Peiyi and h. Siping, "*research on data augmentation for image classification based on convolution neural networks,*" 2017 chinese automation congress (cac), jinan, 2017, pp. 4165-4170.

[9] Arnal Barbedo, J. G. (2019). *Plant disease identification from individual lesions and spots using deep learning*. Biosystems Engineering, 180, 96–107.

[10] Durmus, H., Gunes, E. O., & Kirci, M. (2017). *Disease detection on the leaves of the tomato plants by using deep learning*. 2017 6th International Conference on Agro-Geoinformatics.

[11] Rath, A. K., & Meher, J. K. (2020). *Disease detection in infected plant leaf by computational method*. Archives of Phytopathology and Plant Protection, 1–11.

[12] Marcos, a. P., rodovalho, n. L. S., & backes, a. R. (2019). *Coffee leaf rust detection using convolutional neural network*. 2019. Xv workshop de visão computacional (wvc). Doi:10.1109/wvc.2019.8876931.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:48:41 UTC from IEEE Xplore. Restrictions apply.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/342678861

Dual-Modal Emotion Recognition Using Discriminant Correlation Analysis

Conference Paper · July 2020

CITATIONS
0

READS
11

5 authors, including:



Utkarsh Bisaria
Delhi Technological University
1 PUBLICATION **0** CITATIONS
SEE PROFILE



Suhail Saifi
Delhi Technological University
1 PUBLICATION **0** CITATIONS
SEE PROFILE

All content following this page was uploaded by Suhail Saifi on 09 July 2020.

The user has requested enhancement of the downloaded file.

Proceedings of the International conference on Electronics and Sustainable Communication Systems (ICESCS 2020)
DVD Part Number: CFP20V66-DVD; ISBN: 978-1-7281-4107-7

Dual-Modal Emotion Recognition Using Discriminant Correlation Analysis

Dr. Rajiv Kapoor
Dept. of Electronics and Communication
Delhi Technological University
New Delhi. India
rajivkapoor@dce.ac.in

Suhail Saifi
Dept. of Electronics and Communication
Delhi Technological University
New Delhi, India
suhail_bt2k16@dtu.ac.in

Tarun Kapoor
Dept. of Electronics and Communication
Delhi Technological University
New Delhi, India
tarun_bt2k16@dtu.ac.in

Utkarsh Bisaria
Dept. of Electronics and Communication
Delhi Technological University
New Delhi, India
2utkarsh_bt2k16@dtu.ac.in

Nikhil Singh
Dept. of Electronics and Communication
Delhi Technological University
New Delhi, India
nikhil.gbpec2016@gmail.com

***Abstract*— A dual-modal framework is presented for human emotion recognition based on Discriminant Correlation Analysis. For visual features, key-frames are selected from the video sequences and features are extracted using a Residual Neural Network. The Tuning Network (T-Net) is used for fine-tuning of these visual features. From the audio channel, Intensity, MFCC, pitch, F0 envelope, Line Spectral Frequency (LSF), and Zero Crossing Rate (ZCR) is evaluated along with their delta regression coefficients. Discriminant Correlation Analysis based feature fusion is used to fuse audio and visual modalities.**

Keywords*— Dual-Modal Framework, Discriminant Correlation Analysis, T-Net, Residual Neural Network, Feature Fusion

I. INTRODUCTION

One of the key tasks for Human-Computer interaction is detecting and recognizing human emotions. Humans express emotions using different modalities, such as facial expressions, hand movements, speech, body posture, etc. Out of all these modalities, facial expression and speech are studied the most for human emotion recognition. For recognizing facial expression a lot of different challenges are faced due to head pose variations, illumination changes, 2D and 3D rotations, etc. For speech emotion recognition, major challenges include the extraction of significant features and finding appropriate classifiers. To improve the emotion recognition process, both audio and visual information extracted from the video clips are used . For detecting and recognizing facial expressions first, the face is detected in each frame using a Multi-Task CNN [7]. 2-D rotation is eliminated by aligning the face using facial landmarks. After face detection and alignment, key-frames are extracted from the frame sequence. The key-frames remove the redundant information inside video frames and make the process computationally less expensive by reducing the number of frames for processing.

For visual feature extraction, Xiaobai Li [1] used LBP histograms on three planes (LBP-TOP). To solve the problem of the variable number of key-frames, temporal interpolation was used, but it is computationally more expensive as compared to our approach. Junkai Chen [2] used Histograms of Oriented Gradients over three orthogonal planes (HOG-TOP) and concluded that compared with LBP-TOP, HOG-TOP is more robust and efficient. LBP-TOP and HOG-TOP, although maintains the temporal information, does not provide the extent of quality seen in features extracted using deep neural networks. In our framework, visual features are extracted from the key-frames using a Deep Residual Neural Network. The extracted features are fine-tuned using another Neural Network to reduce dimensionality and improve accuracy.

S. Sahoo [3] also proposed an audio-visual based emotion recognition system in which they used MFCC as audio and LBP as a visual feature and applied decision level fusion. The classification was done using HMM (Hidden Markov Model) and SVM classifier for audio and visual features respectively. For the audio part, the following features is extracted along with their delta regression coefficients: Intensity, MFCC, pitch, F0 envelope, Line Spectral Frequency (LSF), and Zero Crossing Rate (ZCR). Some statistical functions are applied to these features to extract a large and relevant amount of features from audio modality.

Junkai Chen [2] compared the accuracy of feature and decision level fusion and obtained better results with the former technique. In our framework, after extracting features from both modalities, Discriminant Correlation Analysis (DCA) [26] based feature level fusion is performed, which gives us highly information-rich fused data capable of representing both modalities together.

The final prediction is made on the feature vector obtained after the fusion of audio-visual modalities. The method was tested on SAVEE [4], CREMA-D [5], and eNTERFACE'05 [6] dataset and results were compared for audio, visual and fused audio-visual features.

II. FRAMEWORK

A. Visual Feature Extraction

1) Face Tracking and Face Alignment

One very important part of the emotion detection and recognition process is face detection and alignment. Significant challenges are faced due to 2D-3D rotation, different pose, lighting conditions, and different camera angles. Recently CNN's have proved to be very useful in solving a lot of computer vision tasks. A three-stage multi-task deep cascaded convolution network [7] is used for face detection and face alignment. In stage one, for the generation of candidate windows, fast Proposal Network is used. A refinement network (R-Net) is used in the second stage and the last stage The Output Network (O-Net) is used to produce the facial-key landmark position.

This 3-stage cascaded network has proved to be more accurate and less computationally expensive over the state-of-the-art techniques. The facial-key landmarks produced by the Output Network are used for face detection and alignment.

2) Key Frame Extraction

Rapid facial movements [8] between frames is analysed by studying the difference observed in LBP features.

a) Feature Difference analysis (FDA)

FDA is used to compare the difference in LBP features in sequential video frames within an interval.

The currently observed frame is referred to as the Target Frame (TF) when an interval of H frames is analysed. Start Frame (SF) and End Frame (EF) are the κth frames before and after TF respectively. Where κ is 1/2 × (*H*-1). The average of SF and EF is referred to as the Mean Feature Frame (MFF).

LBP histograms from each TF are compared to its MFF using the Chi-Squared distance. Any rapid facial movement results in high dissimilarity.

b) Threshold and peak estimation

The face is divided into 36 blocks and FDA is done for every block. Average of α largest FD values are used to describe each frame.

$$L = \frac{1}{\alpha}\sum_{j=1}^{\alpha}(value_j) \tag{1}$$

Contrasting is required to avoid errors due to local peak variations and noise:

$$L_{new\,i} = L_i - \frac{1}{2}(L_{i+k} + L_{i-k}) \tag{2}$$

Negative values are set to 0 as they represent that there are no sudden movements. Finally, the threshold is computed as:

$$th = L_{new\,mean} + \tau \times (L_{new\,max} - L_{new\,mean}) \tag{3}$$

Where τ is a percentage parameter and lies between 0 and 1. Subsequent peaks are separated by a distance of κ/2.

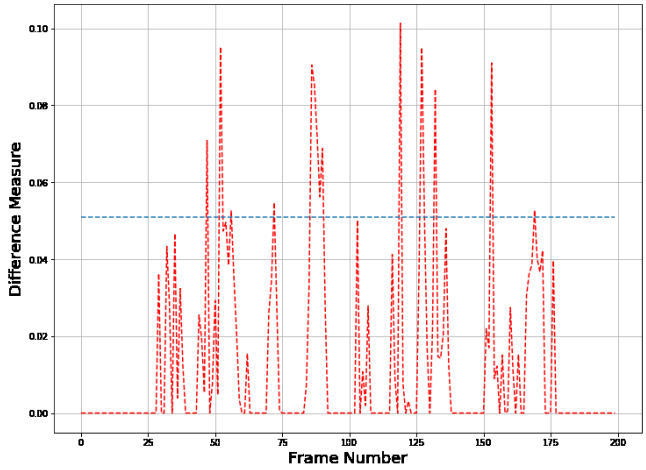


Fig.1. Plot showing FD vector vs. frame number

To obtain a constant number of key-frames from different videos, a new vector F is calculated as:

$$F_i = L_{new\,i} - th \tag{4}$$

λ frames with the highest corresponding values in F are selected as key-frames.

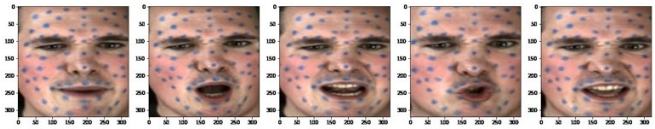


Fig.2. Key-frames extracted from a video of person JK in SAVEE expressing Anger

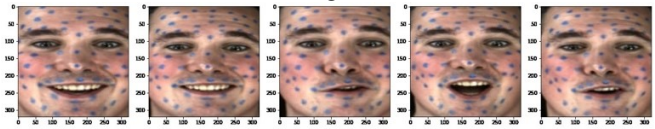


Fig.3. Key-frames extracted from a video of person JK in SAVEE expressing Happiness

3) ResNet-50

ResNet-50 is used [9], which is a 50 layer Residual Neural Network on each key-frame for feature extraction. The network can be trained on the emotion recognition datasets but better results are obtained using transfer learning on ImageNet [10] weight and bias.

Frames are fed to the Residual Network and feature vectors with 2048 dimensions are obtained. This method is repeated for each key-frame in a video sequence. For example, if there are λ key-frames, then the final feature vector with shape (λ×2048) is obtained for each video sequence.

4) Fine Tuning Video Features, T-Net

ResNet-50 operates on individual frames extracted from a video sequence, to represent each video with a single feature

vector, different approaches can be taken. The simplest way is to take mean overall the key-frames to obtain the average-feature set. This method is the least expensive in terms of computational power but the temporal information is not retained because the order of occurrence of frames inside the video sequence is not taken into consideration.

Temporal-information can be preserved using a feature descriptor that works in both the spatial and temporal domains. For example, Xiaobai Li [1] used LBP over three orthogonal planes (LBP-TOP) proposed by Zhao and Pietkäinen [11] as a feature descriptor, but according to previous studies, traditional LBP-TOP feature which uses all three histogram XY, XT, YT does not always yield the best performance. Xiaobai Li has used five different combinations of LBP histograms on three planes to increase performance. These methods, although maintains the temporal information does not provide the extent of quality seen in features extracted using deep neural networks like ResNet-50.

Taking mean overall key-frames not only results in loss of temporal information but also reduces the training data by a factor of λ, where λ is the number of key-frames per video. To solve this problem a method is proposed where instead of taking the mean overall key-frames, each key-frame is treated as an individual sample for training by assigning it the label of the video it is extracted from. This feature vector is referred to as the extended feature vector. The average and extended feature vector have same dimensions and only differ in the number of samples, in extended feature vector one video are split into λ samples and in the average feature vector, one video is represented by only one sample by taking mean overall the key-frames. A neural network is trained using the extended feature vector. This network is referred to as the tuning-network (T-NET). The output of the last dense layer of T-Net on the average feature set is used as the final visual feature vector.

B. Audio Feature Extraction

Audio have been used as our second modality for analysis. Human beings make use of both their facial expressions as well as their speech to express their emotions. Hence audio analysis is also very important which can give us some really important features to judge the emotion of a human being. For audio analysis, multiple features have been extracted that are quite relevant to express human emotions.

The feature set extracted consists of Intensity, MFCC, pitch, F0 envelope, Line Spectral Frequency (LSF), and Zero Crossing Rate (ZCR). It has been seen that by using time derivatives of these parameters, which are also called delta regression coefficients, the overall performance of the system can be improved hence delta regression coefficients have also been added for each of the above-mentioned parameters in our features set.

But there is a problem of variable feature-length of each audio sample depending on the duration of the audio clip (total number of frames will be different per audio clip). Thus, the coefficients of these features cannot be considered as it is, so to tackle this, some statistical functions have been applied on each of the parameters such that a fixed-length feature vector per audio clip can be obtained. The following statistical functions are applied: arithmetic mean, max/min value, range, linear regression coefficients, standard deviation, linear and quadratic error, kurtosis, skewness, and inter-quartile ranges.

Let us see what are the above-mentioned features mathematically are and how they are relevant features for detecting human emotions:

1) Pitch

It is the perceptual feature for a human listening to a particular sound. So in the emotion detection task, it is one of the important features to be determined. There are various time domain and frequency domain algorithms to determine pitch, also known as algorithms to find fundamental frequency(F0). Mathematically, the pitch for speech signal s is given as:

$$P(s) = D\{log|D(s.{w_n}^H\|s\|)|\} \quad (5)$$

Where D stands for DFT function, $\|s\|$ for the length of the signal and ${w_n}^H$ for hamming Window.

2) Intensity

Physically, the Intensity of a sound is defined as the amount of energy carried by an audio signal per unit time through a unit area. Mathematically, Intensity for a given sound signal s can be calculated as:

$$I_i(s) = \frac{\sum_{n=1}^{N}(s_{i+n}).{w_n}^H}{\sum_{n=1}^{N}{w_n}^H} \quad (6)$$

3) MFCC (Mel-frequency Cepstral Coefficient)

Mel-frequency Cepstral Coefficient (MFCC) is one of the most popular feature extractors in the field of audio analysis. It is the state-of-the-art audio feature extractor till now. The basic ideology behind this method is just to mimic the behaviour of human ears.

MFCC is calculated by a series of steps which have been explained in [12] very clearly. Finally, MFC coefficients are calculated by taking Discrete Cosine Transform (DCT) to decorrelate the filter bank coefficients, which is given by:

$$M_i = \sum_{k=1}^{N}\cos[i(k-1)\frac{\pi}{N}], \qquad i = 1,2,3 \dots P \quad (7)$$

Where P represents the number of cepstral coefficients $X_k, k = 1,2, \dots N$ is the log energy output of the k^{th} filter and N represent the number of triangular filter bank filters.

4) ZCR (Zero Crossing Rate)

The Zero-Crossing Rate (ZCR) of an audio frame is determined as the rate of change of sign of the signal in one frame duration.

Mathematically, ZCR for a frame $(Z(i))$ of length l can be calculated as:

$$Z(i) = \frac{1}{2l}\sum_{n=1}^{l}|sgn[s_i(n)] - sgn[s_i(n-1)]| \quad (8)$$

Where,

$$sgn[x_i(n)] = \begin{cases} 1, s_i(n) \geq 0 \\ -1, \ s_i(n) < 0. \end{cases}$$

C. Feature Fusion Technique

After extracting audio and visual features, the Discriminant Correlation Analysis (DCA) have been applied based feature fusion.

DCA is one of the robust fusion techniques. This technique is applied at the feature level, where a rich amount of information is available. In this fusion technique both of the feature, vectors are scaled in such a manner that inter-class correlation can be minimized within feature vectors and further scaled so that pair-wise correlation between two feature vectors can also be maximized. For more understanding, consider the following example:

Assume that η number of data samples have been collected that correspond to m different class labels. Now, let us divide this η columns matrix into m sections. In this matrix $u_{ij} \in U$ is j^{th} sample vector in the i^{th} class, $\overline{u_i}$ represents mean of the u_{ij} vectors belong to i^{th} class and $\bar{u}$ represents mean of the whole feature set i.e., $\overline{u_i} = \frac{1}{\eta_i}\sum_{j=1}^{\eta_i} u_{ij}$ and $\bar{u} = \frac{1}{\eta}\sum_{j=1}^{m} \eta_i \overline{u_i}$. Therefore, the inter-class scatter matrix can be described as :

$$S_{bu_{(p\times p)}} = \sum_{i=1}^{m} \eta_i (\overline{u_i} - \bar{u})(\overline{u_i} - \bar{u})^T = \boldsymbol{\varphi}_{bu}\boldsymbol{\varphi}_{bu}{}^T \quad (9)$$

Where p denotes the total number of features,

$$\boldsymbol{\phi}_{bu_{(p\times m)}} = \left[\sqrt{\eta_1}\,(\overline{u_1} - \bar{u}), \sqrt{\eta_2}\,(\overline{u_2} - \bar{u}), \ldots, \sqrt{\eta_m}\,(\overline{u_m} - \bar{u})\right]$$

So, if $p>>m$, which will be the case most of the time in real-world situations it will always be simple to compute $(\boldsymbol{\varphi}_{bu}{}^T\boldsymbol{\varphi}_{bu})_{(m\times m)}$ as covariance matrix instead of evaluating $(\boldsymbol{\varphi}_{bu}\boldsymbol{\varphi}_{bu}{}^T)_{(p\times p)}$. In [13] it has been shown that the most important eigenvectors of $\boldsymbol{\varphi}_{bu}\boldsymbol{\varphi}_{bu}{}^T$ can be acquired by simply mapping the eigenvectors of $\boldsymbol{\varphi}_{bu}{}^T\boldsymbol{\varphi}_{bu}$. Hence, just need to obtain eigenvectors of the matrix $(\boldsymbol{\varphi}_{bu}{}^T\boldsymbol{\varphi}_{bu})_{(m\times m)}$.

Now, to get well-separated classes, $\boldsymbol{\varphi}_{bu}{}^T\boldsymbol{\varphi}_{bu}$ should be a diagonal matrix. To diagonalize it, transformation can be used , given as:

$$A^T(\boldsymbol{\varphi}_{bu}{}^T\boldsymbol{\varphi}_{bu})A = \hat{\Delta} \quad (10)$$

Where A represents a matrix of orthogonal eigenvectors and $\hat{\Delta}$ represents a diagonal matrix where eigenvalues are sorted in descending order.

Now, consider $B_{(m\times c)}$ a matrix which consists of c eigenvectors obtained from A matrix by selecting c largest non-zero eigenvalues, such that:

$$B^T(\boldsymbol{\varphi}_{bu}{}^T\boldsymbol{\varphi}_{bu})B = \Delta_{(c\times c)} \quad (11)$$

Hence, the corresponding c most important eigenvectors for scatter-matrix in Eq. 7 can be acquired by simply mapping: $B \rightarrow \boldsymbol{\varphi}_{bu}B$:

$$(\boldsymbol{\varphi}_{bu}B)^T S_{bu}(\boldsymbol{\varphi}_{bu}B) = \Delta_{(c\times c)} \quad (12)$$

Therefore, $\theta_{bu} = \boldsymbol{\varphi}_{bu}B\Delta^{-1/2}$, is the weight which diagonalizes S_{bu} and lowers the dimensions of our matrix U from p to c, i.e.,

$$\theta_{bu}{}^T S_{bu}\theta_{bu} = I, \quad (13)$$

$$U'_{(c\times\eta)} = \theta_{bu_{(c\times p)}}{}^T U_{(p\times\eta)} \quad (14)$$

Where, U' is the transformed version of U in which classes are well-separated and inter-class correlation is minimized.

Similarly, another feature setcan be solved for, V and find transforming weight θ_{bv}, which minimize inter-class correlation in V feature vector and lowers the dimensions from q to c.

$$\theta_{bv}{}^T S_{bv}\theta_{bv} = I, \quad (15)$$

$$V'_{(c\times\eta)} = W_{bv_{(c\times q)}}{}^T V_{(q\times\eta)}\,. \quad (16)$$

Till now transformed U to U' and V to V', such that between class correlation has been minimized but now need to further transform U' and V' such that features of U' will have non-zero correlation only with corresponding features in another set, in V'. To accomplish this need to diagonalize another matrix, which is called the between-set covariance matrix, $S'_{uv} = U'V'^T$. For this task, singular value decomposition (SVD) can be used :

$$S'_{uv_{(c\times c)}} = X\,\Sigma Y^T \Rightarrow X^T S'_{xy} Y = \Sigma \quad (17)$$

Where, U' and V' have the rank of order c and $S'_{uv\ (c\times c)}$ is non-degenerate. Thus, Σ is a diagonal matrix in which only diagonal elements are non-zero. So, have $\theta_{du} = X\Sigma^{-1/2}$ and $\theta_{dv} = Y\Sigma^{-1/2}$ such that:

$$\left(X\Sigma^{-\frac{1}{2}}\right)^T S'_{uv}\left(Y\Sigma^{-\frac{1}{2}}\right) = I, \quad (18)$$

Where, S'_{uv} gets unitized. Therefore, transformed finally our feature vectors as:

$$U^* = \theta_{du}{}^T U' = \underbrace{\theta_{du}{}^T\theta_{bu}{}^T}U = \theta_u U \quad (19)$$

$$V^* = \theta_{dv}{}^T V' = \underbrace{\theta_{dv}{}^T\theta_{bv}{}^T}V = \theta_v V \quad (20)$$

Where $\theta_u = \theta_{du}{}^T\theta_{bu}{}^T$ and $\theta_v = \theta_{dv}{}^T\theta_{bv}{}^T$ are the final transformation weights for U and V, respectively.

Hence after transforming our audio feature vector, say U to U' and visual feature vector, say V to V', feature level

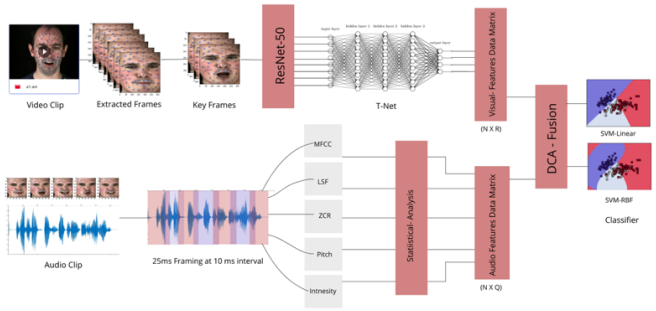


Fig.4. Proposed framework

fusion can be performed by simply concatenating or taking the summation of the transformed feature vectors as shown:

$$Z_1 = \begin{pmatrix} U^* \\ V^* \end{pmatrix} = \begin{pmatrix} W_u{}^T U \\ W_v{}^T V \end{pmatrix} = \begin{pmatrix} W_v & 0 \\ 0 & W_u \end{pmatrix}^T \begin{pmatrix} U \\ V \end{pmatrix} \quad (21)$$

Or

$$Z_2 = U^* + V^* = W_u{}^T U + W_v{}^T V = \begin{pmatrix} W_u \\ W_v \end{pmatrix} \begin{pmatrix} U \\ V \end{pmatrix} \quad (22)$$

However, the concatenation method have been used to fuse audio (U) and visual (V) feature vectors i.e., Z_1.

III. Experiment and Results

In this section, qualitative and quantitative analysis is done in the terms of accuracy with other existing techniques on three datasets i.e. SAVEE, CREMA-D, and eNTERFACE'05 and results are compared for audio, visual and fused audio-visual features.

A. SAVEE

The SAVEE database contains recordings from 4 different actors expressing 7 different emotions. The database has a total of 480 audio-visual clips recorded at 60fps. We merged classes 'Sadness' and 'Disgust' due to their high co-relation and to maintain a constant 6-class classification on each dataset.

For key-frame extraction, the value of H is set to H = 19 (to maintain 0.3 sec). The default value of 0.5 is used for τ. For LBP, radius r is set to r = 3 and neighboring points to $8 \times r = 24$. λ is set to 20. The output of the third dense layer of T-Net on the average feature vector is used as the final visual-feature set for fusion. For audio features, a frame size of 25ms at the rate of 10ms is used. 12 MFCC coefficients, 8 line spectral frequencies along with other features with their respective delta regression coefficients are extracted from each frame. Hence in total $24 \times 2 = 48$ different types of features are calculated and 17 statistical functions are applied on each of these features, therefore in total $48 \times 17 = 816$ types of audio features are calculated per audio clip.

After getting feature data matrices from both the modalities i.e., $X_{(480\times816)}$ from audio and $Y_{(480\times512)}$ from visual, DCA fusion is applied which gives us the fused feature vector of dimension (48×10) for classification. SVM [14] with linear and RBF kernel is used for classification at each stage.

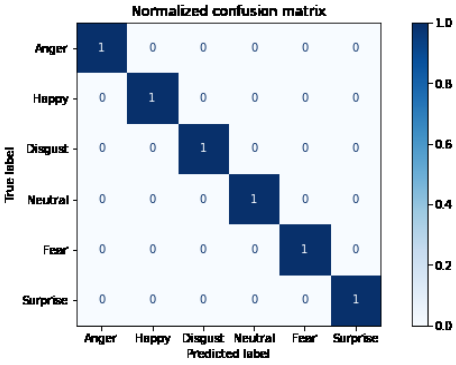


Fig. 5. Confusion matrix for 6 classes evaluated using audio-visual features with SVM classifier on the SAVEE dataset.

TABLE 1
Accuracy on SAVEE dataset

Feature type	Classifier	Accuracy
Visual	SVM (RBF)	99.00%
Audio	SVM (RBF)	90.70%
Audio-Visual	SVM (linear)	100%

This table represents the accuracy of our classification model when only the audio feature vector is used i.e. $X_{(480\times816)}$ which is 90.70%, when only visual feature vector is used i.e. $Y_{(480\times512)}$ which is 99.00% and when a DCA based fused vector is used which is 100%.

TABLE 2
Comparison of SAVEE

Method	Accuracy
ISLA [15]	86.01%
F. Noroozi, SVM [16]	98.10%
F. Cid [17]	96.20%
S. Haq [18]	99.00%
D. Gharavian [19]	96.88%
Ours	**100%**

This table compares the emotion recognition rates from our proposed framework with some other pre-existing frameworks for emotion detection.

B. CREMA-D

CREMA-D contains 7,442 clips from 91 actors. The sentences are presented using 6 different emotions at four different levels. For our application, all levels of emotions are considered as one.

For key-frame extraction, the value of H is set to H = 11 (to maintain 0.3 sec). The default value of 0.5 is used for τ. For LBP, radius r is set to r = 3 and neighboring points to $8 \times r = 24$. λ is set to 20. The output of the third dense layer of T-Net on the average feature vector is used as the final visual-feature set for fusion. For audio features, a frame size

of 25ms at the rate of 10ms is used. 14 MFCC coefficients, 10 line spectral frequencies along with other features with their respective delta regression coefficients are extracted from each frame. Hence in total $28 \times 2 = 56$ different types of features are calculated and 17 statistical functions are applied on each of these features, therefore in total $56 \times 17 = 952$ types of audio features are calculated per audio clip.

After getting feature data matrices from both the modalities i.e., $X_{(7442\times952)}$ from audio and $Y_{(7442\times512)}$ from visual, DCA fusion is applied which gives us the fused feature vector of dimension (7442×10) for classification. SVM with linear and RBF kernel is used for classification at each stage.

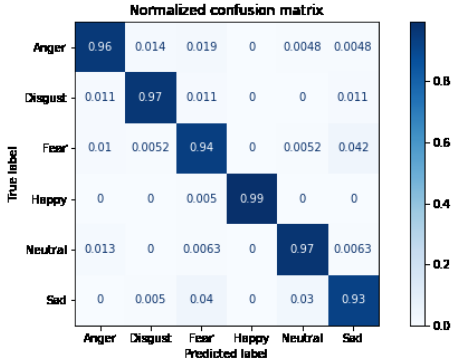


Fig. 6. Confusion matrix for 6 classes evaluated using audio-visual features with SVM classifier on CREMA-D dataset

TABLE 3
Accuracy on CREMA-D dataset

Feature type	Classifier	Accuracy
Visual	SVM (RBF)	83.80%
Audio	SVM (RBF)	62.10%
Audio-Visual	SVM (linear)	95.90%

This table represents the accuracy of our classification model when only the audio feature vector is used i.e. $X_{(7442\times952)}$ which is 62.10%, when only visual feature vector is used i.e. $Y_{(7442\times512)}$ which is 83.80% and when a DCA based fused vector is used which is 95.90%.

TABLE 4
Comparison on the CREMA-D dataset

Method	Accuracy
E. Ghaleb [20]	74.00%
R. Beard [21]	65.00%
E. Ghaleb [22]	66.50%
Ours	**95.90%**

This table compares the emotion recognition rates from our proposed framework with some other pre-existing frameworks for emotion detection.

C. eNTERFACE'05

The eNTERFACE'05 contains audio-visual clips from 46 subjects reacting to six different situations and expressing 6 different emotions. The videos are recorded at a 25fps frame rate.

For key-frame extraction, the value of H is set to $H = 7$ (to maintain 0.3 sec). The default value of 0.5 is used for τ. For LBP, radius r is set to $r = 2$ and neighboring points to $8 \times r = 16$. λ is set to 20. The output of the third dense layer of T-Net network on the average feature vector is used as the final visual-feature set for fusion. For audio features, a frame size of 25ms at the rate of 10ms is used. 12 MFCC coefficients, 10 line spectral frequencies along with other features with their respective delta regression coefficients are extracted from each frame. Hence in total $26 \times 2 = 52$ different types of features are calculated and 17 statistical functions are applied on each of these features, therefore in total $52 \times 17 = 884$ types of audio features are calculated per audio clip.

After getting feature data matrices from both the modalities i.e., $X_{(1166\times884)}$ from audio and $Y_{(1166\times512)}$ from visual, DCA fusion is applied which gives us the fused feature vector of dimension (1166×10) for classification. SVM with linear and RBF kernel is used for classification at each stage.

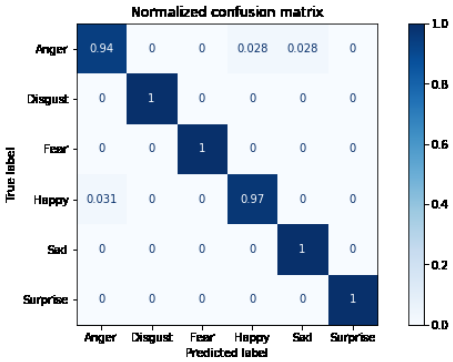


Fig. 7. Confusion matrix for 6 classes evaluated using audio-visual features with SVM classifier on eNTERFACE'05 dataset

TABLE 5
Accuracy on eNTERFACE'05 dataset

Feature type	Classifier	Accuracy
Visual	SVM (RBF)	84.10%
Audio	SVM (linear)	67.30%
Audio-Visual	SVM (RBF)	98.50

This table represents the accuracy of our classification model when only the audio feature vector is used i.e. $X_{(1166\times884)}$ which is 67.3%, when only visual feature vector is used i.e. $Y_{(1166\times512)}$ which is 84.10% and when a DCA based fused vector is used which is 98.50%.

View publication stats

TABLE 6
Comparison on the eNTERFACE'05 dataset

Method	Accuracy
M. ShamimHossain [23]	86.40%
F. Noroozi [16], SVM	94.92%
Subhasmita Sahoo [3]	81.00%
AVEF [24]	86.40%
Mansoorizadeh [25]	84.02%
Ours	**98.50%**

This table compares the emotion recognition rates from our proposed framework with some other pre-existing frameworks for emotion detection.

IV. CONCLUSION

Our proposed framework based on Discriminant Correlation Analysis on three different standard datasets have been successfully applied: SAVEE, CRAMA-D, and eNTERFACE'05. Our results on CREMA-D outperform other previously proposed frameworks in [20, 21, 22] and achieve state-of-the-art results with a recognition rate of 95.90%. On SAVEE and eNTERFACE'05, our results are equivalent to state-of-the-art results which are 100% and 98.5% respectively as compared to the results in [16]. In future, neural networks will be used for audio feature extraction.

V. REFERENCES

[1] X. Li et al., "Towards Reading Hidden Emotions: A Comparative Study of Spontaneous Micro-Expression Spotting and Recognition Methods," in IEEE Transactions on Affective Computing, vol. 9, no. 4, pp. 563-577, 1 Oct.-Dec. 2018.

[2] J. Chen, Z. Chen, Z. Chi and H. Fu, "Facial Expression Recognition in Video with Multiple Feature Fusion," in IEEE Transactions on Affective Computing, vol. 9, no. 1, pp. 38-50, 1 Jan.-March 2018.

[3] S. Sahoo and A. Routray, "Emotion recognition from audio-visual data using rule based decision level fusion," 2016 IEEE Students' Technology Symposium (TechSym), Kharagpur, 2016, pp. 7-12.

[4] P. Jackson and S. Haq, "Surrey audio-visual expressed emotion (savee) database," 2015, http://kahlan.eps.surrey.ac.uk/savee/

[5] H. Cao, D. G. Cooper, M. K. Keutmann, R. C. Gur, A. Nenkova and R. Verma, "CREMA-D: Crowd-Sourced Emotional Multimodal Actors Dataset," in IEEE Transactions on Affective Computing, vol. 5, no. 4, pp. 377-390, 1 Oct.-Dec. 2014.

[6] O. Martin, I. Kotsia, B. Macq and I. Pitas, "The eNTERFACE' 05 Audio-Visual Emotion Database," 22nd International Conference on Data Engineering Workshops (ICDEW'06), Atlanta, GA, USA, 2006, pp. 8-8.

[7] K. Zhang, Z. Zhang, Z. Li and Y. Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," in IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499-1503, Oct. 2016.

[8] A. Moilanen, G. Zhao, and M. Pietikäinen, "Spotting rapid facial movements from videos using appearance-based feature difference analysis," in ICPR. IEEE, 2014, pp. 1722–1727.

[9] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 770-778.

[10] J. Deng, W. Dong, R. Socher, L. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, 2009, pp. 248-255.

[11] G. Zhao and M. Pietikäinen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," PAMI, vol. 29, no. 6, pp. 915–928, 2007.

[12] Koduru, A., Valiveti, H.B. & Budati, A.K. Feature extraction algorithms to improve the speech emotion recognition rate. *Int J Speech Technol* **23,** 45–55 (2020). https://doi.org/10.1007/s10772-020-09672-4

[13] M. Turk and A. Pentland, "Eigenfaces for recognition," Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71–86, 1991.

[14] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt and B. Scholkopf, "Support vector machines," in IEEE Intelligent Systems and their Applications, vol. 13, no. 4, pp. 18-28, July-Aug. 1998.

[15] Y. Kim and E. M. Provost, "ISLA: Temporal Segmentation and Labeling for Audio-Visual Emotion Recognition," in IEEE Transactions on Affective Computing, vol. 10, no. 2, pp. 196-208, 1 April-June 2019.

[16] F. Noroozi, M. Marjanovic, A. Njegus, S. Escalera and G. Anbarjafari, "Audio-Visual Emotion Recognition in Video Clips," in *IEEE Transactions on Affective Computing*, vol. 10, no. 1, pp. 60-75, 1 Jan.-March 2019.

[17] F. Cid, L. J. Manso, and P. Nunez, "A novel multimodal emotion recognition approach for affective human robot interaction," Proc. FinE, pp. 1–9, 2015.

[18] S. Haq, T. Jan, A. Jehangir, M. Asif, A. Ali, and N. Ahmad, "Bimodal human emotion classification in the speaker-dependent scenario," Pakistan Academy Sci., Art. no. 27, vol. 52, no. 1, pp. 27– 38, 2015.

[19] D. Gharavian, M. Bejani, and M. Sheikhan, "Audio-visual emotion recognition using FCBF feature selection method and particle swarm optimization for fuzzy ARTMAP neural networks," Multi- media Tools Appl., vol. 76, pp. 1–22, 2016.

[20] E. Ghaleb, M. Popa and S. Asteriadis, "Multimodal and Temporal Perception of Audio-visual Cues for Emotion Recognition," 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII), Cambridge, United Kingdom, 2019, pp. 552-558.

[21] R. Beard, R. Das, R. W. Ng, P. K. Gopalakrishnan, L. Eerens, P. Swietojanski, and O. Miksik, "Multi-modal sequence fusion via recursive attention for emotion recognition," in Proceedings of the 22nd Conference on Computational Natural Language Learning, 2018, pp. 251–259.

[22] E. Ghaleb, M. Popa and S. Asteriadis, "Metric Learning Based Multimodal Audio-visual Emotion Recognition," in *IEEE MultiMedia*.

[23] Emotion recognition using deep learning approach from audio–visual emotional big data M.Shamim Hossain , Ghulam Muhammad , Emotion Recognition Using Deep Learning Approach from Audio-Visual Emotional Big Data, Information Fusion (2018), doi: https://doi.org/10.1016/j.inffus.2018.09.008.

[24] Y. Ma, Y. Hao, M. Chen, J. Chen, P. Lu, A. Košir Audio-visual emotion fusion (AVEF): a deep efficient weighted approach. Inf. Fusion, 46 (2019), pp. 184-192

[25] Mansoorizadeh, M., Moghaddam Charkari, N. Multimodal information fusion application to human emotion recognition from face and speech. Multimed Tools Appl 49, 277–297 (2010). https://doi.org/10.1007/s11042-009-0344-2

[26] M. Haghighat, M. Abdel-Mottaleb and W. Alhalabi, "Discriminant Correlation Analysis: Real-Time Feature Level Fusion for Multimodal Biometric Recognition," in IEEE Transactions on Information Forensics and Security, vol. 11, no. 9, pp. 1984-1996, Sept. 2016.

Iranian Journal of Science and Technology, Transactions of Mechanical Engineering
https://doi.org/10.1007/s40997-020-00366-5

RESEARCH PAPER

Effect of Distributed Damping on the Dynamic Behavior of Flexible Chassis of Heavy Road Vehicle Under Standardized Random Road Responses

Ashish Gupta[1] · **Vikas Rastogi**[2]

Received: 22 April 2019 / Accepted: 23 May 2020
© Shiraz University 2020

Abstract
This paper evaluates the dynamics of a heavy road truck on different random road conditions, whereas the issue of beam flexibility is also being accompanied. Heavy vehicles like trucks and buses having a large structure to carry their load ride in all kinds of terrains, for thousands of miles without failing structure. Structure (chassis or frame) of the truck should be able to sustain all types of vibration (i.e., longitudinal, transverse and flexural). This study starts with developing an analytical model with consideration of the Rayleigh beam approach, where the rotary inertia of the beam has also been admitted. This work also considered internal damping of the chassis, which has been neglected in past studies specifically for flexural vibration. Further bond graph model of the flexible structure of a heavy road vehicle is developed, where this model is derived through modal expansion approach. The dynamic response of the structure is analyzed under various random road conditions at different vehicle speeds. These random models of the road are developed according to ISO 8608 standard, whereas four (H1–H4) kinds of road categories are considered. Results show the dynamic response of the chassis under a real road response so that parameters of the various parts of the structure can be evaluated, which further raises the ride comfort and road holding capability of the structure. Further, this work also examines the effect of structural damping or internal damping on the dynamics of the system.

Keywords Lagrangian approach · Bond graph · Frame flexibility of the chassis · Structural damping · Random road conditions

1 Introduction

Ride vibration is a principal concern for vehicle dynamics, where frame flexibility plays a very significant role. Numerous studies have been conducted on problems of ride vibration. However, very few studies have been carried out to study the effect of vibration, which includes a flexible frame. It has been generally noticed that the flexural vibration can be neglected for smaller vehicles, relatively for stiffer automobiles. However, trucks and the long heavy vehicle experience significant beaming mode vibration (Margolis and Edeal 1989). Ibrahim (1996) has created a model of truck frame through FEM where he calculated the modal properties of the frame. The results were presented in terms of PSD and RMS of the vehicle vertical response. Most of the analyses on the effects of chassis flexibility in vehicle dynamics were based on the use of particular model developed, whereas the FEM (finite element method) and the modal superposition theory have been extensively used to calculate the modal properties of the frame structure. Generally, the problem occurs due to the beam flexibility of the vehicle, which is termed as 'beaming response of the vehicle' (Rideout 2012). Rideout and Khan (2010) presented the model of the frame flexibility issue for truck model. They created a pitch plane model which involved longitudinal dynamics and transverse frame vibrations. Huang and Zeng (2010) studied about flexural vibration problems for high-speed passenger rail vehicle, where car body was modeled through Euler beam approach. This research underlined

✉ Ashish Gupta
ashish.gupta@glbitm.ac.in; ashish.j001@gmail.com

[1] Mechanical Engineering Department, G.L Bajaj Institute of Technology and Management, Greater Noida 201306, India

[2] Design Centre, Department of Mechanical Engineering, Delhi Technological University, Delhi 110042, India

Published online: 23 June 2020

the effect of internal damping in ride comfort of rail vehicle. Kumar et al. (2017) analyzed the whole-body vibration for railway system. They have considered biodynamic model of human for the assessment of ride comfort, where whole-body model was integrated with flexible bogie model system. The flexible bogie system was modeled through Euler–Bernoulli beam approach. Similarly in other studies, Euler–Bernoulli approach was attempted to introduce beam flexibility in heavy truck chassis modeling (Zhou et al. 2009; Kim and Yim 1994). Heavy long vehicles have a ladder-type chassis structure across the whole length of the vehicle. These structures are stiffer in bending, which can respond to bending frequencies more than 20 Hz. When the engine and other elements are mounted on the structure, these bending frequencies are reduced up to 8 Hz, for highly stiffer frames (Cai-hong and Zeng 2010; Tomioka et al. 2017). Tomioko and Takigami (2015) have conducted the experimental study on flexural beam to evaluate the effect of frame flexibility on passenger comfort. Some other studies apparently show that the beaming frequency is higher than the rigid frequency of the vehicle (Tomioka et al. 2006; Agostinacchio et al. 2014). However, in large vehicles, beaming frequency is influenced by road inputs.

A lot of research has been conducted considering the issue of frame flexibility in the analysis of railway vehicle. But, limited studies were reported on dynamic analysis of heavy road vehicles considering random road disturbances. A country like India, where multifarious road conditions are present, which creates an infirm condition before all leading automobile industries to maintain comfort level as per standard on every road conditions. Some studies have mentioned that structural damping can participate in ride quality improvement. So, in the present work frame flexibility has been incorporated by considering frame as Raleigh beam, which also includes the effect of the rotary inertia. Lagrangian approach has been applied to obtain the energy equations for the vehicle system, which resulted from the well-known rigid body modes and the modal parameters of the frame flexibility. The work involves the importance of structural or internal damping of beam-type chassis structure on flexural vibration and ride quality of the heavy vehicle with random road disturbance prescribed by ISO 8608 (Meirovitch 1980) as input. The bond graph approach has been utilized to model the vehicle with flexible frame. Besides the beam response, the effect of different road conditions as per ISO 8608 and vehicle speeds on the dynamics of vehicle has also been presented. The work begins with deriving Lagrange equation of frame considered as Rayleigh beam for truck flexural response. Then, this formulation is being extended for bond graph modeling with or without consideration of structural damping effect. Simulations were carried out to visualize the beaming mode response with respect to random road conditions as prescribed by ISO 8608.

2 Analytical Framework for Heavy Vehicle System Through Classical Mechanics

2.1 Generalized Lagrange's Equation for a Continuous System

It is a well-known fact that the Lagrange formulation provides one of the most convenient ways of writing down the equation of motion for a wide range of mechanical systems. The principal advantage of Lagrange's equation is that it is easier to apply to dynamical systems other than the simplest. This formulation is created from scalar quantities of kinetic energy, potential energy and work expressed in terms of generalized coordinates and developed for universal handling of dynamical systems (Mukherjee and Karmakar 2000). The equation is formulated equations of motion from the principle of least action. This density function is a function comprised of time, transverse displacements and transverse velocities, and up to second derivatives of displacements to space coordinates x, which may be expressed as

$$\widehat{L}\big(w(t,x),\dot{w}(t,x),w'(t,x),w''(t,x),\dot{w}'(t,x)\big)=\hat{T}(t,x)-\hat{V}(t,x) \tag{1}$$

where $\widehat{L}(\cdot)$ is the Lagrangian density function and $\hat{T}(\cdot)$ is the kinetic energy density, which may be expressed as,

$$\widehat{T}(x,t)=\widehat{T}(\dot{w}(x,t),\dot{w}'(x,t)) \tag{2}$$

Moreover, $\hat{V}(t,x)$ is the potential energy density, which has the following functional dependence,

$$\widehat{V}(x,t)=\widehat{V}(w'(x,t),w''(x,t)) \tag{3}$$

The next step is used for finding the equation of motion and the boundary conditions for the beam through the variational principle (see Appendix 2). This formulation yields

$$\delta\int_{t1}^{t2}\widehat{L}\big(w(t,x),\dot{w}(t,x),w'(t,x),w''(t,x),\dot{w}'(t,x)\big)\mathrm{d}t=0 \tag{4}$$

where $\delta(\cdot)$ is infinitesimal variation operator and it is very similar to the total derivative operator; the only difference is that $\delta(\cdot)$ does not vary with time. Another significant property of infinitesimal operator is that it commutes with the differential operator. Using the extremization condition, one may obtain

$$\int_0^L\int_{t1}^{t2}\delta\widehat{L}\big(w(t,x),\dot{w}(t,x),w'(t,x),w''(t,x),\dot{w}'(t,x)\big)\,\mathrm{d}x\mathrm{d}t=0 \tag{5}$$

or

$$\int_0^L \int_{t1}^{t2} \left[\frac{\partial \widehat{L}}{\partial w}\delta w + \frac{\partial \widehat{L}}{\partial \dot{w}}\delta \dot{w} + \frac{\partial \widehat{L}}{\partial w'}\delta w' + \frac{\partial \widehat{L}}{\partial \dot{w}'}\delta \dot{w}' + \frac{\partial \widehat{L}}{\partial w''}\delta w'' \right] \mathrm{d}x\,\mathrm{d}t = 0 \tag{6}$$

The detailed derivation of the variational formulation is presented in "Appendix 1."

2.2 Analytical Formulation for Flexural Behavior of the Truck Chassis Structure

An analytical formulation for flexural behavior of the truck structure is being attempted in this section. The vehicle base is modeled as a rigid body with a local coordinate reference frame (w, x) attached to the center of mass (M) and aligned with the inertia principal axes as shown in Fig. 1. It has density (ρ) and $p$itch inertia (J) with respect to the body $W_1(t) - W_2(t)$ displacement axis, EI flexural rigidity and $\theta(t)$ pitch velocity of the base. It is connected with the suspension system to different elements of the vehicles. The spring stiffness characterizes the spring and damper in the suspension system k_s and damping coefficient R_s, respectively. The various displacements of the vehicle are described $w(x, t)$ with respect to the equilibrium positions. As the vehicle is assumed to be rigid, its motion can be described by the vertical displacement (bounce) as $w(x, t)$ and the rotation about the transverse horizontal axis (pitch) as $\dot{\theta}(t)$.

The model of the vehicle is being created with the following assumptions:

- The components of the vehicle body act as a rigid body.
- The springs and dampers of the suspension system have linear characteristics.
- The spring damper system is assumed to be massless.
- The tire is assumed to provide stiffness and thus modeled as a spring element.
- The positions of the two ends of the spring connecting two rigid bodies or connecting one rigid body and one contact point are required as input data.
- The spring damper system connects each rigid body.
- The traveling road is assumed to be straight.
- The road inputs are assumed to be random.

The analytical framework of truck chassis can be assumed as a free–free beam structure, where Rayleigh beam approach is being adopted. Let us consider a transverse deflection of the beam $w(x,t)$, where x varies from 0 to beam length L. The formalism of beam flexibility is easily obtained through the well-known Lagrangian equation. Lagrangian density variation may be obtained from the following expression as

$$L = \left[\frac{1}{2}\int_0^L \left(\rho A \dot{w}(x,t)^2 + \rho I \dot{w}'(x,t)^2\right)\mathrm{d}x - \frac{1}{2}\int_0^L \left(EIw''(x,t)^2\right)\mathrm{d}x \right] \tag{7}$$

where ρ is the mass density, A is the area of cross section, EI is the rigidity and I is the second moment of area of the cross section of the beam about the neutral axis. In Eq. (7), the first term represents kinetic energy of the beam, the second term presents as inertial energy and the final term is strain energy. The variational formulation presents a refined and classical method of deriving the equation of motion of a dynamical system, where all the boundary conditions of a system can be revealed. However, fundamentals of variational formulation for the continuous system are detailed in "Appendix 2."

The equation of motion is obtained after applying a variational formulation and boundary conditions in free vibration condition, which may be expressed as

$$\rho A \ddot{w}(x,t) - \left(\rho I \dot{w}''(x,t)\right)' + \left(EIw''(x,t)\right)'' = 0 \tag{8}$$

This equation of motion is based on Rayleigh beam model. The term $\left(EIw''(x,t)\right)''$ is usually referred to as the

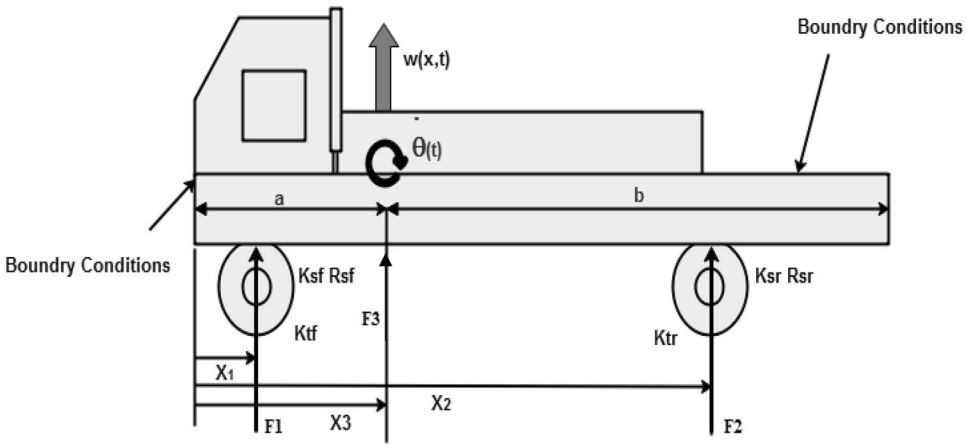


Fig. 1 Schematic diagram of a continuous frame of vehicle

flexural term, where EI is called the flexural stiffness and $\left(\rho I\dot{w}''(x,t)\right)'$ is known as the rotary inertia term. If this term is zero, then Eq. (8) will become a Euler beam model. Thus, the term $\left(\rho I\dot{w}''(x,t)\right)'$ can also be named as a Rayleigh beam term.

Assume a modal solution for Eq. (8) in the form of

$$w(x,t) = W(x)e^{i\omega t} \tag{9}$$

where ω is the eigenfrequency and $W(x)$ is the eigenfunction. The actual solution of the equation may be obtained by taking the real part of the solution. Substituting the modal solution in the field Eq. (8) yields after rearrangement as

$$-\omega^2\left[\rho AW - \left(\rho IW'\right)'\right] + \left(EIW''\right)'' = 0 \tag{10}$$

which along with the corresponding boundary condition represents the eigenvalue problem for a Rayleigh beam.

2.3 Modal Analysis for Uniform Rayleigh Beam

Considering Eq. (10) as a uniform beam and substituting in Eq. (9) a solution of the form

$$W(x) = Ce^{\lambda x} \tag{11}$$

where C and λ are constants, Eq. (10) can be rewritten as

$$EI\lambda^4 - \omega^2\rho I\lambda^2 - \omega^2\rho A = 0 \tag{12}$$

yielding

$$\lambda^2 = \frac{1}{2EI}\left[\omega^2\rho I \pm \sqrt{\omega^4\rho^2I^2 + 4\omega^2EI\rho A}\right]$$

Thus, the spatial equation (real form) of Eq. (12) has the general solution, which may be expressed as

$$W(x) = A\cosh\lambda x + B\sinh\lambda x + C\cos\lambda x + D\sin\lambda x \tag{13}$$

Here, one may consider a uniform Rayleigh beam for which the free–free boundary conditions are given by

$$W(0) = 0, \quad W'(0) = 0, \quad W''(l) = 0, \quad W'''(l) = 0, \tag{14}$$

Using, Eq. (14) in Eq. (13) yields the frequency equation as

$$\cosh\lambda_n l\cos\lambda_n l - 1 = 0 \tag{15}$$

The solution of Eq. (15) yields

$$\lambda_n = \left(\frac{2n+1}{2}\pi\right)\frac{1}{l}$$

Moreover, the mode shape function may be expressed as

$$\begin{aligned} W_n(x) = {} & (\cos\lambda_n l - \cosh\lambda_n l)(\sin\lambda_n x + \sinh\lambda_n x) \\ & - (\sin\lambda_n l - \sinh\lambda_n l)(\cos\lambda_n x + \cosh\lambda_n x) \end{aligned} \tag{16}$$

Thus, one may obtain the natural frequency of Eq. (6), which may be written as

$$\omega_n = \lambda_n^2\frac{1}{\left[1+\lambda_n^2\frac{I}{A}\right]^{1/2}}\sqrt{\frac{EI}{\rho A}} \tag{17}$$

It is recognized that conditions are force-free at this point, which permits $\omega_n = 0$ to be a mode frequency. Thus, Eq. (10) will be written as

$$\frac{\mathrm{d}^4W}{\mathrm{d}x^4} = 0 \tag{18}$$

or

$$W = c_1x^3 + c_2x^2 + c_1x + c_4 \tag{19}$$

There are two possible solutions for Eq. (13), which satisfied all boundary conditions. They are $W = \text{const}$ and $W = ax + b$. These are called rigid body modes, and it is convenient to assume them as rigid body rotation of the beam about the centrally located center of mass. Thus,

$$W_{00}(x) = 1 \tag{20}$$

and

$$W_0(x) = x - \frac{L}{2} \tag{21}$$

Let us consider an analysis for truck chassis, which is presented in Fig. 1. In this figure, the half model of the road truck is supported by front and rear suspension, which exerted an external force on the beam over the various road responses. From equation of motion, Eq. (8) will be rewritten as

$$\begin{aligned} & \rho A\ddot{w}(x,t) - \left(\rho I\dot{w}''(x,t)\right)' + \left(EIw''(x,t)\right)'' \\ & = F_1\delta(x-x_1) + F_2\delta(x-x_2) + F_3\delta(x-x_3) \end{aligned} \tag{22}$$

where F_1 and F_2 are the force exerted by front and rear suspension systems, respectively, and F_3 is dead weight on the beam, $\delta(\cdot)$ is the Dirac delta function, x_1 and x_2 are the respective distance from left end to the front and rear suspension of the vehicle, respectively, and x_3 is the distance of dead weight from the left end. One may assume that the solution of Eq. (22) is separable in space and time, and the forced solution has the following form,

$$w(x,t) = \sum_{n=0}^{\infty} W_n(x)U_n(t) \tag{23}$$

where $W_n(x)$ depends on the spatial position and $U_n(t)$ depends on time. Introducing Eq. (23) into Eq. (22), and multiplying each term by $W_m(x)$ and integrating with respect

to $x=0$, $x=l$, one may then use the orthogonal property to the modes and obtain

$$\int_0^l \rho A W_n^2 \mathrm{d}x\ddot{U}(t) + \omega_n^2(\rho I\lambda_n^2 - \rho A)\int_0^l W_n^2 \mathrm{d}xU(t) = F_1 W_n(x_1) + F_2 W_n(x_2) + F_3 W_n(x_3) \tag{24}$$

$$\rho A\left(1 - \frac{I}{A}\lambda_n^2\right)\int_0^l W_n^2 \mathrm{d}x\ddot{U}(t) + \rho A\left(1 - \frac{I}{A}\lambda_n^2\right)\omega_n^2\int_0^l W_n^2 \mathrm{d}xU(t) = F_1 W_n(x_1) + F_2 W_n(x_2) + F_3 W_n(x_3) \tag{25}$$

Equation (25) can be represented in general form as

$$m\ddot{U}(t) + kU(t) = F_1 W_n(x_1) + F_2 W_n(x_2) + F_3 W_n(x_3) \tag{26}$$

where

$$m = \rho A\left(1 - \frac{I}{A}\lambda_n^2\right)\int_0^l W_n^2 \mathrm{d}x$$

$$k = m\omega_n^2\int_0^l W_n^2 \mathrm{d}x$$

Now one may write the first frequency modes (Karnopp et al. 2012)

$$\left[\rho A\int_0^l (1)_n^2 \mathrm{d}x\right]\ddot{U}_{00} = F_1 + F_2 + F_3 \quad \text{or} \quad m\ddot{U}_{00} = F_1 + F_2 + F_3 \tag{27}$$

which states that the external force accelerates the center of mass of the beam. The other zero frequency mode yields

$$\left[\rho A\int_0^l \left(x - \frac{l}{2}\right)^2 \mathrm{d}x\right]\ddot{U}_{00} = F_1 + F_2 + F_3 \tag{28}$$

or

$$J\ddot{U}_0 = F_1 + F_2 + F_3$$

This equation states that the moment of the external forces about the center of mass produces angular acceleration $\ddot{U}_{00}$, where J is the centroid moment of inertia of beam. The next subsection will present the formulation, which includes structural damping.

2.4 Incorporation of Structural Damping

Now one may consider internal damping (the detailed derivation of internal damping is presented in "Appendix 2" for ready reference) of the beam and Eq. (22) may be rewritten as

$$\rho A\ddot{w}(x,t) - \left(\rho I\dot{w}''(x,t)\right)' + \left(EIw''(x,t)\right)'' - \mu_I I\left(\dot{w}'''\right)' = F_1\delta(x - x_1) + F_2\delta(x - x_2) + F_3\delta(x - x_3) \tag{29}$$

$$\rho A\ddot{w}(x,t) - \left(\rho I\dot{w}''(x,t)\right)' + \omega^2(\rho I\lambda^2 - \rho A) - \mu_I I\left(\dot{w}'''\right)' = F_1\delta(x - x_1) + F_2\delta(x - x_2) + F_3\delta(x - x_3) \tag{30}$$

where μ_I is the structural damping coefficient, δ is the Dirac delta function, x_1 and x_2 are the respective distance from left corner to the front and rear suspension of the vehicle and x_3 is the distance of dead weight from the extreme left corner of the vehicle. F_1 and F_2 are the suspension forces of the front and rear suspension system, respectively, and F_3 is the dead weight. The front and rear suspension forces of the vehicle may be expressed as

$$F_1 = C_{\mathrm{s}}^{\mathrm{f}}(V_{\mathrm{f}} - \dot{w}_1) + K_{\mathrm{s}}^{\mathrm{f}}\left(\int_0^t V_{\mathrm{f}}\mathrm{d}t - w_1\right) \tag{31}$$

$$F_2 = C_{\mathrm{s}}^{\mathrm{r}}(V_{\mathrm{r}} - \dot{w}_2) + K_{\mathrm{s}}^{\mathrm{r}}\left(\int_0^t V_{\mathrm{r}}\mathrm{d}t - w_2\right) \tag{32}$$

where $C_{\mathrm{s}}^{\mathrm{f}}$ and $C_{\mathrm{s}}^{\mathrm{r}}$ are, respectively, damping coefficient of the front and rear damper; V_{f} and V_{r} are the road velocity input at front and rear wheel, respectively; $K_{\mathrm{s}}^{\mathrm{f}}$ and $K_{\mathrm{s}}^{\mathrm{r}}$ are the suspension stiffness of front and rear suspension system, respectively; $\dot{w}_1$ and w_1 are the velocity and displacement of front sprung mass; and $\dot{w}_2$ and w_2 are the velocity and displacement of rear sprung mass.

One may assume solution similar to Eq. (24); applying Eqs. (23)–(30), one may obtain a solution as

$$\int_0^l \rho A W_n^2 \mathrm{d}x\ddot{U}(t) + \omega_n^2(\rho I\lambda_n^2 - \rho A)\int_0^l W_n^2 \mathrm{d}xU(t) - \rho I\lambda_n^2\int_0^l W_n^2 \mathrm{d}x\ddot{U}(t) - \mu_I I\lambda_n^4\int_0^l W_n^2 \mathrm{d}x\dot{U}(t) = F_1 W_n(x_1) + F_2 W_n(x_2) + F_3 W_n(x_3) \tag{33}$$

Rearranging Eq. (33), one may obtain

$$\rho A\left(1 - \frac{I}{A}\lambda_n^2\right)\int_0^l W_n^2 \mathrm{d}x\ddot{U}(t) + \rho A\left(1 - \frac{I}{A}\lambda_n^2\right)\omega_n^2\int_0^l W_n^2 \mathrm{d}xU(t) - \mu_I I\lambda_n^4\int_0^l W_n^2 \mathrm{d}x\dot{U}(t) = F_1 W_n(x_1) + F_2 W_n(x_2) + F_3 W_n(x_3) \tag{34}$$

Finally, Eq. (34) may be represented as

$$m_n\ddot{U}(t) + R_n\dot{U}(t) + k_nU(t) = F_1W_n(x_1) + F_2W_n(x_2) + F_3W_n(x_3) \tag{35}$$

where

$$m_n = \rho A\left(1 - \frac{I}{A}\lambda_n^2\right)\int_0^l W_n^2\mathrm{d}x, \tag{36}$$

$$k_n = m\omega_n^2\int_0^l W_n^2\mathrm{d}x, \tag{37}$$

$$R_n = \mu_I I\lambda_n^4\int_0^l W_n^2\mathrm{d}x \tag{38}$$

Equations (36–38) show that the rotary inertia affects the dynamics of the beam, which is not being taken in case of Euler beam. However, in this case, the value of $I/A \cdot \lambda_n^2$ is always less than 1. Generally, distributed elements that one may obtain to involve in almost all the system models are lightly damped. The location of structure damping cannot be precisely identified from where it arises. The mechanism of energy dissipation is very typical and complicated; some dissipation takes place because of the plastic deformation and some of the energy radiation from material surfaces. It is not assured in the modeling that individual conventional damper will attach to model for representing the damping of distributed elements. It can happen if the damping includes functionally by incorporating it into individual modes. This is accompanied by merely appending R-elements to the mode oscillators of the modal bond graph. R-element is represented in Eq. (30), where μ_I is varied from 0.01 to 0.1. The next subsection will present the bond graph modeling to visualize the simulation for different parameters.

3 Bond Graph Modeling of the Truck Chassis Structure

The physical system of a truck chassis (pitch plane model) with front and rear suspension system driven over the random road is presented in Fig. 1. The boundary condition of the structure is taken as free–free. The dynamic disturbance in vertical direction of the vehicle due to random road conditions is the main attraction of this modeling. The longitudinal and lateral dynamics are not considered in the modeling. The bond graph technique is being used as a modeling tool, and the bond graph of the system is created through formulation, which was developed in the previous section. Figure 2 presents the bond graph model of a flexible infinitesimal beam, whereas Fig. 3 shows the bond graph model of a flexible beam with internal damping along with suspension system. The rigid body modes appear simply as inertia elements with associated modal stiffness and damping. The first inertia parameter is the beam mass (M), and second is the beam moment of inertia (Jg). The transformer (-TF-) elements are connected to the rigid body modes, which correctly apply the forces and moments to these elements. The modal masses are taken from Eq. (36), whereas modal stiffness and modal damping can be taken from Eqs. (37) and (38), respectively.

The system is entirely flexible in transverse and bending, but torsionally rigid. Each 0-junction is attached to each 1-junction with modal components through transformer (TF) elements with moduli equal to appropriate mode function. The transformer function is moduli with shape function (W_n), and the values of this function can be determined from Eq. (16). The attached systems are then appropriately appended to the external 0-junction bonds to form a complete, low-order, very accurate model. Figure 3 shows a general finite mode representation with two input (front and rear) location, where Eq. (26) represents a governing contribution of *i*th mode.

In the real world, distributed elements have some damping properties, which resist vibration level up to a certain level. No mechanism can quickly identify the value of damping on specific. However, only the quantity of dissipative energy can be determined. It is customary not to attempt to perform detailed modeling of the damping mechanism, but instead to include damping functionally by incorporating its individual modes. This is accomplished by simply appending R-elements to the mode oscillator of the bond graph. The bond graph model of the distributed system including damping modes is presented in Fig. 3. The governing equation of this configuration is mentioned in Eq. (23).

4 Numerical Simulation

Computer simulation can compress the performance of a system over the years into a few minutes of the computer running time. The parameters used for simulation for truck chassis system are presented in Table 1. Simulation models are relatively flexible and can be modified to accommodate the changing environment to the real situation. At present, most of the simulation models are made using differential equations. In this research, a half car flexible model with suspension systems is investigated at

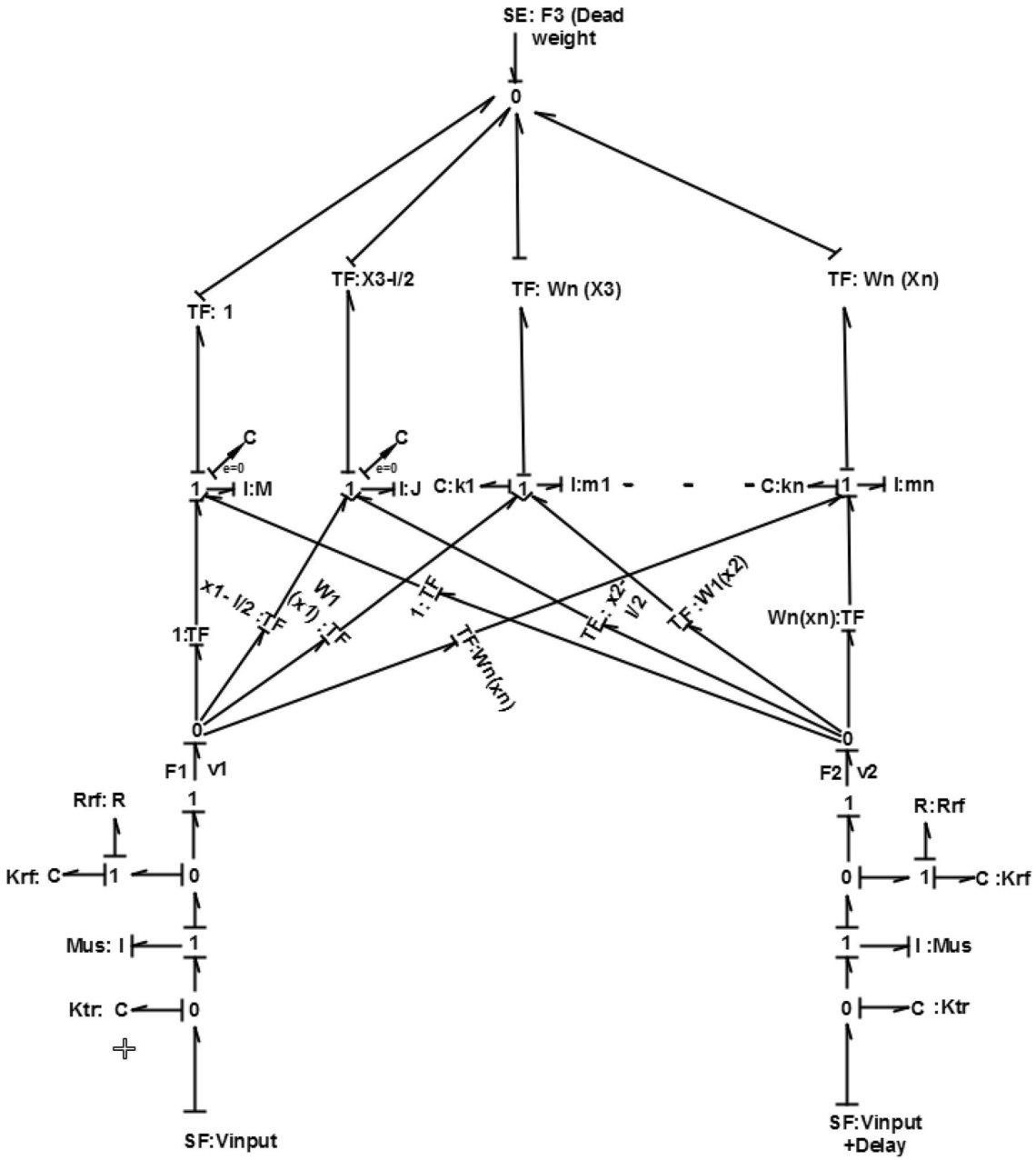


Fig. 2 Bond graph model of truck chassis considering internal damping

different operating speeds on different road profile inputs using bond graph technique and simulators of SYMBOLS Sonata® and MATLAB/Simulink® software are used (Gupta and Rastogi 2016; Hassani 2013; Mukherjee and Samantaray 2006). The bond graph model of the vehicle is simulated for 10 s to obtain different output responses. In total, 1024 records are used in the simulation and error is kept in the order of 5.0×10^{-4}. Fifth-order Runge–Kutta Gill method is used in this present work to solve the various differential equations generated through bond graph model.

4.1 Road Inputs

Two types of road profile inputs of the simulation of this system are used, viz. steady-state input and random road input.

4.1.1 Sine Wave Input

The steady-state input has been taken in a sinusoidal form, and the following expression is used for simulation (Hagedorn and DasGupta 2007):

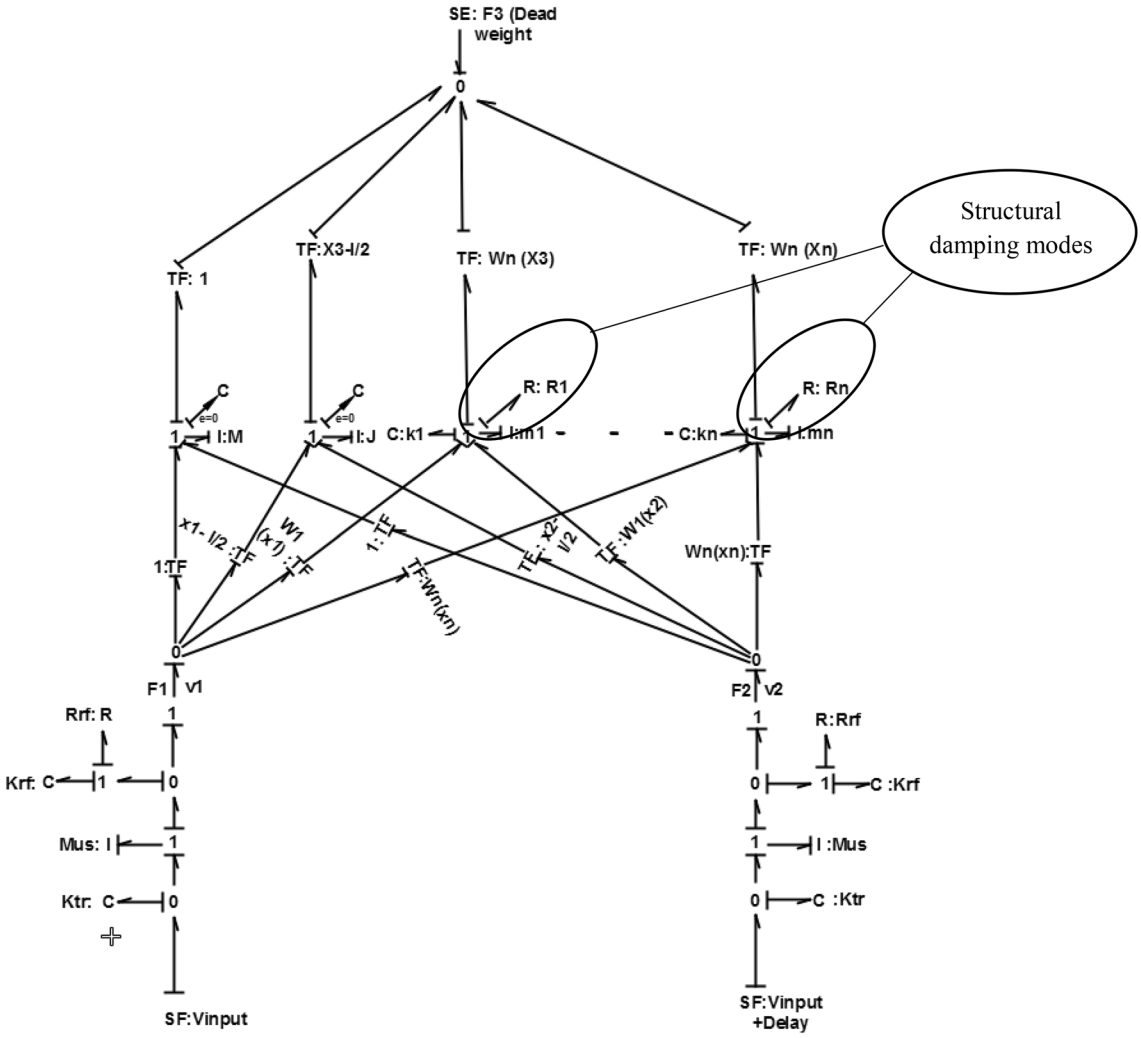


Fig. 3 Bond graph model of truck chassis considering internal damping

Table 1 Simulation parameters for flexible chassis system (Rideout and Khan 2010)

Parameter	Value	Unit
Vehicle length	10	m
Frame mass (Mg)	4350	kg/m
Frame inertia (J)	24,000	kg/m^2
Suspension stiffness (K_{sf}, K_{sr})	375, 870	kN/m
Suspension damping (R_{sf}, R_{sr})	31,895, 33,384	Ns/m
Flexural rigidity (EI)	1.05e+08	N m^2
Unsprung mass (Mus)	50	kg
x_1, x_2, x_3, x_4	2.95, 5.75, 7, 3	m

$$V_{\text{input}} = A_{\text{jump}} * \omega * \sin(\omega * t), \quad \text{for front wheel} \tag{39}$$

$$V_{\text{input}} = A_{\text{jump}} * \omega * \sin(\omega * (t + d/v)), \quad \text{for rear wheel} \tag{40}$$

where A_{jump} is the amplitude of vibration and ω is the excitation frequency.

4.1.2 Random Road Condition

A random road profile is generated according to the International Organization for Standardization (ISO 8608). It gives a depiction of the road profile through estimation of the PSD of the vertical displacements G_d, as a function of spatial frequency $n(n = \Omega/2\pi(\text{cycles})/\text{m})$ and also of angular spatial frequency Ω. ISO 8608 introduces a classification, which is evaluated in accordance with conventional values of spatial frequency $n_0 = 0.1$ cycles/m and angular spatial frequency $\Omega_0 = 1$ cycles/m. There are eight classes of roads mainly

identified from classes H1 to H8 according to the values of $G_d(n)$ and $G_d(\Omega)$ established in ISO 8608, which are shown in Table 2 (Meirovitch 1980).

In the simulation, ISO 8608 provides the roughness of the road surface profile, which may be stated using the following equations:

$$G_d(n) = G_d(n_0)\left(\frac{n}{n_0}\right)^{-w} \tag{41}$$

$$G_d(\Omega) = G_d(\Omega_0)\left(\frac{n}{n_0}\right)^{-w} \tag{42}$$

where w denotes the waviness and its value is taken to be 2 in this case. In this research, random road inputs have been developed by taking into consideration the PSD of vertical displacement G_d as a function of spatial frequency n.

Beginning with a distributed road profile to specify the value of spatial frequency n, looped within a frequency band Δn, the PSD function value is represented through the following expression as

$$G_d(n) = \lim_{\Delta n \to 0}\left(\frac{\psi_x^2}{\Delta n}\right) \tag{43}$$

where the mean square value is denoted by ψ_x^2 which signifies the component of the signal for spatial frequency n contained by the frequency band Δn. Accordingly, the signal of road profile is discretized, and thus, it is characterized by a series of elevation points along with evenly spaced ones. Here, road profile length is denoted by L, whereas the sampling interval is denoted by B. Further, the maximum value of sampling spatial frequency c is $1/B$, the maximum effective sampling spatial frequency (n_{eff}) is $n_{\max}/2$ and the discretized spatial frequency values n_i are equally spaced under the frequency domain, with an interval of (Δn) as $1/L$.

So herein, generic spatial frequency value n_i can be regarded as $i{\cdot}\Delta n$ and Eq. (43) may be written in the discrete form as

$$G_d(n) = \frac{\psi_x^2(n_i, \Delta n)}{\Delta n} = \frac{\psi_x^2(i \cdot \Delta n, \Delta n)}{\Delta n} \tag{44}$$

where i varying from 0 to N as $\frac{n\max}{\Delta n}$, one may obtain the road profile by a simple harmonic function

$$h(x) = A_i \cos(2\pi \cdot n_i \cdot x + \phi) = A_i \cos(2\pi \cdot i \cdot x \cdot \Delta n + \phi) \tag{45}$$

where the amplitude is denoted by A_i, spatial frequency is denoted by n_i and phase angle is φ. One may generate a harmonic signal through mean square value

$$\psi_x^2 = \frac{A_i^2}{2} \tag{46}$$

Therefore,

$$G_d(n_i) = \frac{\psi_x^2(n_i)}{\Delta n} = \frac{A_i^2}{2 \cdot \Delta n} \tag{47}$$

It has been shown in several works that the development of an artificial road profile using Eq. (45) is only possible if the PSD function of vertical displacements is well known, presuming a random phase angle φ_i following a uniform probabilistic distribution within range 0 to 2π. The artificial profile can be given as

$$\begin{aligned} h(x) &= \sum_{i=0}^{N} A_i \cos(2\pi \cdot n_i \cdot x + \phi_i) \\ &= \sum_{i=0}^{N} \sqrt{2 \cdot \Delta n \cdot G_d(i \cdot \Delta n)} \cdot A_i \cos(2\pi \cdot i \cdot x \cdot \Delta n + \phi) \end{aligned} \tag{48}$$

Using Eq. (47) in Eq. (45), a random road profile can be generated according to ISO classification by the following equation,

Table 2 ISO 8608 values of $G_d(n_0)$ and $G_d(\Omega_0)$

Road class	$G_d(n_0)$ (10^{-6} m^3)		$G_d(\Omega_0)$ (10^{-6} m^3)		Category
	Lower limit	Upper limit	Lower limit	Upper limit	
H1	–	32	–	2	New roadways with asphalt layers
H2	32	128	2	8	New roadways with concrete layers
H3	128	512	8	32	Old roadways layers which are not maintained
H4	512	2048	32	128	Roadways layer consisting if cobble stone layers or similar materials
H5	2048	8192	128	512	Very poor road condition
H6	8192	32,768	512	2048	Off-road condition
H7	32,768	131,072	2048	8192	–
H8	131,072	–	8192	–	–
	$n_0 = 0.1$ cycles/m		$\Omega_0 = 1$ cycles/m		

$$h(x) = \sum_{i=0}^{N} \sqrt{\Delta n.} \cdot 2^{k} \cdot 10^{-3} \cdot \left(\frac{n_0}{i \cdot \Delta n}\right) \cos(2\pi \cdot i \cdot x \cdot \Delta n + \phi) \tag{49}$$

where x is the abscissa variable from 0 to L; Δn is $1/L$; $n_{\max}$ is taken as $1/B$; N is $\frac{^{n}\max}{\Delta n}$ or L/B; where $L = 250$, $N = 100$; and constant is denoted as k, depending on ISO road profile classification. It has also assumed integers augmenting from 3 to 9, which corresponds to the profiles from classes H1 to H8 (as shown in Table 3); here, k is assumed to take the value 3 corresponding to the class H1 road profile and $G_d(n_0)$ is taken to be 32, where n_0 is 0.1 cycles/m and φ_i is the random phase angle following an uniform probabilistic distribution within range 0 to 2π. In this work, four road conditions (H1-H4) are considered, which is shown in Figs. 4, 5, 6 and 7.

The simulation of different road categories is obtained by MATLAB/Simulink of the previously discussed ISO 8608 random input. This road profiles are simulated for five different speeds, viz. 40 kmph, 60 kmph, 80 kmph, 90 kmph and 100 kmph. Figure 4 shows H1 road condition, where unevenness of the random-type road with maximum magnitude is 0.03 m in an upward direction at 40 km/h and − .035 m in a downward direction at 100 km/h. Similarly, H2, H3 and H4 road categories have a maximum magnitude 0.06 m, 0.12 m and 0.24 m, respectively, in an upward direction at 40 km/h and − 0.65 m, − 0.12 m and − 0.24 m, in a downward direction at 100 km/h.

Table 3 k values for ISO road roughness classification

Road class		k
Lower limit	Upper limit	
H1	H2	3
H2	H3	4
H3	H4	5
H4	H5	6
H5	H6	7
H6	H7	8
H7	H8	9

5 Results and Discussion

The model is exhibited first on sine wave input from start to end and runs for 20 s with zero initial conditions. The amplitude of sine wave is 0.01 m, and the input frequency is taken 50 Hz. In Fig. 8, the magnitude of the vertical acceleration is 0.8 m/s^2 in un-damped condition, whereas acceleration amplitude has been reduced to 0.7 m/s^2 in damped condition. Thus, the magnitude of vibration has been reduced from 12 to 15% due to the impact of structural damping of the vehicle chassis. Similarly, in Fig. 9, the amplitude of pitching acceleration has been reduced up to 20% due to the influence of structural damping. However, frequencies in both conditions have almost similar values. The dead load on the center of chassis has been taken around 500 N, which lies in the downward direction.

The RMS acceleration responses of truck chassis at the first three modes at sine wave condition are shown in Fig. 10a–c. In these figures, the un-damped condition of the beam is superimposed over damped condition. The frequencies of the first three modes are 34.4 Hz, 86.89 Hz and 174.8 Hz, respectively. Figure 10a–c shows that the acceleration magnitude is continuously reduced, whereas maximum acceleration is found at the first mode, and minimum

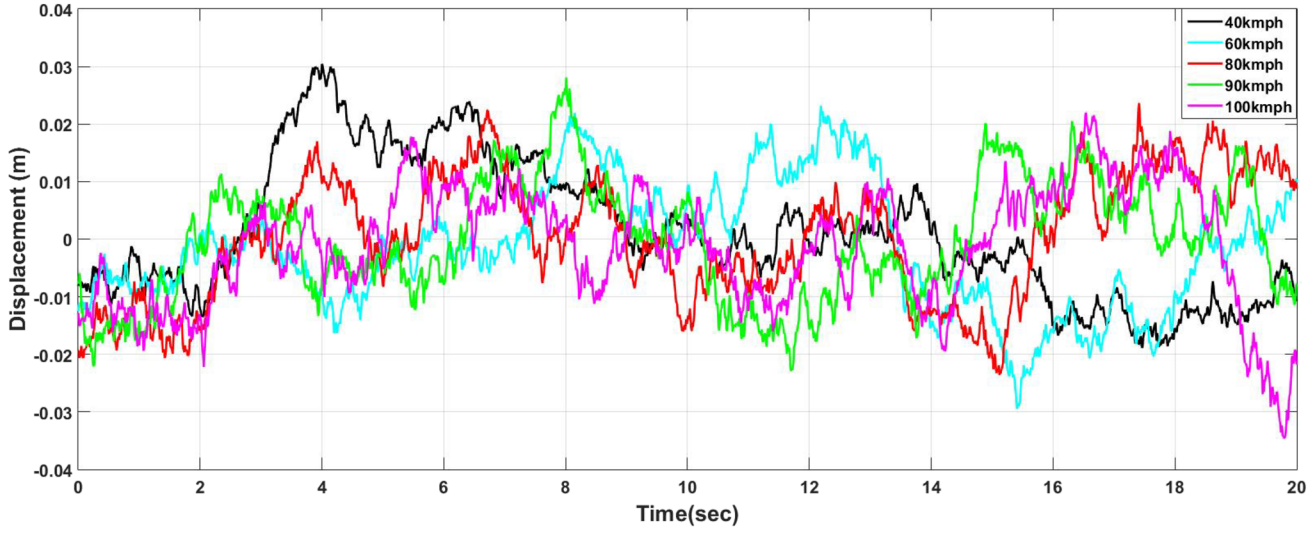


Fig. 4 Random road profile of H1 road category at 40 kmph, 60 kmph, 80 kmph, 90 kmph and 100 kmph

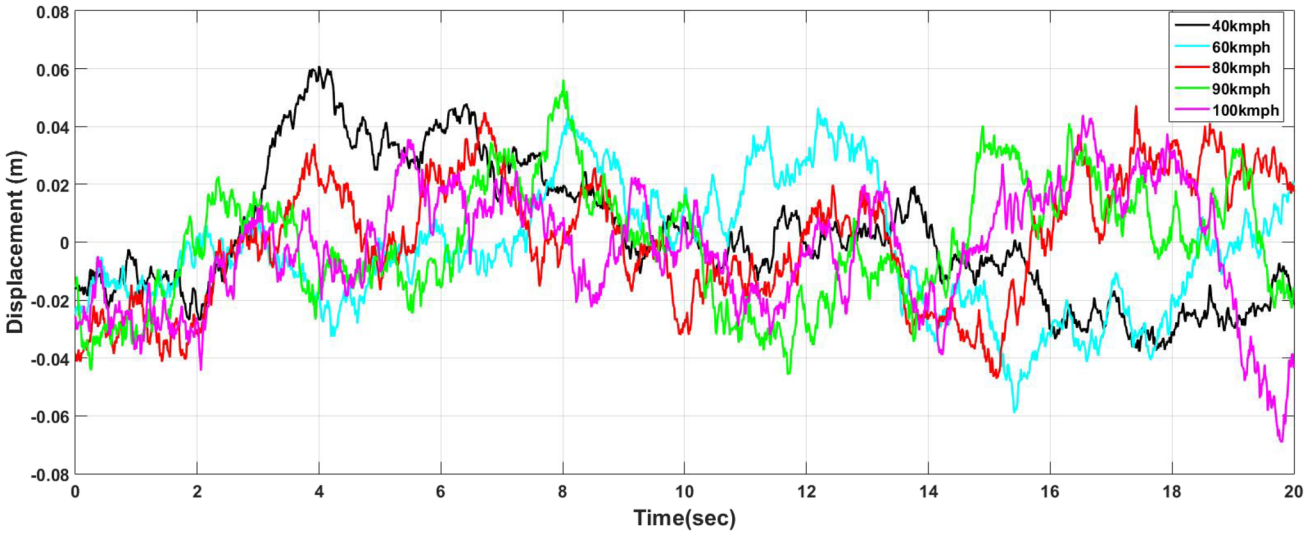


Fig. 5 Random road profile of H2 road category at 40 kmph, 60 kmph, 80 kmph, 90 kmph and 100 kmph

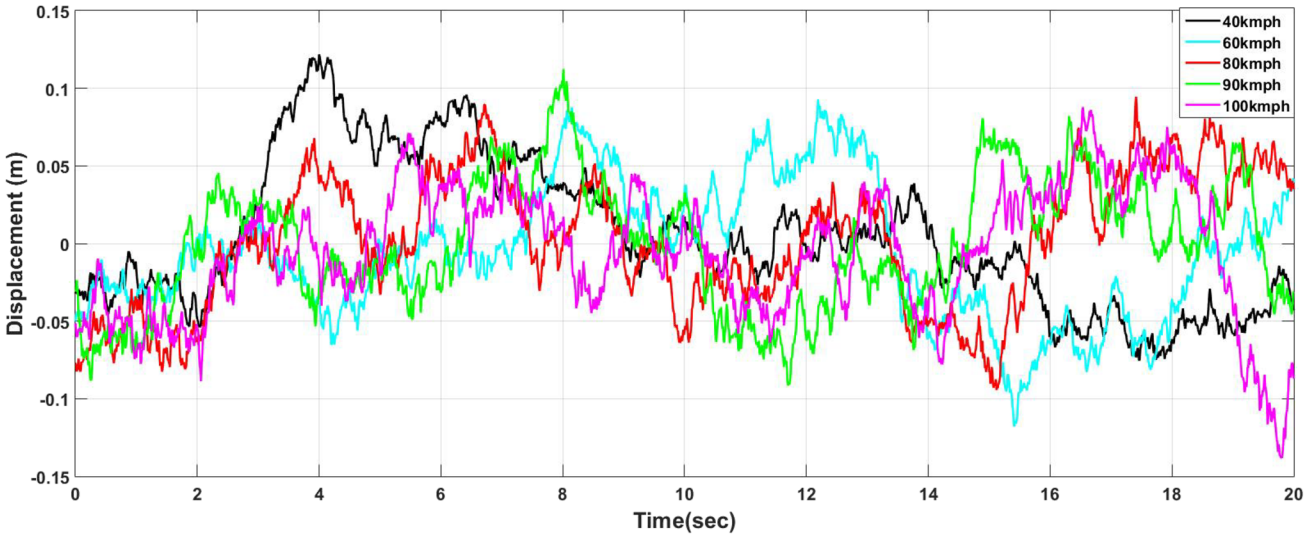


Fig. 6 Random road profile of H3 road category at 40 kmph, 60 kmph, 80 kmph, 90 kmph and 100 kmph

acceleration is in the third mode. The influence of internal damping is presented in terms of RMS acceleration response, whereas the amplitude has reduced due to the effect of internal damping at the same frequency.

The ride comfort analysis may be evaluated through RMS acceleration of the vehicle chassis. Figures 11, 12, 13 and 14 present the RMS acceleration response of the truck CG for four different classes of roads (H1–H4) as prescribed by ISO 8608. The H1 is the best class, and H4 represents the worst road class in these four classes. The model is simulated for each road class at different speeds ranging from 40 to 100 kmph with and without consideration of the internal damping of the chassis. Figures 11, 12, 13 and 14 show that the magnitude of vertical acceleration increases with vehicle speed and with road class. Figure 11a demonstrates the truck response neglecting the chassis damping, where maximum RMS acceleration magnitude is observed as 0.65 m/s^2 at 100 km/h and 0.25 m/s^2 at 40 km/h on H1 class road. This peak acceleration magnitude gets reduced up to 0.58 m/s^2 at 100 km/h and 0.22 m/s^2 at 40 km/h with due consideration to chassis structural damping, as shown in Fig. 11b. A similar kind of result has been noticed in other plots shown in Figs. 12, 13 and 14. Thus, it is evident from the above results that internal structural damping of the base structure of heavy vehicle plays an important role

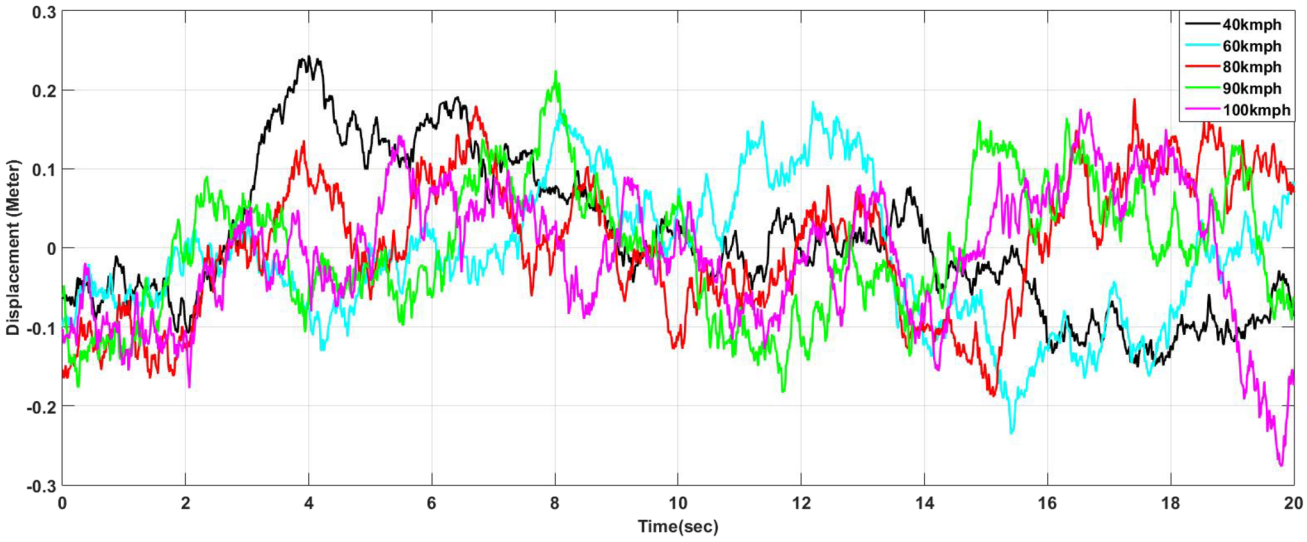


Fig. 7 Random road profile of H4 road category at 40 kmph, 60 kmph, 80 kmph, 90 kmph and 100 kmph

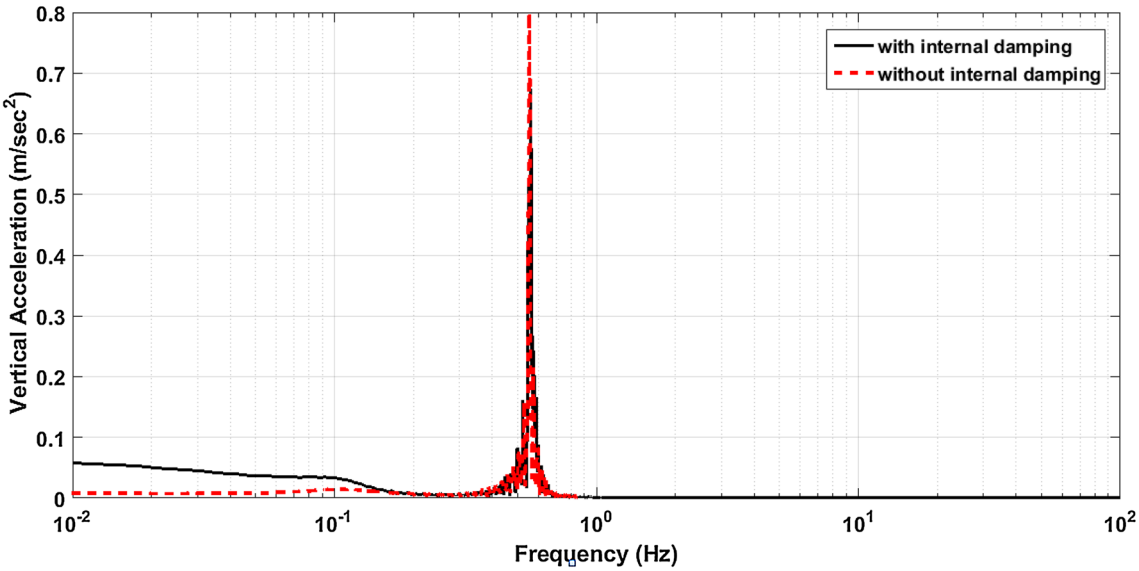


Fig. 8 Vertical response of vertical acceleration of chassis at steady-state input

and it needs to be included in the dynamical model of the vehicle to get the true dynamic behavior of the vehicle.

It has been observed that from the RMS curves that the H1 road condition is perfect for driving the vehicle for long duration, whereas it is not advisable to drive vehicles for long duration on H4 road condition. Indian roads lie between H2 and H3 conditions. So, a driver can comfortably drive the given vehicle at 60–80 km/h in single stretch without a break.

6 Conclusions

Numerous models have been constructed for the heavy vehicle system, but most of the models are developed with the lumping process. In this work, an analytical framework for truck chassis has been developed through classical mechanics and through extended Lagrangian approach. The following conclusion has been drawn from present research work:

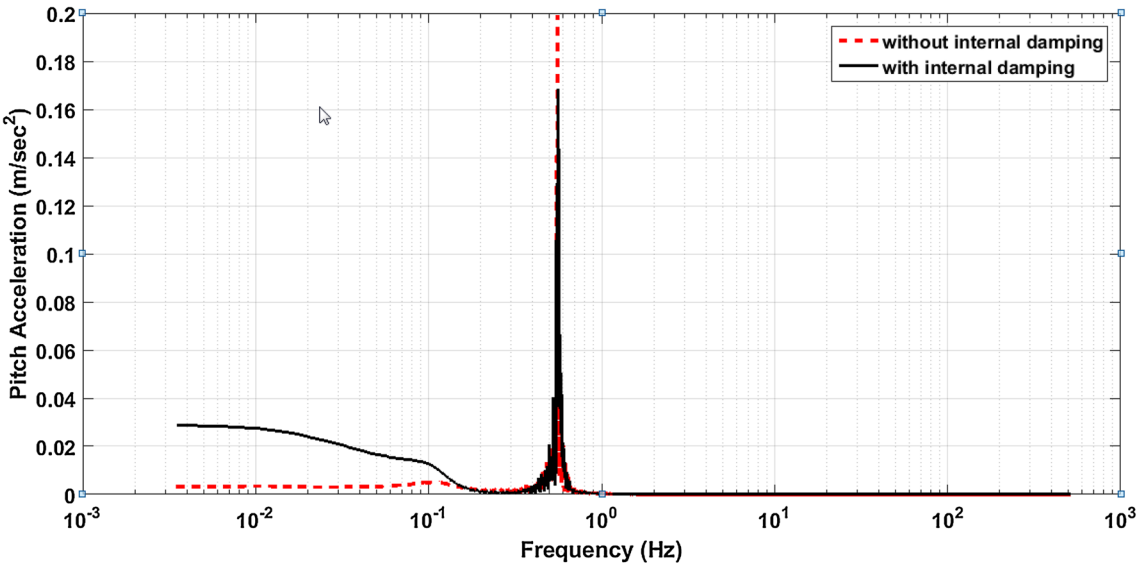


Fig. 9 Response of pitch acceleration of chassis at steady-state input

- In the classical mechanics approach, the idea of lumped–distributed modeling has been explored within the framework of bond graphs.
- The Lagrange variation of truck chassis, including flexural behavior has been constructed, where it was integrated with lumped suspension system. However, the structural damping effects were also involved in terms of fifth-order partial differential term in generalized equation of motion. The detailed derivation for fifth-order differential term is presented in "Appendix 1."
- The integrated bond graph model of flexural beam was the extended form of the analytical model. Results have been presented under the two types of road conditions, i.e., sine wave input and random road input. In sine wave road conditions, the vertical response of flexural beam is reduced up to 12–15% and pitch response of beam is reduced up to 20% due to the structural damping effect. In random road conditions, four types of road category (H1, H2, H3, H4) have been used at four different speeds (40 kmph, 60 kmph, 80 kmph and 100 kmph).
- It is depicted from the result that maximum acceleration at maximum speed (100 kmph) of chassis has been increased up to 0.5 m/s^2 in H1–H2 conditions, 1 m/s^2 in H2–H3 conditions and 2.25 m/s^2 in H3–H4 conditions. Similarly, the solution has been extended for structural or internal damping, whereas maximum acceleration at maximum speed has been reduced up to 10%, 5%, 7% and 5% (approx.) in H1, H2, H3 and H4 road conditions, respectively.

Appendix 1

This appendix gives a brief introduction to the variational formulation of dynamics of continuous systems. Consider for simplicity a one-dimensional continuous system with the field variable $w(x, t)$ which uniquely represents the configuration of the system at any time t. In the course of temporal evolution of the system, let the configurations at two time instants $t = t_1$ and $t = t_2$ be recorded as, respectively, $w(x, t_1)$ and $w(x, t_2)$. The actual path is that any infinitesimal variation over that path should leave the value of I unchanged (Hagedorn and DasGupta 2007). Now the action integral may be expressed as

$$I = \int_{t_0}^{t_1} \int_{x_0}^{x_1} \delta L(\ldots)\, \mathrm{d}t\mathrm{d}x \tag{50}$$

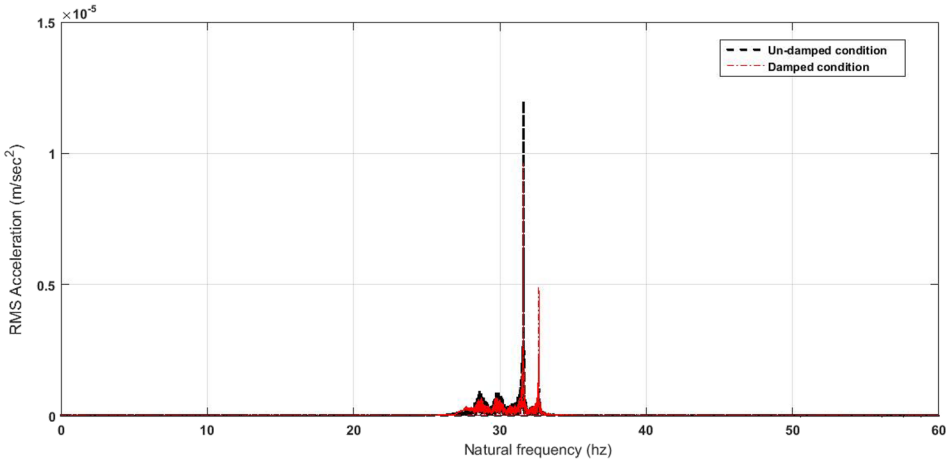
Using the extremization condition of Eq. (50), one may obtain

$$\int_{t_0}^{t_1} \int_{x_0}^{x_1} \delta \hat{L}\left(w, w', \dot{w}, w'', t\right) \mathrm{d}t\mathrm{d}x = 0 \tag{51}$$

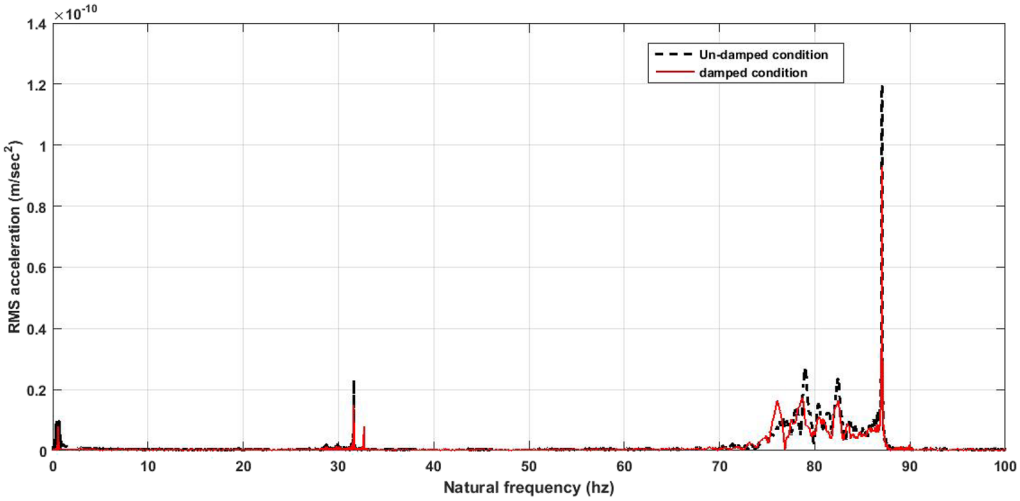
One may write Lagrangian density variation,

$$\delta \hat{L} = \frac{\partial \hat{L}}{\partial w}\delta w + \frac{\partial \hat{L}}{\partial \dot{w}}\delta \dot{w} + \frac{\partial \hat{L}}{\partial w'}\delta w' + \frac{\partial \hat{L}}{\partial w''}\delta w'' \tag{52}$$

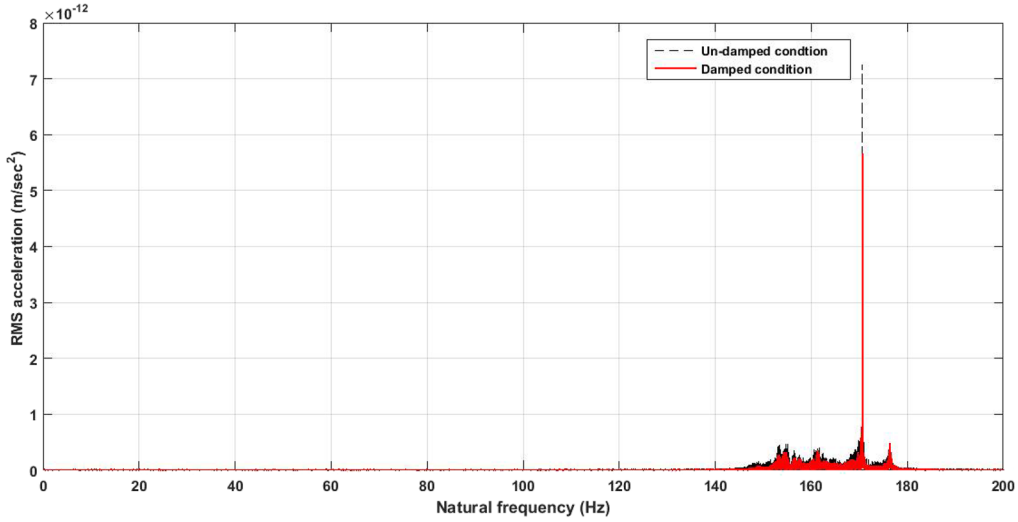
Introducing Eqs. (51) and (52), one may have


×10^-5
Un-damped condition
Damped condition
RMS Acceleration (m/sec^2)
Natural frequency (hz)

(a)


×10^-10
Un-damped condition
damped condition
RMS acceleration (m/sec^2)
Natural frequency (hz)

(b)


×10^-12
Un-damped condtion
Damped condition
RMS acceleration (m/sec^2)
Natural frequency (Hz)

(c)

◂**Fig. 10** RMS acceleration response of different modes of chassis under sine wave condition: **a** first mode = 34.4 Hz; **b** second mode = 86.89 Hz; **c** third mode frequency = 174.8 Hz

$$\int_{t_0}^{t_1}\int_{x_0}^{x_1}\left[\frac{\partial\hat{L}}{\partial w}\delta w+\frac{\partial\hat{L}}{\partial\dot{w}}\delta\dot{w}+\frac{\partial\hat{L}}{\partial w'}\delta w'+\frac{\partial\hat{L}}{\partial w''}\delta w''\right]\mathrm{d}t\mathrm{d}x=0 \tag{53}$$

The next step is to transform the integrand in Eq. (52) into one containing only δw, i.e., one that is free of $\delta\dot{w},\delta w',\delta w'',\delta\dot{w}'$. This can be accomplished by integration by parts, with respect to both space and time.

$$\begin{aligned}&\int_0^l\frac{\partial\hat{L}}{\partial w}\delta w\Big|_{t_1}^{t_2}\mathrm{d}x+\int_{t_0}^{t_1}\left[\frac{\partial\hat{L}}{\partial w''}\delta w''+\left\{\frac{\partial\hat{L}}{\partial w'}-\frac{\partial}{\partial x}\left(\frac{\partial\hat{L}}{\partial w''}\right)\right\}\delta w\right]\Bigg|_0^l\mathrm{d}t\\&+\int_{t_0}^{t_1}\int_{x_0}^{x_1}\left[\frac{\partial\hat{L}}{\partial w}-\frac{\partial}{\partial t}\left(\frac{\partial\hat{L}}{\partial\dot{w}}\right)-\frac{\partial}{\partial x}\left(\frac{\partial\hat{L}}{\partial w'}\right)+\frac{\partial^2}{\partial x^2}\left(\frac{\partial\hat{L}}{\partial w''}\right)+\frac{\partial^2}{\partial x\partial t}\left(\frac{\partial\hat{L}}{\partial\dot{w}'}\right)\right]\delta w\,\mathrm{d}t\mathrm{d}x=0\end{aligned} \tag{54}$$

But the vertical displacement δw is arbitrary by definition, which implies that it can be assigned values at will provide these values are compatible with the system constraint, such as geometric conditions at the end points. Let $\delta w=\delta w'=0$ at $x=0$ and $x=l$. One may conclude that Eq. (54) can be satisfied for all values of δw in the open domain $0<x<l$.

Thus,

$$\frac{\partial\hat{L}}{\partial w}-\frac{\partial}{\partial t}\left(\frac{\partial\hat{L}}{\partial\dot{w}}\right)-\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial\hat{L}}{\partial w'}\right)+\frac{\partial^2}{\partial x^2}\left(\frac{\partial\hat{L}}{\partial w''}\right)+\frac{\partial^2}{\partial x\partial t}\left(\frac{\partial\hat{L}}{\partial\dot{w}'}\right)=0 \tag{55}$$

Equation (55) represents the *Lagrange differential equation of motion* for this continuous system and must be satisfied at every point of open domain $0<x<l$.

$$\frac{\partial\hat{L}}{\partial w''}\equiv 0\quad\text{or}\quad\delta w|_{x=0}\equiv 0 \tag{56a}$$

$$\left.\frac{\partial\hat{L}}{\partial w''}\right|_{x=l}\equiv 0\quad\text{or}\quad\delta w|_{x=l}\equiv 0 \tag{56b}$$

$$\left.\frac{\partial\hat{L}}{\partial w'}-\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial\hat{L}}{\partial w''}\right)\right|_{x=0}=0\quad\text{or}\quad\delta w|_{x=0}\equiv 0 \tag{57a}$$

$$\left.\frac{\partial\hat{L}}{\partial w'}-\frac{d}{dx}\left(\frac{\partial\hat{L}}{\partial w''}\right)\right|_{x=l}=0\quad\text{or}\quad\delta w|_{x=l}\equiv 0 \tag{57b}$$

Equations (56–57) are known as the boundary conditions and only two boundary conditions must be satisfied at either end, one from condition Eq. (56) and other from Eq. (57), which shows that clearly the selection is not arbitrary, but it must reflect physical conditions at two ends.

Appendix 2

In this formulation, the vertical translation of a uniform free–free beam is considered. If the internal damping of the material is involved in the system, then the strain displacement relationship or Hook's law equation can be written as Karnopp et al. (2012) and Hagedorn and DasGupta (2007).

$$\sigma(x,t)=E\varepsilon(x,t)+\mu_I\varepsilon(x,t) \tag{58}$$

The strain displacement relation at any height from the plane of the neutral fibers can be written as

$$\varepsilon(x,t)=\frac{zw''(x,t)}{\left[1+w'^2(x,t)\right]^{\frac{3}{2}}}\approx -zw''(x,t)\quad(\text{Assuming } w'<<1) \tag{59}$$

The bending moment at any section (from Fig. 15) can be written as

$$M(x,t)=-\int_{h/2}^{-h/2}z\sigma(x,t)\mathrm{d}A=-\left(E\varepsilon(x,t)+\mu_I\varepsilon(x,t)\right)\int_{h/2}^{-h/2}z^2\mathrm{d}A \tag{60}$$

$$\Rightarrow\quad -\left(EIw''(x,t)+\mu_I I\dot{w}''(x,t)\right) \tag{61}$$

where I is the moment of inertia cross section of beam about the neutral axis. The equation of the vertical force of an infinitesimal beam is

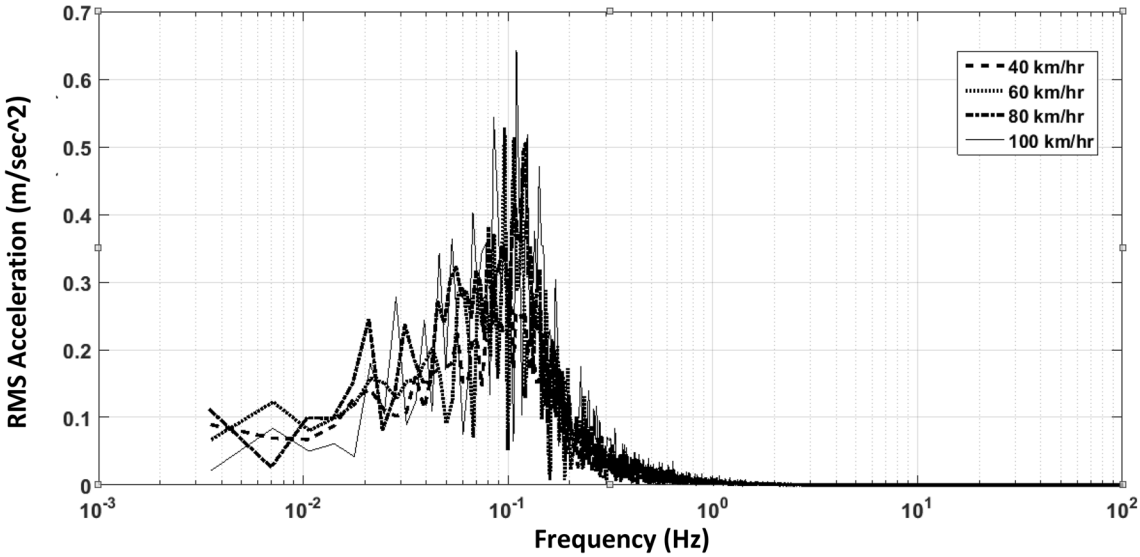
$$(\rho A\mathrm{d}x)\ddot{w}(x,t)=p(x,t)\mathrm{d}x+(V+\mathrm{d}V)\cos(\psi+\mathrm{d}\psi)-V\cos\psi$$

or

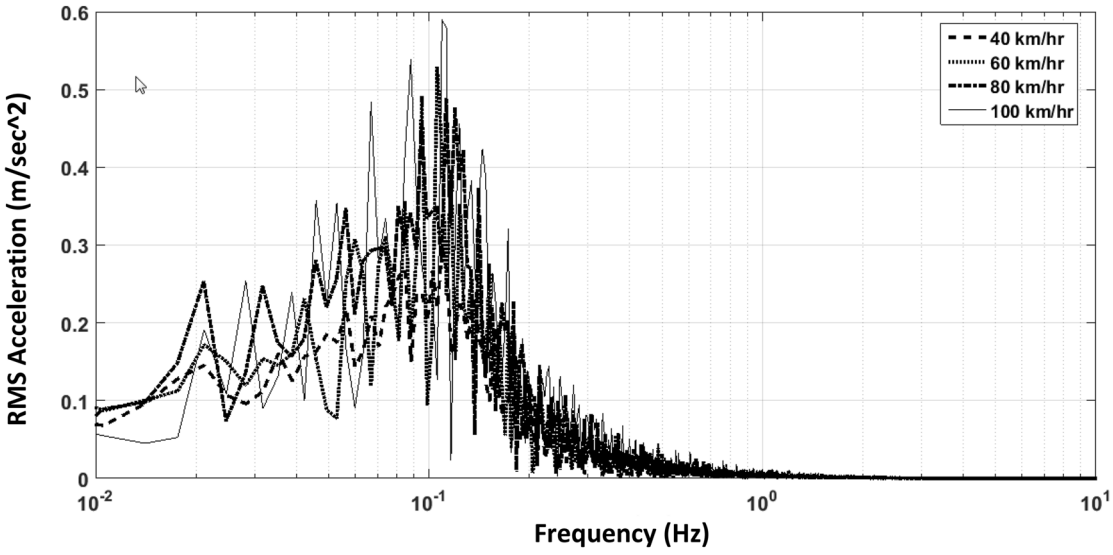
$$pA\ddot{w}(x,t)=p(x,t)+V' \tag{62}$$

The rotational dynamics of infinitesimal beam may be written as

$$(\rho I\mathrm{d}x)\ddot{\psi}=(M+\mathrm{d}M)-M(V+\mathrm{d}V)\frac{\mathrm{d}x}{2}+V\frac{\mathrm{d}x}{2}$$



(a)



(b)

Fig. 11 **a** Response of chassis acceleration at various random road inputs in H1 road as per ISO 8608. **b** Response of chassis acceleration consideration of internal damping at different random inputs in H1 road condition as per ISO 8608

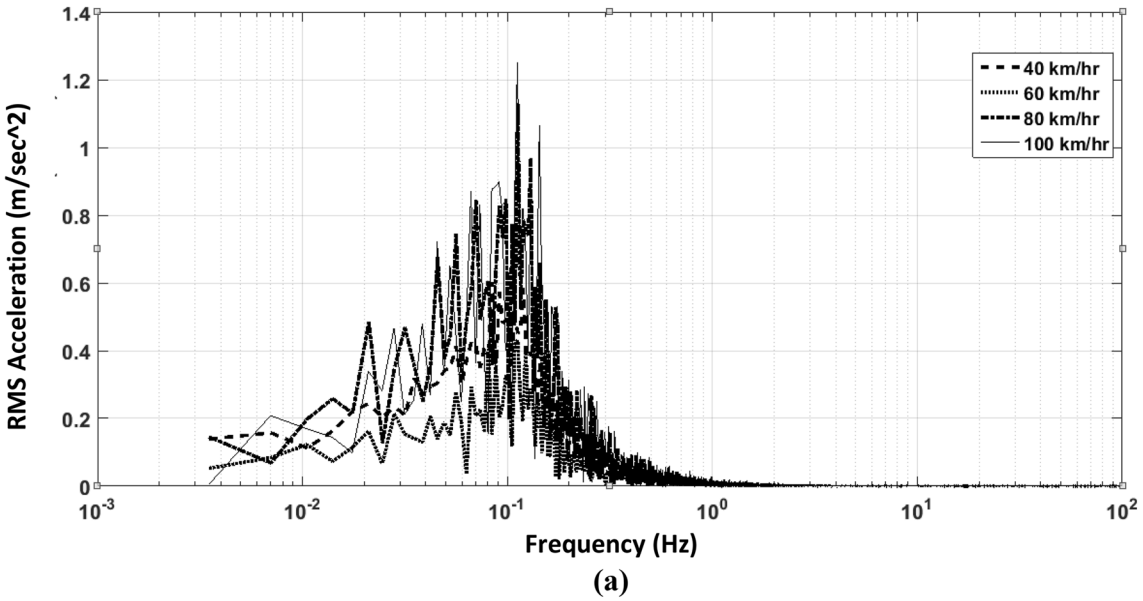
or

$$\rho I\ddot{\psi} = M' + V \tag{63}$$

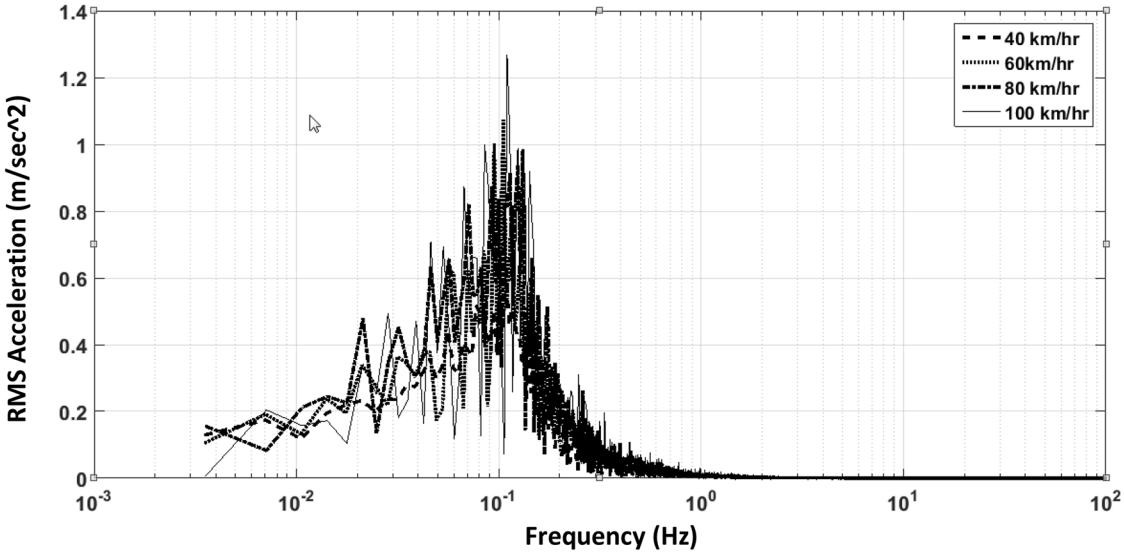
One may write $\psi = \frac{w'(x,t)}{1+w'^2(x,t)} \approx w'(x,t)$Thus, from Eqs. (62) and (63), one may have

$$\rho A\ddot{w}(x,t) - \rho I\ddot{w}''(x,t) + \mu_I I\dot{w}''''(x,t) + EIw''''(x,t) - p(x,t) = 0 \tag{64}$$

The third term of Eq. (64) represents the structural damping or internal damping of the beam.



(a)



(b)

Fig. 12 **a** Response of chassis acceleration at various random road inputs in H2 road as per ISO 8608. **b** Response of chassis acceleration consideration of internal damping at different random inputs in H2 road condition as per ISO 8608

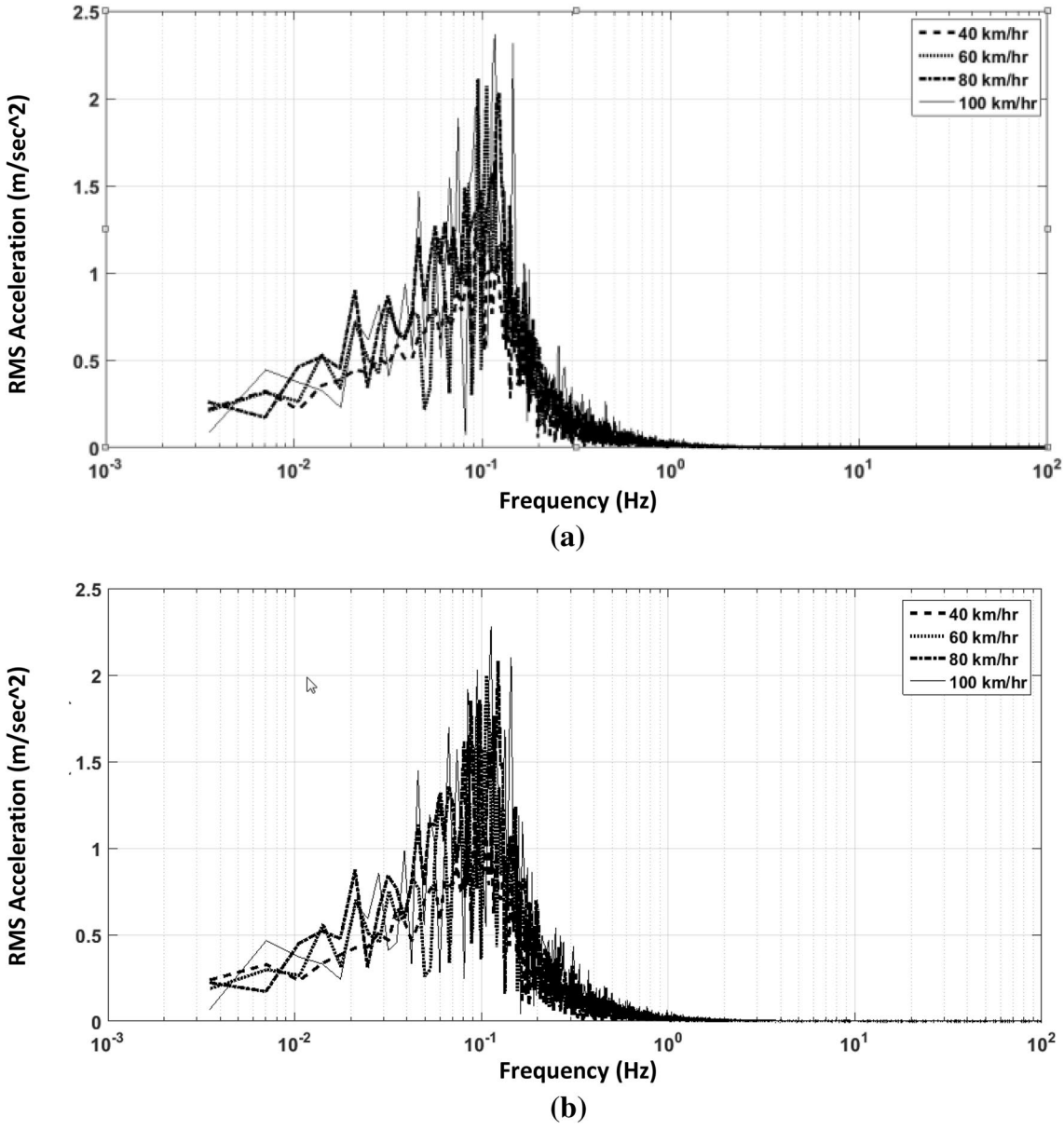


Fig. 13 **a** Response of chassis acceleration at various random road inputs in H3 road as per ISO 8608. **b** Response of chassis acceleration consideration of internal damping at different random inputs in H3 road condition as per ISO 8608

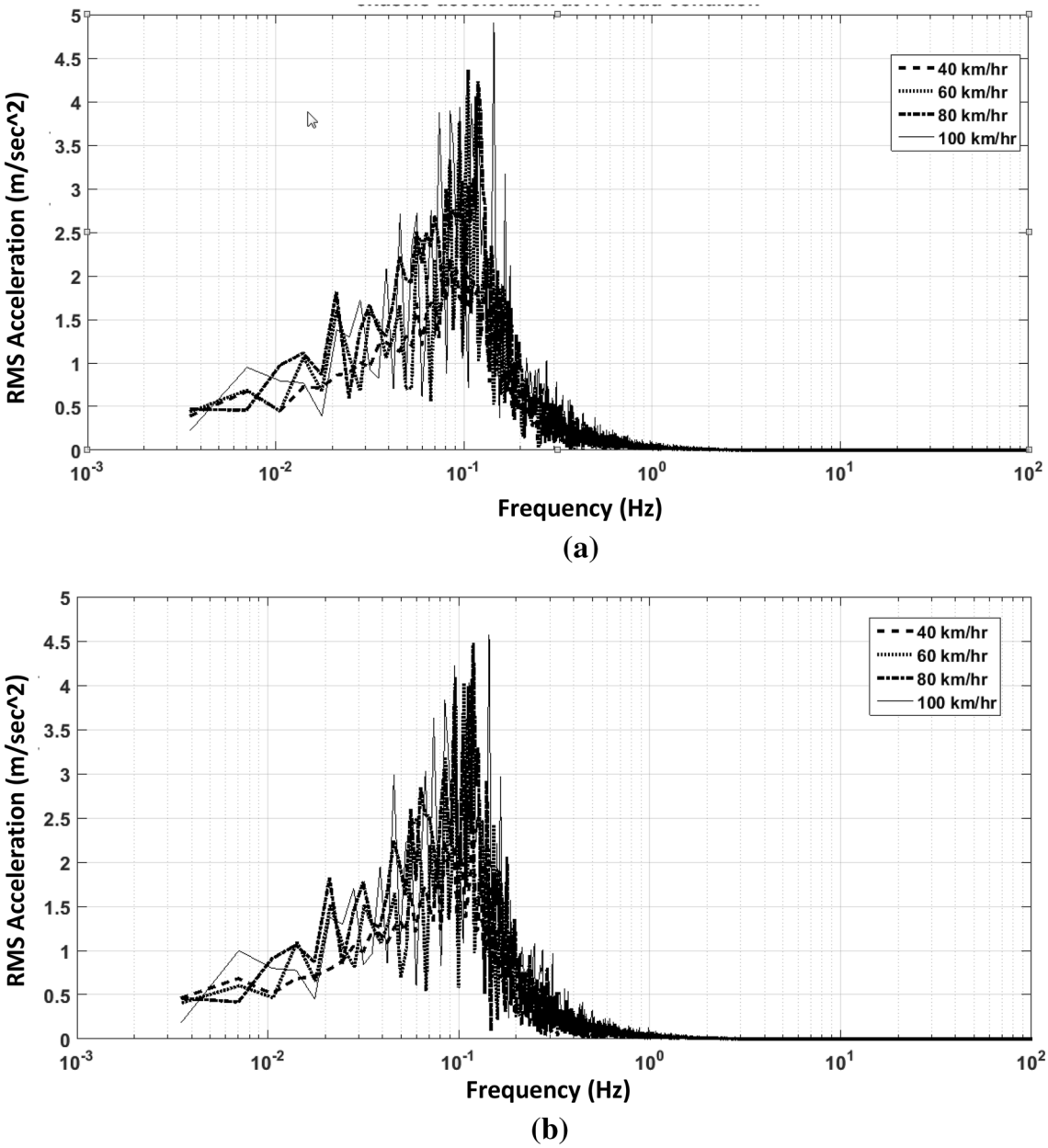


Fig. 14 **a** Response of chassis acceleration at various random road inputs in H4 road as per ISO 8608. **b** Response of chassis acceleration consideration of internal damping at different random inputs in H4 road condition as per ISO 8608

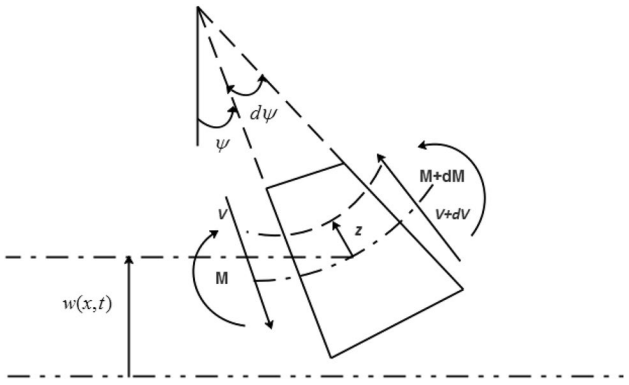


Fig. 15 Infinitesimal element of a deflected beam

References

Agostinacchio M, Ciampa D, Olita S (2014) The vibrations induced by surface irregularities in road pavements—a Matlab® approach. Eur Transp Res Rev 6(3):267–275

Gupta A, Rastogi V (2016) Effects of various road conditions on dynamic behavior of heavy road vehicle. Procedia Eng 144:1129–1137

Hagedorn P, DasGupta A (2007) Vibrations and waves in continuous mechanical systems. Wiley, New York

Hassani S (2013) Mathematical physics: a modern introduction to its foundations. Springer, Berlin

Huang CH, Zeng J (2010) Flexural vibration suppression of car body for high-speed passenger car based on constrained damping layers. J Traffic Transp Eng 10(1):36–41

Ibrahim IM (1996) Effect of frame flexibility on the ride. Vib Heavy Trucks 58(4):709–713

Karnopp DC, Margolis DL, Rosenberg RC (2012) System dynamics: modeling, simulation, and control of mechatronic systems. Wiley, New York

Kim JH, Yim HJ (1994) Influence of chassis flexibility on dynamic behavior of engine mount systems. SAE Paper, 942269

Kumar V, Rastogi V, Pathak P (2017) Simulation for whole-body vibration to assess ride comfort of a low–medium speed railway vehicle. Simulation 93(3):225–236

Margolis D, Edeal D (1989) Modeling and control of large flexible frame vehicles using bond graphs SAE Paper 892488. Society of Automotive Engineers, Warrendale

Meirovitch L (1980) Computational methods in structural dynamics. Springer, Netherlands

Mukherjee A, Karmakar R (2000) Modelling and simulation of engineering system through bond graph, Narosa Publishing House, New Delhi, reprinted by CRC Press for North America and by Alpha Science for Europe

Mukherjee A, Samantaray AK (2006) SYMBOLS-Shakti user's manual. High-Tech Consultants, STEP Indian Institute of Technology, Kharagpur

Rideout DG (2012) Simulating coupled longitudinal, pitch and bounce dynamics of trucks with flexible frames. Mod Mech Eng 2(4):176–189

Rideout D, Khan TM (2010) Flexible truck modelling and investigation of coupling between rigid and flexible dynamics. In: Proceeding of the ICBGM 2010, Orlando, USA

Tomioka T, Takigami T, Suzuki Y (2006) Numerical analysis of three-dimensional flexural vibration of railway vehicle car body. Veh Syst Dyn 44(Supp-1):272–285

Tomioka T, Takigami T, Aida K (2017) Experimental investigations on the damping effect due to passengers on flexural vibrations of railway vehicle car body and basic studies on the mimicry of the effect with simple substitution. Veh Syst Dyn 55(7):995–1011

Tomioko T, Takigami T (2015) Experimental and numerical study on the effect due to passengers on flexural vibrations in railway vehicle car bodies. J Sound Vib 343:1–19

Zhou J, Goodall A, Ren L, Zhang H (2009) Influences of car body vertical flexibility on ride quality of passenger railway vehicles. F J Rail and Rapid Transit 2009:223

Defence Technology xxx (xxxx) xxx

Contents lists available at ScienceDirect

Defence Technology

journal homepage: www.keaipublishing.com/en/journals/defence-technology

Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075

Husain Mehdi*, R.S. Mishra

Department of Mechanical Engineering, Delhi Technological University, Delhi, India

ARTICLE INFO

Article history:
Received 7 January 2020
Received in revised form
4 March 2020
Accepted 22 April 2020
Available online xxx

Keywords:
Residual stress
TIG+FSW
Heat transfer
Micro-hardness
Tensile strength

ABSTRACT

Tungsten inert gas (TIG) welding is the most commonly used joining process for aluminum alloy for AA6061 and AA7075 which are highly demanded in the aerospace engineering and the automobile sector, but there are some defects occur during TIG welding like micro-crack, coarse grain structure, and porosity. To improve these defects, the TIG welded joint is processed using friction stir processing (FSP).

This paper presents the effect of friction stir processing on TIG welding with filler ER4043 and ER 5356 for dissimilar aluminum alloy AA6061 and AA7075. The mechanical characterization, finite element formulation and mathematical equations of heat transfer of TIG + FSP welded joints are investigated using ANSYS Fluent software by adjusting process parameters of FSP. The results show that the maximum compressive residual stress 73 MPa was obtained at the fusion zone (FZ) of the TIG weldment with filler ER4043, whereas minimum compressive residual stress 37 MPa was obtained at stir zone (SZ) of the TIG + FSP with filler 5356. The maximum heat flux 5.33 × 106 W/m2 and temperature 515 °C have observed at tool rotation 1600 rpm with a feed rate of 63 mm/min. These results give a satisfactory measure of confidence in the fidelity of the simulation

© 2020 China Ordnance Society. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Friction stir welding (FSW) is a solid-state joining process that uses a non-consumable tool to join similar or dissimilar materials. The FSW does not involve the melting of parent material, which have many advantages on conventional fusion welding processes [1]. The FSW process involves the heat generation, which produced by pressure and friction between the tool and the workpiece. As a result of the generated heat, the weldment surfaces in contact soften become plasticized [2]. The modern development in the automobile sector and defense sector have been transformed from conventional materials to light material such as aluminum alloy. Due to excellent physical and mechanical properties of Al-alloy such as high corrosion resistance, low density, high strength to weight ratio and high thermal conductivity, it is mostly used for making various components such as military aircraft, rocket, and rocket launcher, axle shafts, rims, bumpers, and car bodies. FSW is used in the fabrication of automotive and shipbuilding components, aircraft structural components and reducing structural weight [3–7]. To study material flow and temperature distribution in friction stir welding of similar or dissimilar aluminum alloys many authors developed the CFD model to analyze the plastic flow, heat transfer and heat generation in the welding process [8–10], and the effect of tool profile on the material flow [11–13]. Due to low heat generation in the friction stir welding, the residual stresses are also low in the weldment. The transverse force of friction stir welding tools plays an important role in stress measurement [14–16]. The residual stresses in the weldment have a big impact on the performance of the welded structure. So far, information on the residual stress distribution of FSW has been limited [17]. The effects of tool feed rate on residual stresses of FSW of Al-alloy joints were studied with synchrotron X-ray measurement and analyze the residual stresses in longitudinal and transverse directions [16]. Many researchers have used finite element methods for the residual stress distribution in FSW. The computational fluid dynamic supports the heat transfer model and heat source of friction stir welding. The residual stress depends on the empathetic of heat transfer during FSW, although the correct prediction of temperature distribution depends on the heat source [18–20]. FSW can affect the performance of the residual stresses in

* Corresponding author.
E-mail address: husainmehdi4u@gmail.com (H. Mehdi).
Peer review under responsibility of China Ordnance Society

https://doi.org/10.1016/j.dt.2020.04.014
2214-9147/© 2020 China Ordnance Society. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Please cite this article as: Mehdi H, Mishra RS, Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, Defence Technology, https://doi.org/10.1016/j.dt.2020.04.014

Nomenclature

u	Plastic flow velocity	τ	Max. Shear stress
μ	Non-Newtonian viscosity	η	Mechanical efficiency
p	Pressure	ω	Angular velocity
ρ	Density	P_N	Normal pressure
U_1	Welding velocity	q_1	Total heat generation rate
z	Zener Hollomon parameter	T_a	Ambient temperature
a and n	Material constant	σ	5.67×10^{-16} J/(K^4 m^2 s)
Q	Temperature activation energy	ε	Convection heat transfer coefficient
Z	Temperature effective strain rate	h_b	Heat transfer coefficient
K	Thermal conductivity	σ_2	Friction stress
C_p	Specific heat	σ_1	Tensile stress
S_i	Heat generation rate	d	Main grain diameter
h_t	Emissivity at top surface	k	Constant
r	Radial distance from tool axis	R_S	Shoulder radius
A_r	Small area on tool pin	R_P	Tool pin radius
V	Control volume under A_r	h_{b0}	Heat transfer parameter of bottom.surface
θ	Angle of negative x-axis	μ_f	Spatially variable fraction slip.between the work piece and tool
		S_b	Heat generation rate due to plastic.deformation

the weldment and the processing parameters such as rotating tool may affect the development of residual stresses [21,22]. The mechanical properties, grain size are dependent on tool geometry, processing parameters and chemical composition of alloying elements. A new grain structure could be formed in friction stir welding by controlling the process parameter and heat input [23–27]. Comparison of residual stresses of TIG and FSW has been done for AA6061 and AA7075, and found that the longitudinal residual stress is larger than the traverse residual stress in the welded joint [28]. There are no successful investigations have been done to find out residual stress and heat transfer in TIG welded joint with friction stir processing. The small plate having the dimension of 150 mm × 25 mm × 0.8 mm, have been investigated, as a result, the properties at the beginning and end of the welded joints can be significantly different. The tensile strength first increases and then decreases with the growth of rotational speed. Three layers i.e. top, middle, and bottom, have considered measuring the microhardness of the welded joint. The top layer of the weld has a higher recrystallization degree and dynamic recovery and the bottom layer is strongly stirred under mechanical action of the tool pin such that each layer of welded joint shows the different mechanical properties [29]. A new welding approach of TIG+FSP was successfully applied to the AA2024 to enhance the mechanical properties of the TIG welded joints. They conclude that the defects and porosities in the TIG welded joints are completely reduced by the FSP process and modified the microstructure and mechanical properties of the TIG+FSP welded joint [30]. The application of FSP on the TIG welded joint improved the ductility and tensile strength of the FSW and TIG welded joint of AA5083-H111 and results revealed best-reduced dimple size [31].

In this work, the effect of FSP on TIG welding for dissimilar Al-alloy AA6061 and AA7075 with filler wire ER4043 and ER5356 was carried out and analyze the mechanical properties, residual stress distribution and heat transfer in (TIG+FSP) welded joint. The present study was focus on the improvement of mechanical properties and the residual stress distribution of the TIG-welded joint with different filler wire (ER4043 and ER5356) by using a friction stir processing approach. This topic is carefully selected after noticing that there is little literature review available on the TIG+FSP welding approach. The outcomes of this paper will give the new approach for enhancing the mechanical properties and heat transfer of TIG+FSP welded joints which play a huge role in welded structures.

2. Materials and methods

Two AA6061 and AA7075 aluminum alloy plates having dimensions (150 mm × 40 mm × 6.2 mm) were welded together using the TIG+FSP welding approach. V butt joints of 45° were prepared for TIG welding with the help of milling cutter and power hacksaw. There are two types of samples were prepared for TIG welded joint. In the first sample, ER4043 filler wire was used and for the second sample, ER 5356 filler wire was used. The chemical composition of filler rod and base materials are shown in Table 1. The processing parameters and filler wire are strappingly affected by the mechanical properties of the welded joint. In this work, the filler wire of ER4043 and ER5356 of diameter 2.4 mm were used to join a single V groove plate at voltage 22 V and current 50 amp for all welding processes. Argon (25 L/min) is used as a shielding gas and travel speed 3.6 mm/s to fabricate the joint.

After TIG welding, the FSP technique was used on TIG weldment to improve the welding quality and mechanical properties. The non-consumable H13 steel tool with pin diameter, shoulder diameter and a pin length of 3 mm, 19.5 mm, and 5.5 mm respectively was used. 0° tilt angle was used during friction stir processing on the AA6061 and AA7075 plate at a rotational speed of 800 rpm−1600 rpm and a traverse speed of 63 mm/min is shown in Fig. 1(a). The tool nomenclature is shown in Fig. 1(b). The processing parameters of TIG welding and friction stir processing are shown in Table 2 and Table 3.

Fig. 2 demonstrate the dimension of the tensile sub test specimen as per ASTM E8. The tensile test specimens were sliced and machined from the welded joint using a milling cutter and shaper machine. Single-pass FSP is used to fabricate TIG+FSP welded joint. Three sub tensile specimens were tested on a UTM machine at

Table 1
Chemical composition of filler wire and parent material.

Material	Si	Cu	Fe	Zn	Mg	Mn	Cr	Ti	Al
ER 4043	5.300	0.025	0.80	0.10	0.05	–	–	0.02	Bal.
ER 5356	0.025	0.015	0.15	0.01	4.90	0.14	0.11	0.09	Bal.
AA7075	0.580	1.200	0.35	5.10	2.10	0.12	0.20	0.05	Bal.
AA6061	0.400–0.800	0.200	0.70	0.25	0.80	0.10	0.20	0.15	Bal.

Please cite this article as: Mehdi H, Mishra RS, Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, Defence Technology, https://doi.org/10.1016/j.dt.2020.04.014

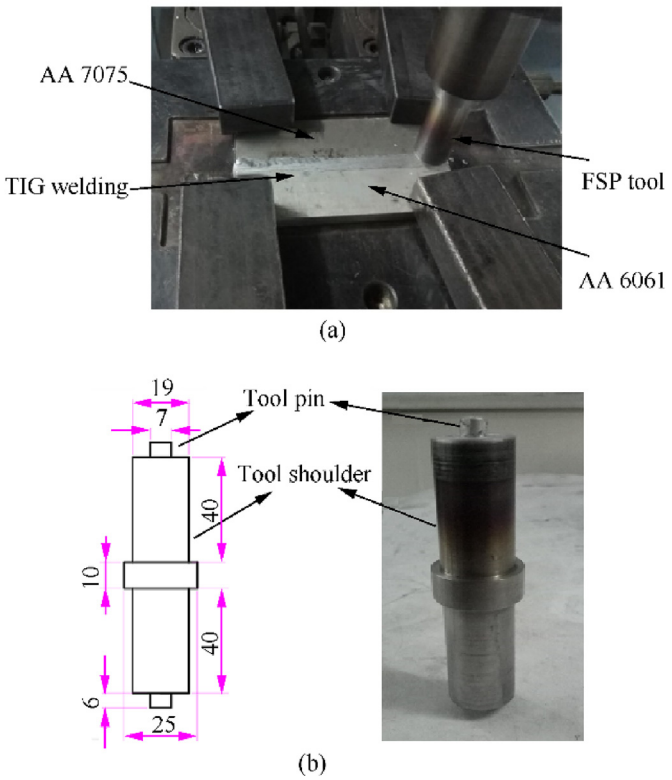


Fig. 1. (a) FSP process after TIG welding, (b) Friction stir processing tool of H13 tool steel.

room temperature and the average of these three results is taken. Vicker microhardness machine was used for measuring the hardness across the welded joints with a load of 100 g and dwell time 30 s. Microstructure observations were carried out using a scanning electron microscope (SEM) for TIG and TIG+FSP welded joints.

3. Numerical modeling [10]

For incompressible flow, the continuity equation in the direction of x, y and z is given by

$$\frac{\partial u_i}{\partial x_i} = 0 \tag{1}$$

The momentum conservation equation of the heat source is given as

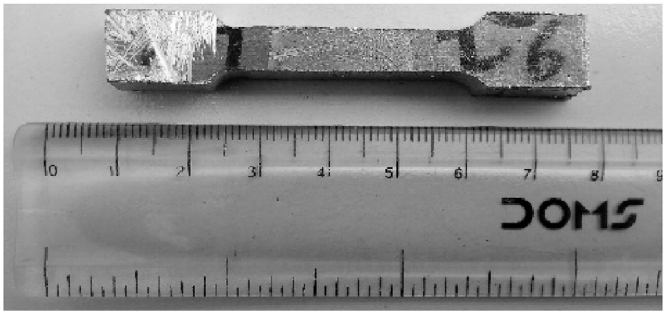


Fig. 2. Dimension of tensile test specimen as per ASTM E8.

$$\rho \frac{\partial u_I\, u_j}{\partial x_i} = -\frac{\partial P}{\partial x_j} + \frac{\partial}{\partial x_I}\left(\mu \frac{\partial U_j}{\partial x_i} + \mu \frac{\partial U_i}{\partial x_j}\right) - \rho\, U_1 \left(\frac{\partial U_j}{\partial x_I}\right) \tag{2}$$

The flow stress can be calculated as

$$\sigma_e = \frac{1}{\alpha} \sinh^{-1}\left[\left(\frac{z}{A}\right)^{1/n}\right] \tag{3}$$

$$A = 1.80 \times 106 + 1.74 \times 10^8 (\%C) - 6.5 \times 10^7 (\%C)^2 \tag{4}$$

$$\alpha = 1.07 + 1.70 \times 10^{-4} T - 2.8 \times 10^{-7} T^2 \tag{5}$$

$$n = 0.2 + 3.96 \times 10^{-4} T \tag{6}$$

$$z = \varepsilon \exp\left(\frac{Q}{RT}\right) \tag{7}$$

The effective strain may be written as

$$\varepsilon = \left(\frac{2}{3}\, \varepsilon_{ij} \varepsilon_{ij}\right)^{1/2} \tag{8}$$

Where ε_{ij} is given by

$$\varepsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \tag{9}$$

So, the viscosity can be calculated as

$$\mu = \frac{\sigma_e}{3\varepsilon} \tag{10}$$

The thermal energy conservation equation at steady state is given by

Table 2
Processing parameters of TIG welded joint [32].

Working Parameters	Shielding Gas/(L·min^{-1})	Current/A	Electrode Diameter/mm	Voltage/V	Welding Speed/(mm·s^{-1})
Description	25 (Argon)	50	3.2	22	3.6

Table 3
Processing parameters of friction stir processing [33].

Sample No	Current	Welding Speed/(mm·min^{-1})	Tool rotational speed/rpm	Frequency	Tilt angle/(°)
1	6.9	63	800	28.52	0
2			1000	33.36	
3			1250	40.00	
4			1600	52.61	

Please cite this article as: Mehdi H, Mishra RS, Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, Defence Technology, https://doi.org/10.1016/j.dt.2020.04.014

$$\rho C_P \frac{\partial (u_i T)}{\partial x_i} = -\rho C_P U_1 \frac{\partial T}{\partial x_i} + \frac{\partial}{\partial x_i}\left(k \frac{\partial T}{\partial x_i}\right) + S_i + S_b \tag{11}$$

Then S_i may be written as

$$S_i = \left[(1-\delta)\eta\tau + \delta\mu_f P_N\right]\ (\omega_r - U_1 \sin\theta)\ \frac{A_r}{V} \tag{12}$$

During friction stir welding, mixing is not atomic, only grains structure is deformed not mixing of atoms.

$$\phi = 2\left[\left(\frac{\partial u_1}{\partial x_1}\right)^2 + \left(\frac{\partial u_2}{\partial x_2}\right)^2 + \left(\frac{\partial u_3}{\partial x_3}\right)^2\right] + \left(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1}\right)^2 + \left(\frac{\partial u_1}{\partial x_3} + \frac{\partial u_3}{\partial x_1}\right)^2 + \left(\frac{\partial u_3}{\partial x_2} + \frac{\partial u_2}{\partial x_3}\right)^2 \tag{13}$$

3.1. Boundary conditions

When the work-piece top surface away from the tool shoulder edge, the boundary condition for heat exchange involves both convection and radiation heat transfer. However, at the sides and the bottom surface of the work-piece, the boundary condition for heat exchange involves only convection heat transfer. The heat generation rate at the interface between tool and the work-piece can be given be given by

$$f = \frac{J_w}{J_T} = \frac{\sqrt{(k_\rho C_P)_W}}{\sqrt{(k_\rho C_P)_T}} \tag{14}$$

Where T and W denote the tool and work piece respectively. The heat flux continuity on the shoulder interface yields

$$k \left.\frac{\partial T}{\partial z}\right|_{top} = \frac{J_w}{J_w + J_T}\ q1 \quad R_P \le r \le R_s \tag{15}$$

Where J_w and J_T are the heat conducted to the work-piece and tool respectively

$$J = \sqrt{k\rho\ C_p} \tag{16}$$

Where C_p is heat capacity, ρ and K are density and thermal conductivity respectively.

The total heat generation rate may be written as

$$q_1 = [\eta(1-\delta)\tau + \delta\mu f.P_H]\ (\omega r - U_1 \sin\theta) \tag{17}$$

The heat transfer coefficient may be calculated as.

$$k \left.\frac{\partial T}{\partial z}\right|_{bottom} = h_b (T - T_a) \tag{18}$$

Value of h_b can be calculated as

$$h_b = h_{b0}(T - T_a)^{0.25} \tag{19}$$

The heat transfer due to radiation and convection is written as

$$k \left.\frac{\partial T}{\partial z}\right|_{top} = h_t (T - T_a) + \sigma\varepsilon^1 \left(T^4 - T_a^4\right) \tag{20}$$

In this model, the computational region is considered as a single-phase visco-plastic non-Newtonian fluid and the FSW tool is considered as rotating in a fixed position as shown in Fig. 3. The material flows into the computational domain from the inlet velocity and out from outlet velocity at the welding speed (traverse speed). Top, bottom and side surfaces of the work-pieces are considered equivalent to the wall surface, having the same velocity as welding speed but opposite in direction.

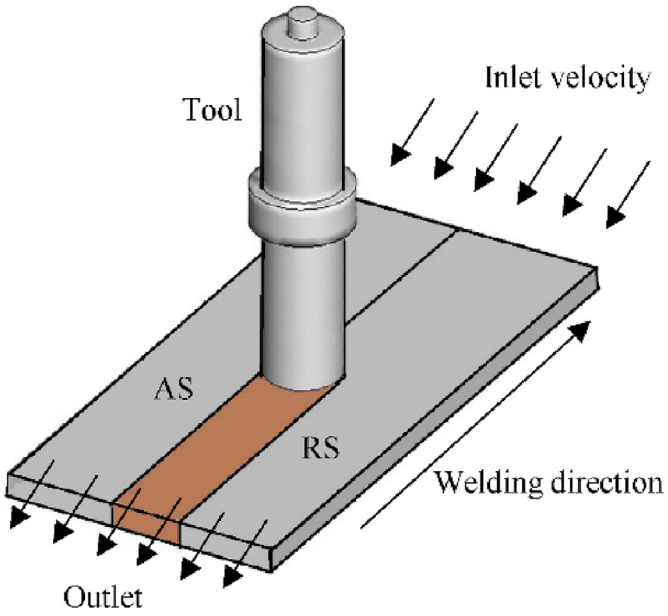


Fig. 3. Boundary condition.

Velocity at the tool pin periphery have been defined in terms of tool translation velocity and the tool pin angular velocity

$$\left.\begin{aligned} u &= (1-\delta)(\omega R_P \sin\theta - U_1) \\ v &= (1-\delta)\omega R_P \cos\theta \\ w &= k\ \frac{\omega}{2\pi} R_P \end{aligned}\right\},\ R_P \le r \le R_S \tag{21}$$

Similarly, at the shoulder contact, the velocity boundary condition may be written as

$$\left.\begin{aligned} u &= (1-\delta)(\omega r \sin\theta - U_1) \\ v &= (1-\delta)\omega r \cos\theta \end{aligned}\right\},\ R_P \le r \le R_S \tag{22}$$

3.2. Material properties

The density of AA7075 and AA6061 is taken as constant, equals to 2810 kg/m^3 and 2700 kg/m^3 respectively, while the specific heat and thermal conductivity are considered as temperature dependent properties [34] as given below

$$k = 25.2 + 3.98 \times 10^{-1} T + 7.36 \times 10 - 6T^2 - 2.25 \times 10^{-7} T^3$$
$$C_P = 929 - 6.27 \times 10^{-1} T + 1.48 \times 10 - 3T^2 - 4.33 \times 10^{-8} T^3$$

3.3. Residual stress measurement by cosα method

A mini portable X-ray diffraction apparatus (pulstec μ-X360) used for analyzing the residual stresses in welded joint of AA6061 and AA7075 samples. The X-ray incident angle was set 35° and ±5° oscillation was applied. The X-ray incident time was 4–5 min throughout this process for each sample. Under these conditions, a diffracted beam from the workpiece surface has captured the

Please cite this article as: Mehdi H, Mishra RS, Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, Defence Technology, https://doi.org/10.1016/j.dt.2020.04.014

images of the welded plates at 50 μm resolution, the size of the beam spot is approximately 2.5 mm for 1.2 mm pinhole collimator.

For residual stress determination, the cosα method was described [35]. The translation from the diffractometer space to the sample inherently more complex due to the 2D planar geometry of the measurement and can be expressed as

$$q_s = \begin{bmatrix} \cos\eta \sin\psi_0 + \sin\eta \cos\psi_0 \cos\alpha \\ \cos\eta \sin\psi_0 \sin\varphi_0 + \sin\eta \cos\psi_0 \sin\varphi_0 cos\alpha + \sin\eta \cos\varphi_0 \sin\alpha \\ \cos\eta \cos\psi_0 + \sin\eta \sin\psi_0 cos\alpha \end{bmatrix} \tag{23}$$

The strain projection along (η, α) coordinates can be written as in terms of scattering vector and strain component as

$$\begin{aligned} \varepsilon_\alpha &= q_i \cdot q_j \cdot \varepsilon_{ij} \\ \varepsilon_{ij} &= \frac{1+\nu}{E}\sigma_{ij} - \delta_{ij}\frac{\nu}{E}\sigma_{kk} \end{aligned} \tag{24}$$

So, the strain projection may be written as

$$\varepsilon_\alpha = \frac{1+\nu}{E} q_i q_j \sigma_{ij} - \frac{\nu}{E}\sigma_{kk} \tag{25}$$

Now, defining two parameters a_1 and a_2 for linear determination of σ_{11} and σ_{22}

$$a_1 = \frac{1}{2}[(\varepsilon_\alpha - \varepsilon_{\pi+\alpha}) + (\varepsilon_{-\alpha} - \varepsilon_{\pi-\alpha})] \tag{26}$$

$$a_2 = \frac{1}{2}[(\varepsilon_\alpha - \varepsilon_{\pi+\alpha}) - (\varepsilon_{-\alpha} - \varepsilon_{\pi-\alpha})] \tag{27}$$

After re-expressing of Eq. (26) and Eq. (27) to lead the final relationship for this method.

$$\begin{aligned} a_1 = \frac{1+\nu}{E}\sin\psi_0 \sin 2\eta \cos\alpha [\sigma_{11}(1+\cos 2\varphi_0) \\ + \sigma_{22}(1-\cos 2\varphi_0) + 2\sigma_{12}(\sin 2\varphi_0)] \end{aligned} \tag{28}$$

$$\begin{aligned} a_2 = \frac{1+\nu}{E}\sin\psi_0 \sin 2\eta \sin\alpha [\sigma_{22} \sin 2\varphi_0 - \sigma_{11} \sin 2\varphi_0 \\ + 2\sigma_{12} \cos 2\varphi_0] \end{aligned} \tag{29}$$

at $\varphi_0 = 0$, the above equations will be

$$a_1 = \sigma_{11}\frac{1+\nu}{E}\sin 2\psi_0 \sin 2\eta \cos\alpha \tag{30}$$

$$a_2 = 2\sigma_{12}\frac{1+\nu}{E}\sin\psi_0 \sin 2\eta \sin\alpha \tag{31}$$

Thus the term cosα in Eq. (30) is the origin of the name for this method.

The value of stresses after re-expression maybe written as

$$\sigma_{11} = -\frac{E}{(1+\nu)} \frac{1}{\sin 2\psi_0 \sin 2\eta} \tag{32}$$

$$\sigma_{12} = -\frac{E}{2(1+\nu)} \frac{1}{\sin 2\psi_0 \sin^2\eta} \tag{33}$$

4. Results and discussion

4.1. Tensile strength

In order to analyze the effect of friction stir processing on the TIG-welded joint, the tensile load applied on TIG and TIG+FSP welded joint by universal testing machine (UTM) at room temperature and fractured surfaces obtained from the tensile test were compared with one another. Three test specimens were tested at each condition and the average of these three results are presented in Tables 4–10. The tensile strength of the weldment made by TIG and TIG+FSP techniques are compared. Due to the absence of porosity, small grain size and presence of extra material in the welded region with filler wire, the tensile strength of TIG+FSP obtained higher value than the conventional TIG joining process. Filler ER5356 present in the welded joint makes more compact pressure leading to increase bond strength instead of filler ER4043. The grain size in the fusion zone for ER5356 is smaller than the ER4043. The tensile strength with filler ER5356 was obtained higher than ER4043 [36]. According to Hall Petch equation $\sigma_1 = \sigma_i + kd^{(-1/2)}$, the tensile strength is inversely proportional to the grain size [37]. The tensile strength of the TIG-welded joint using filler 4043 and 5356 was calculated 158.6 MPa and 172.2 MPa respectively which is less than the TIG+FSP as shown in Table 4. The tensile strength of the TIG+FSP welded joint increased with increasing tool rotational speed. The highest tensile strength (233.17 MPa) was observed in TIG+FSP with filler ER5356 at tool rotational 1600 rpm with feed rate 63 mm/min. Three statistical parameters are investigated i.e., standard deviation (SD), standard error (SE) and 95% confidence interval for TIG+FSP welded joint with filler ER4043 and ER5356 as shown in Tables 5–10. The standard deviation (SD) provides the deviance of the experimental

Table 4
Mechanical properties of TIG welded joint.

Processing Parameters	Tensile Strength/MPa	Hardness/HV	Residual stress/MPa
TIG with filler ER4043	158.6	66	71
TIG with filler ER 5356	172.2	74	67

Table 5
Tensile strength of TIG+FSP welded joint with filler ER4043.

Processing Parameters	Tensile Strength/MPa			Mean Tensile Strength/MPa	Standard Deviation	Standard error	95% confidence Interval	
	Specimen-1	Specimen-2	Specimen-3				Minimum	Maximum
TIG+FSP_800	178.4	178.0	178.8	178.40	0.40	0.23	177.95	178.85
TIG+FSP_1000	189.0	189.4	190.4	189.60	0.72	0.42	188.78	190.42
TIG+FSP_1250	204.5	204.8	205.0	204.77	0.25	0.15	204.48	205.05
TIG+FSP_1600	219.0	219.6	220.0	219.53	0.50	0.29	218.96	220.10

Please cite this article as: Mehdi H, Mishra RS, Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, Defence Technology, https://doi.org/10.1016/j.dt.2020.04.014

Table 6
Hardness of TIG+FSP welded joint with filler ER4043.

Processing Parameters	Hardness/HV			Mean Hardness /HV	Standard Deviation	Standard error	95% confidence Interval	
	Specimen-1	Specimen-2	Specimen-3				Minimum	Maximum
TIG+FSP_800	82.5	82.0	83.5	82.6	0.76	0.44	81.80	83.53
TIG+FSP_1000	89.0	90.0	89.0	89.3	0.58	0.33	88.68	89.99
TIG+FSP_1250	92.6	92.0	92.0	92.2	0.35	0.20	91.81	92.59
TIG+FSP_1600	96.0	97.0	96.5	96.5	0.50	0.29	95.93	97.07

Table 7
Residual Stress of TIG+FSP welded joint with filler ER4043.

Processing Parameters	Residual Stress/MPa			Mean Residual Stress/MPa	Standard Deviation	Standard error	95% confidence Interval	
	Specimen-1	Specimen-2	Specimen-3				Minimum	Maximum
TIG+FSP_800	63	63	64	58.67	0.58	0.33	58.01	59.32
TIG+FSP_1000	59	58	57	56.33	1.00	0.58	55.20	57.46
TIG+FSP_1250	49	50	49	51.00	0.58	0.33	50.35	51.65
TIG+FSP_1600	43	43	45	44.67	1.15	0.67	43.36	45.97

Table 8
Tensile strength of TIG+FSP welded joint with filler ER5356.

Processing Parameters	Tensile Strength/MPa			Mean Tensile Strength/MPa	Standard Deviation	Standard error	95% confidence Interval	
	Specimen-1	Specimen-2	Specimen-3				Minimum	Maximum
TIG+FSP_800	187.5	187.0	187.6	187.37	0.32	0.19	187.00	187.73
TIG+FSP_1000	194.2	194.8	195.4	194.80	0.60	0.35	194.12	195.48
TIG+FSP_1250	218.4	219.4	218.8	218.87	0.50	0.29	218.30	219.44
TIG+FSP_1600	233.0	233.5	233.0	233.17	0.29	0.17	232.84	233.49

Table 9
Hardness of TIG+FSP welded joint with filler ER5356.

Processing Parameters	Hardness/HV			Mean Hardness /HV	Standard Deviation	Standard error	95% confidence Interval	
	Specimen-1	Specimen-2	Specimen-3				Minimum	Maximum
TIG+FSP_800	89.5	89.0	89.0	89.2	0.29	0.17	88.84	89.49
TIG+FSP_1000	93.0	94.5	94.0	93.8	0.76	0.44	92.97	94.70
TIG+FSP_1250	99.5	99.0	100.0	99.5	0.50	0.29	98.93	100.07
TIG+FSP_1600	102.0	102.0	102.5	102.2	0.29	0.17	101.84	102.49

Table 10
Residual Stress of TIG+FSP welded joint with filler ER5356.

Processing Parameters	Residual Stress/MPa			Mean Residual Stress/MPa	Standard Deviation	Standard error	95% confidence Interval	
	Specimen-1	Specimen-2	Specimen-3				Minimum	Maximum
TIG+FSP_800	58	58	60	58.7	1.15	0.67	57.36	59.97
TIG+FSP_1000	52	53	52	52.3	0.58	0.33	51.68	52.99
TIG+FSP_1250	39	38	40	39.0	1.00	0.58	37.87	40.13
TIG+FSP_1600	32	31	32	31.7	0.58	0.33	31.01	32.32

values from the mean it may be calculated as $SD = [\Sigma(X_i - M)^2/(N-1)]^{1/2}$ and the standard error is a quantity used to measure how to close the prediction values to the experimental values. *SE* is calculated as $SE = SD/N^{1/2}$, where N= No of observation and M = mean.

The confidence interval has shown that tensile strength and hardness increased with increasing tool rotation, whereas residual stress decreased with increasing tool rotation.

4.2. Micro-hardness

The micro-hardness distribution of the TIG-welded joint of AA6061 and AA7075 with different filler wire and TIG+FSP weldments with different processing parameters were analyzed by Vickers hardness testing results are shown in Fig. 4. Vicker microhardness machine was used for measuring the hardness across the welded joints with a load of 100 g and dwell time 30 s. The hardness distributions are asymmetrical in the weld center due to the microstructure of the advancing side and retreating side introduced by unsteady plastic flow from the base metal to the weld center [38–40]. The hardness slopes downward from the base metal to the welded region. The hardness fluctuates largely on the advancing side while the hardness holds steady in the retreating side [41].

The alloying elements such as Si and Mg existing in the weld center make precipitation reaction and form a strong precipitate of Mg_2Si to give a higher strength as shown in Fig. 5(a). Fig. 4 shows the micro-hardness profile at the heat-affected zone of aluminum alloy 6061 and 7075. It showed a significant difference, where the

Please cite this article as: Mehdi H, Mishra RS, Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, Defence Technology, https://doi.org/10.1016/j.dt.2020.04.014

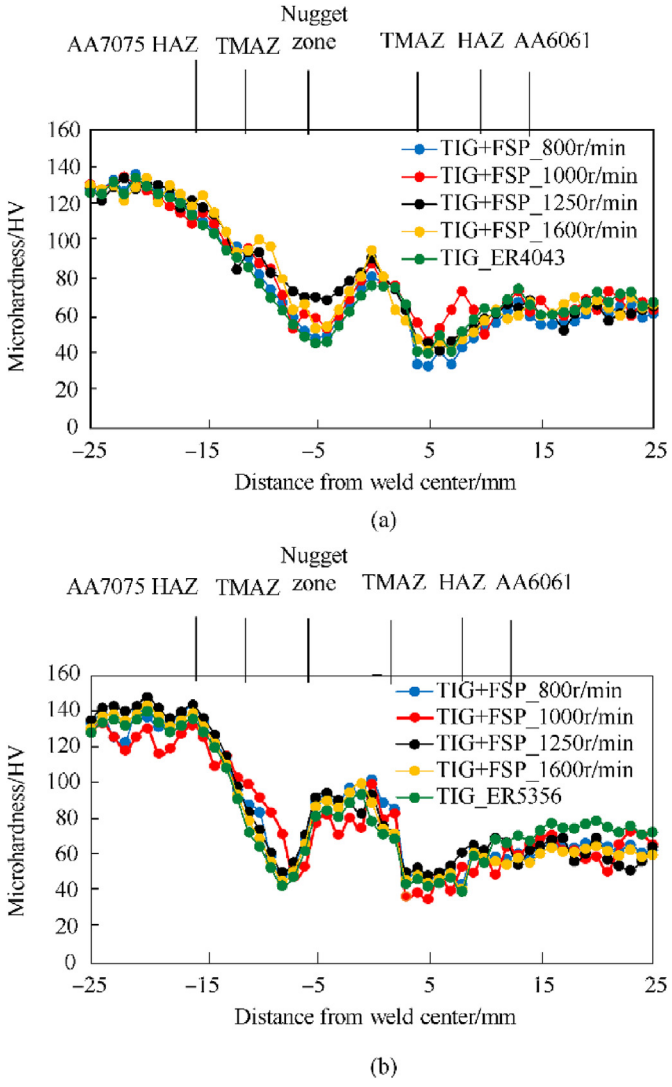


Fig. 4. Comparison of micro-hardness of TIG and TIG+FSP, (a) filler wire ER4043, (b) filler wire ER5356.

welded joint using filler ER4043 showed a lower average hardness value compared to filler ER5356. The hardness value at the center of weldment for filler ER4043 and ER5356 are 96 HV and 102 HV respectively, during welding, the filler wire ER4043 shows the columnar grains while fine equiaxed grains are found in ER5356. It may be analyzed that the fine equiaxed grains improve mechanical properties than the columnar grains [42].

4.3. XRD and EDX analysis

X-ray diffraction (XRD) analysis has been taken on transverse cross-sections of dissimilar TIG and TIG+FSP welded joint of AA6061 and AA7075 with filler ER4043 and ER5356. The corresponding pattern for the welded joints are shown in Fig. 5 and found four major phases i.e. Al, Al_2CuMg, $MgZn_2$, and Mg_2Si. Magnesium (Mg) and Silicon (Si) elements were found in the weldment besides the aluminum (Al), it is found that Mg and Si created the phase after the precipitation reaction in the weldment. The very high intensity was found from aluminum, because of the fragmentation of precipitates the intensity of $MgZn_2$ was decreases after friction stir processing on TIG welded joint with filler ER5356. The alloying elements such as Si and Mg existing in the weld center make precipitation reaction and form a strong precipitate of Mg_2Si to give a higher strength. The same phase of Al_2CuMg was detected in both the cases (filler ER4043 and ER 5356).

The hardness of the FSP joints is based on boundary energy, brittle intermetallic formation, strain hardening and precipitates formation in the joint. Fine recrystallized grains and increase of grain boundaries in the (SZ) of AA6061 and AA7075 joints predict higher hardness. Because of fine precipitates and fine grains structure, the stir zone is associated with plastic deformation and high temperature, due to precipitates formation at high temperature along the grain boundaries the hardness value of TIG+FSP at 1600 rpm with filler ER5356 recorded higher values than the TIG welded joints. The energy-dispersive X-ray spectroscopy (EDX/EDS) of the TIG-welded joint with filler ER4043 and ER 5356 welded joint has been analyzed. It was found that the atomic percentage of Si and Mg in the welded joint is higher than the other elements. Fig. 5(b) and (c) illustrates the EDX image and percentage of element concentration at fusion zone (FZ) in the TIG welded joint with filler ER4043 and ER5356. Zinc (Zn), Magnesium (Mg) and Silicon (Si) elements were found in the weldment besides the aluminum (Al), it is found that Mg, Zn, and Si created the phases after the precipitation reaction in the weldment.

4.4. Residual stresses measurement

A mini portable X-ray diffraction apparatus (Pulstec μ-X360) at Delhi technological university, Delhi, India was used to determine the residual stresses in weldment of AA6061 and AA7075 by the cosα method. The cosα method was introduced in japan for residual stress analysis in 1978 [43]. This method is allowed the stress analysis by capturing the results by a single incident X-ray beam via a 2D detector. It shows the peak center is very stable throughout the welding. Debye ring and distortion ring is obtained by using cosα method. An investigation has been done to analyze the residual stress. The 3D Debye ring and distortion ring of welded samples are shown in Fig. 7. The X-ray incident angle was set 35° and ±5° oscillation was applied. The variation of residual stresses in the transverse direction of weldment with filler ER4043 and ER5356 are shown in Fig. 6(a) and (b). Residual stresses (compressive or tensile) will influence the mechanical behavior of the welded joint. It can reduce brittle fracture strength, buckling strength and cracking in the weldment. Residual stress is also influenced by the prediction of brittle failure and affect the lifetime prediction of the component [44–47]. Residual stress contributes both negative and positive effects to the weldment, generally, the tensile residual stress leads to a negative effect on the weldment [48]. When the fixtures are unconstrained and the welded plate temperature is reduced to room temperature, the material in the nugget zone (NZ) tends to recover, but the material in the heat-affected zone (HAZ) has smaller deformation and will prevent the recovery process in NZ, so the maximum residual stress occurs in the boundaries of the HAZ with a minimum in the NZ. The deformation on the advancing side and retreating side differ, causing the recovery processes to differ too. So the residual stress cannot be symmetry to the welding line [49]. The base material AA7075 on LHS of the weldment shows a minimum compressive residual stress, however, the residual stress gradually increases from the base material to towards the weldment and then decreases till second base material AA6061. The maximum residual stress occurs in the location where the equivalent of the plastic strain is decreased suddenly. The maximum compressive residual stress 71 MPa were located at the fusion zone (FZ) of the TIG weldment with filler ER4043, whereas minimum compressive residual stress 37 MPa was obtained at stir zone (SZ) of the TIG+FSP with filler 5356, the residual stresses profile of TIG+FSP were not symmetrical about the centerline of the

Please cite this article as: Mehdi H, Mishra RS, Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, Defence Technology, https://doi.org/10.1016/j.dt.2020.04.014

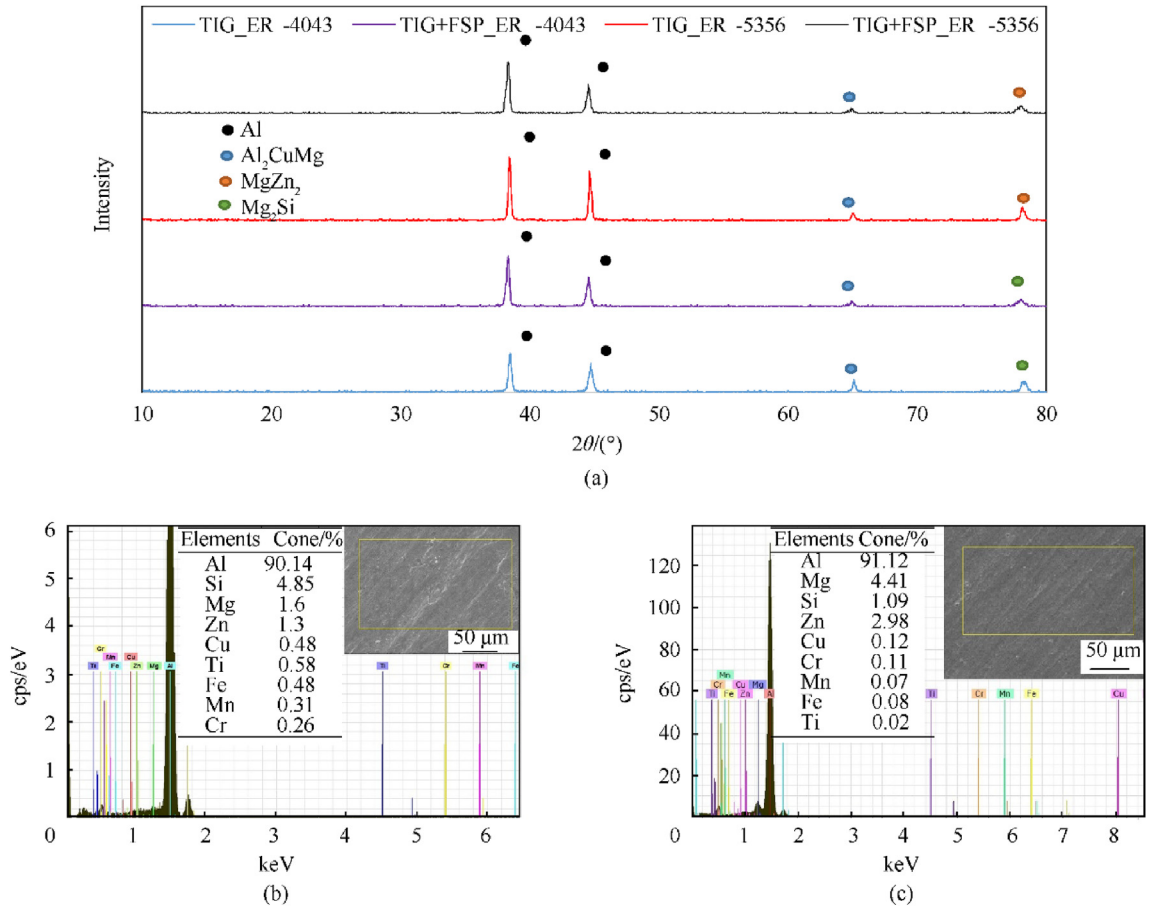


Fig. 5. (a) XRD peaks of TIG and TIG+FSP welded joint, (b&c) EDX analysis of TIG welded joint in fusion zone with filler ER4043 and filler ER 5356.

weldment for both the cases. The left side peak value of the weldment was greater than the right side because the forming processes are different for the retreating side (RS) and advancing side (AS).

4.5. Model validation

To confidently use CFD results for investigating the TIG+FSP process, the CFD model has to be correctly defined and a thorough validation has to be achieved. The model was first applied to simulate the experimental work on friction stir welded joint of Al-alloy 6061 carried out by Hwang el al [50]. The experimental temperatures were measured by thermocouple units placed equally at a distance of 5 mm, along the traverse direction of the rotating tool. The tool rotational speed and traverse speed, was 920 rpm and 20 mm/min respectively. The present simulation result was validated by this experimental results which gives the satisfactory amount of assurance in the fidelity of the simulation of welded joints as shown in Table 11.

4.6. Temperature variation in TIG+FSP process

The experimental temperature results of TIG+FSP measured by thermocouples. The advancing side temperatures in the transverse direction are A_1, A_2, A_3, and A_4 and the retreating side has R_1, R_2, R_3, and R_4. The peak temperature of advancing side is marginally higher than the retreating side [51,52]. The temperature of FSP tool is symmetric about the tool axis. The high temperature in the vicinity of the welding tool is attributed to the localized heat generation. During the TIG+FSP process, the temperature of the base plate around the welding tool is around the 765 K. It is still lower than the melting point of AA6061 and AA7075. When the position weld bead is far away from the FSP tool, the temperature drops quickly as shown in Fig. 9(a) and (b). Fig. 8 shows the temperature distribution plot for TIG+FSP welding at 160 s. During this simulation, eight points were observed to obtain the temperature-time curves which can be compared with the experimental results. The area around the FSP tool reached the maximum temperature. The maximum temperature was observed 760 K at the advancing side whereas minimum temperature was observed 307 K at the retreating side. Higher heat is generated in the SZ at high tool rotation, the prevailing thermal conditions are controlled by the distribution and availability of precipitates in the matrix. Fig. 10 shows the temperature variation profile on the mid-section of the top surface of (TIG+FSP) welded joint with different processing parameters. The region where its peak temperature is higher almost 450 °C in (TIG+FSP) of AA6061 and AA7075 aluminum alloy at tool rotational speed of 1600 rpm. The heat is transmitted to the aluminum alloy which is preheated during tool rotation. The initial heating was predicted as the monitoring thermal contour of the rotating tool. When the rotating tool moves on the monitoring location, the temperature contour leads to slow cooling. Due to the relation between tool pin rotation and material flow, the higher shearing rate was observed at (A.S). Thus the temperature of (RS) is slightly lower than the (AS), while almost the symmetric temperature was observed at the bottom of the welded joint.

The estimated maximum temperature about 515 °C was calculated in the stir zone of the TIG+FSP welded joint at tool rotation of

Please cite this article as: Mehdi H, Mishra RS, Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, Defence Technology, https://doi.org/10.1016/j.dt.2020.04.014

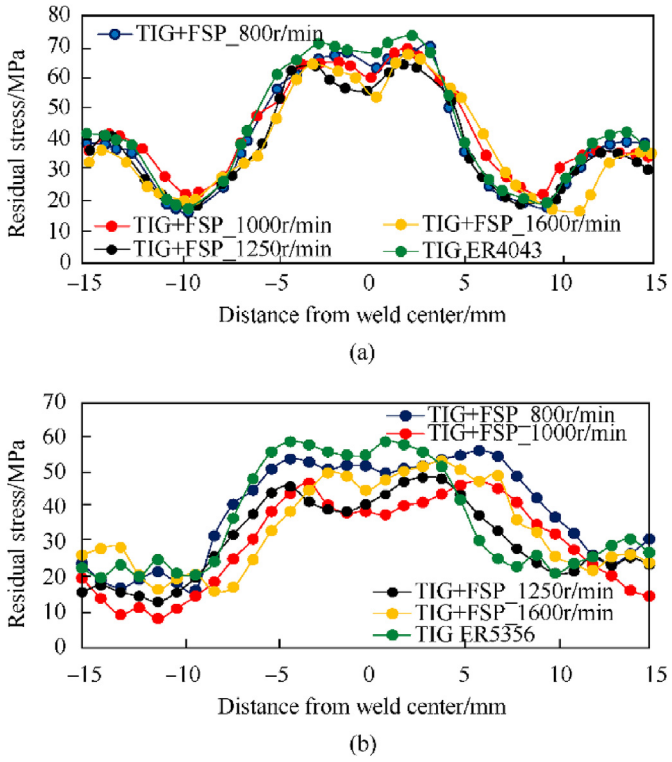


Fig. 6. Variation of transverse compressive residual stress at the welded joint, (a) filler ER4043, (b) filler ER5356.

1600 rpm whereas 408 °C was calculated in the stir zone at 800 rpm. The temperature and heat flux distribution with different processing parameters are shown in Figs. 10 and 11. In general, the temperature values at the (A.S) about 10–25 °C greater than the

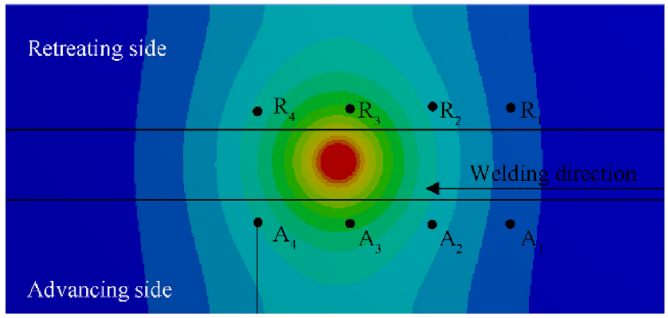


Fig. 8. Temperature contour of TIG+FSP at 160 s.

(R.S), When the tool approaches the target location then there is a rapid change in the temperature, while the slower cooling rate was observed when the tool moves away from the target location [53]. The variation of heat flux at the tool workpiece interface is shown in Fig. 11. It was observed that the heat flux is directly proportional to the tool rotational speed. The maximum heat flux about 5.33×10^6 w/m^{2}was obtained in the SZ at 1600 rpm, whereas minimum heat flux was obtained at 1000 rpm. The non-uniformity was observed in the heat flux pattern at different processing parameters of the tool, because the rapid recirculation of plasticized material, the heat flux rate does not lead to the variation of local temperature.

4.7. Microstructure analysis of TIG and TIG+FSP welded joint

For metallography observation, samples were first ground and mechanically polished and then etched in keller reagent (HNO_3, HF and HCL) and observed by the microscopic machine. The microstructure of TIG welded joint of Al-6061 and 7075 with filler ER4043 and ER 5356 were observed at fusion zone as shown in

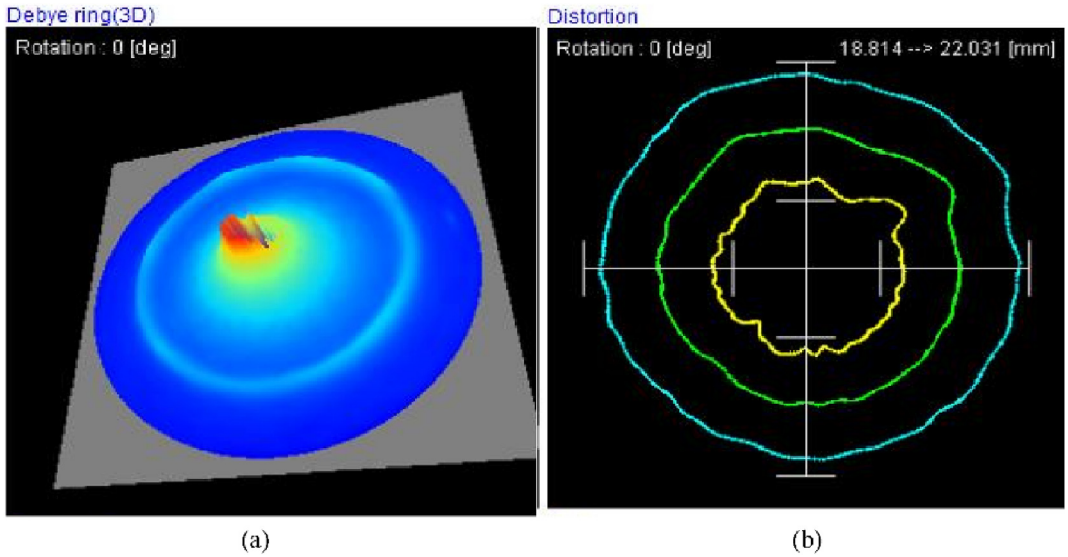


Fig. 7. (a) 3D Debye ring at the center of the TIG+FSP welded joint, (b) 2D Distortion ring at the center of the TIG+FSP welded joint.

Table 11
Validation of temperature distribution of friction stir welded joint of AA6061 [50].

Author	Rotation speed/rpm	Traverse speed/(mm·min^{-1})	Maximum temperature/°C
Present result	920	20	398
Hwang et al. [50]	920	20	386

Please cite this article as: Mehdi H, Mishra RS, Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, Defence Technology, https://doi.org/10.1016/j.dt.2020.04.014

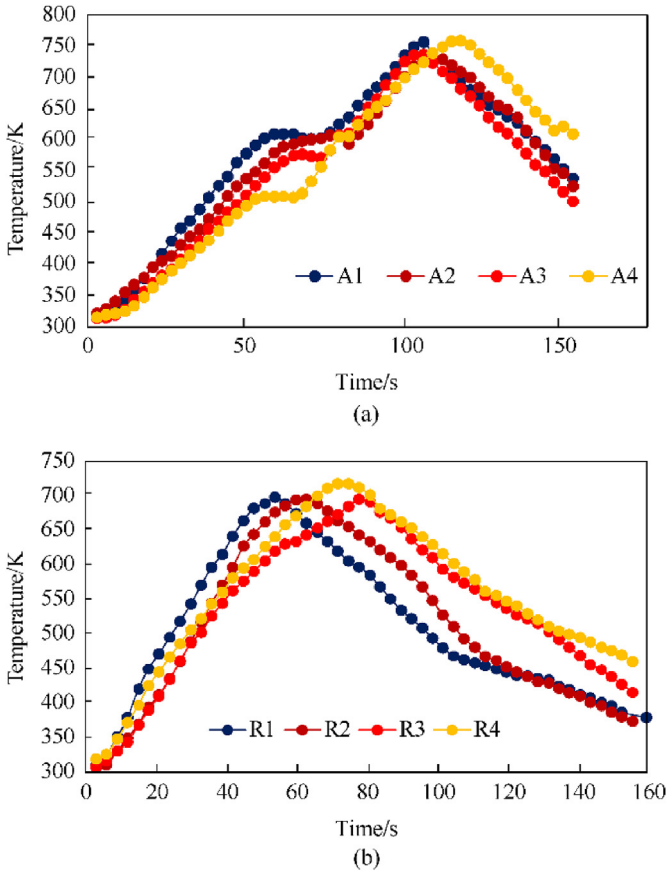


Fig. 9. Variation of temperature during TIG+FSP welding: (a) Advancing side, (b) Retreating side.

Fig. 12. The dispersed precipitates of Al_2CuMg intermetallic compound were observed in welded region and coarse equiaxed grains were also observed in fusion zone with filler ER4043. The equiaxed grains is absent due to the material close to the fusion line provides plenty of sources for the crystal nucleation of the liquid metal during the TIG welding [54]. Fine grains with small precipitates was reported in TIG welded joint with filler ER5356 as compare to ER4043.

Three different zones have been recognized in TIG+FSP weldment at low magnification due to mechanical and thermal stresses caused by the processing parameters. These zones are nugget zone (NZ), thermo-mechanically affected zone (TMAZ), and heat affected zone (HAZ). The formation of nugget zone shape in TIG+FSP welded joint is recognized to the maximum deformation and plasticization in the material which shows the fine recrystallized equiaxed grains. The formation of nugget shapes depends on thermal gradient, processing parameters and tools geometry in the work-piece [55]. Therefore, coarse grain structure of TIG welded joint is transformed into the uniform and fine grains structures in the weld nugget zone due to adequate softening of material revealed the maximum tensile strength and microhardness of the TIG+FSP welded joint as shown in Fig. 13. At high welding speed, the weld nugget zone is more homogenous than those produce low welding speed because high heat input gives the effective recrystallization and more homogenous temperature distribution in weld nugget zone. The grains size in stir zone change crucially which is depend on the heat input and processing parameters [56,57].

The grain size of TIG weldment were analyzed by the image J software and observed grain size in fusion zone with filler ER4043 and ER5356 are 20.4 μm, and 18.2 μm. Fig. 14 shows the effect of tool rotational speed on grain size of TIG+FSP welded joint at nugget zone. It can be conclude that the rotational speed has a significant effect of grain size of the welded joint. When the rotational speed increases, grain size decreases, this observation give the satisfactory amount of assurance with Yupeng Li [58]. The average minimum grain size i.e. 3.4 μm was observed in TIG+FSP with filler ER5356 at tool rotational speed 1600 rpm.

Fig. 15 shows the fractured surface after the tensile test. In this analysis two mode of failure were observed. The fractured occurs in retreating side at low rotational speed (800 rpm), on the other hand failure was observed in advancing side at high tool rotational speed (1600 rpm). This observation gives the satisfactory amount of assurance with Mishra [5]. SEM fractograph has been taken from the fractured tensile test. fracture morphology between the TIG and TIG+FSP shows that TIG welded portion shows the large voids whereas TIG+FSP welded portion shows fine dimples, this is the evidence of crack nucleation and growth 4 mm away from the weld line. The ductile fracture of welded joint occurs, an improvement in

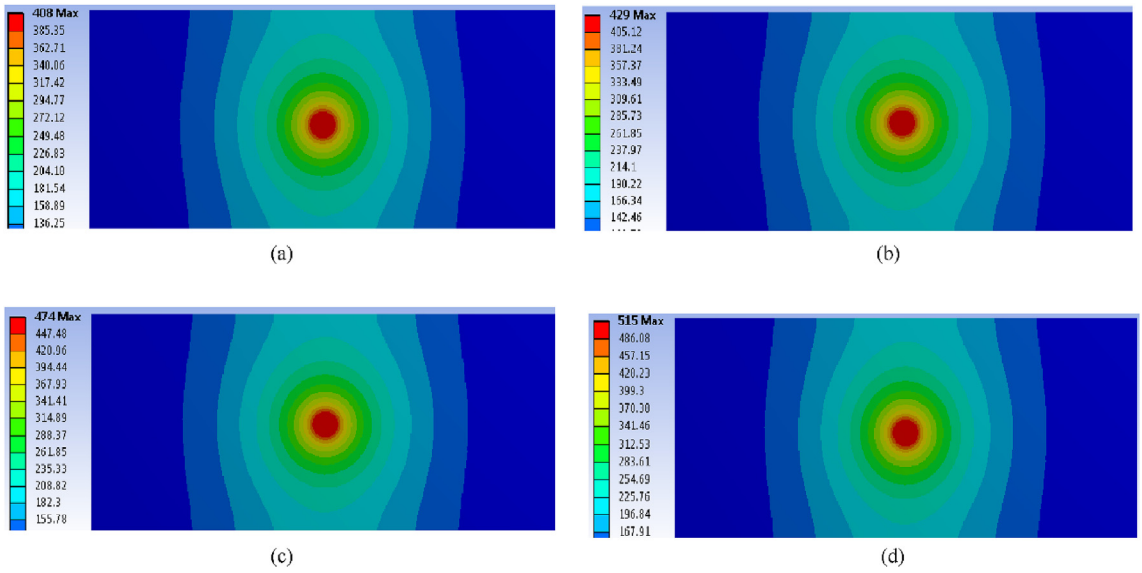


Fig. 10. Temperature distribution at the center of the TIG+FSP welded joint, (a) 800 rpm, (b) 1000 rpm, (c) 1250, (d) 1600 rpm.

Please cite this article as: Mehdi H, Mishra RS, Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, Defence Technology, https://doi.org/10.1016/j.dt.2020.04.014

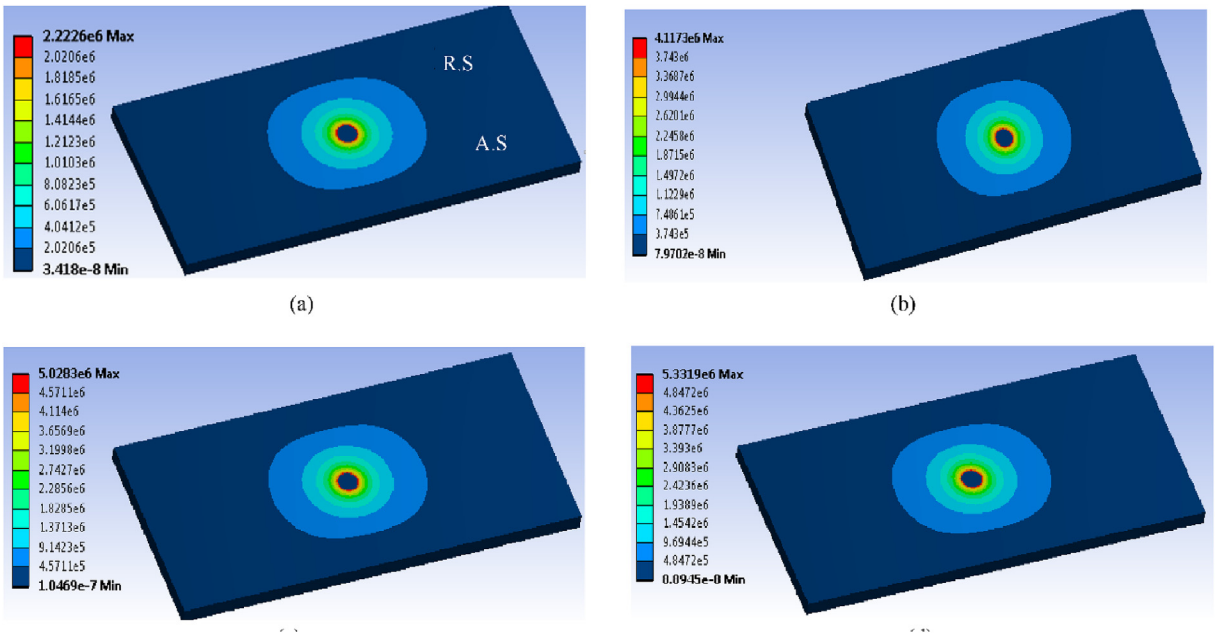


Fig. 11. The heat flux at the center of TIG+FSP welded joint, (a) 800 rpm, (b) 1000 rpm, (c) 1250 rpm, (d) 1600 rpm.

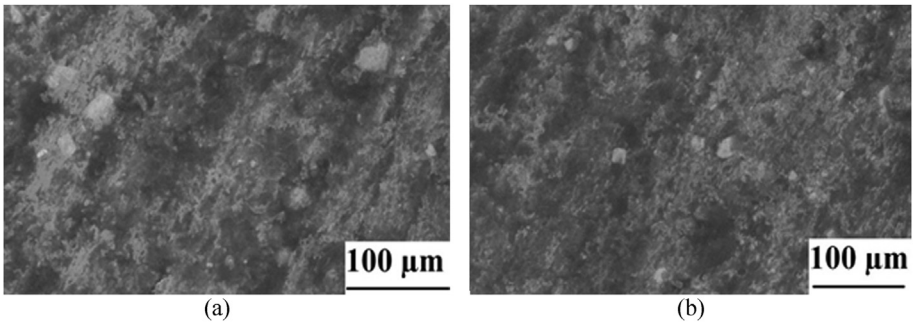


Fig. 12. SEM images of TIG welded joint at fusion zone (a) filler ER 4043, (b) filler ER 5356.

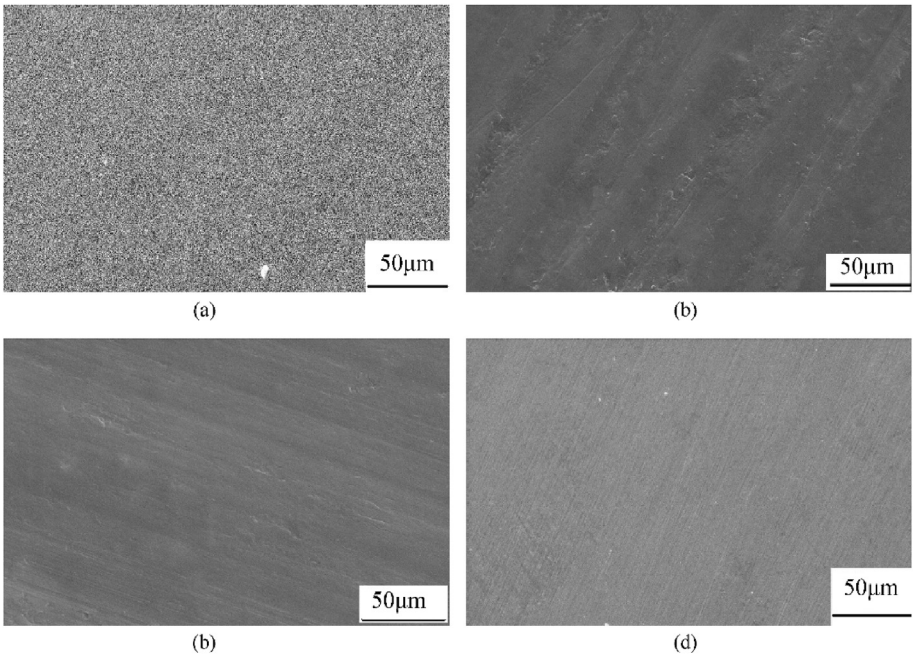


Fig. 13. SEM images of TIG+FSP welded joint at tool rotation (a) 800 rpm, (b) 1000 rpm, (c) 1250 rpm, (d) 1600 rpm.

Please cite this article as: Mehdi H, Mishra RS, Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, Defence Technology, https://doi.org/10.1016/j.dt.2020.04.014

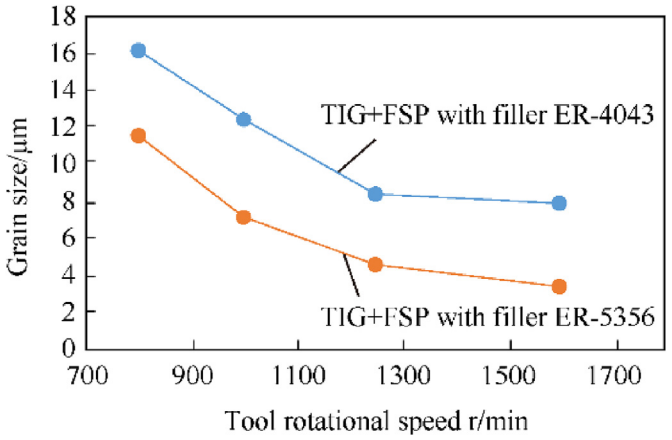


Fig. 14. Effect of tool rotational speed on grain size of TIG+FSP welded joint.

ductility may be achieved when the cavity nucleation could be suppressed [59]. The maximum inter-facial normal stress is depends upon the grain particle size and the volume fraction of the grain particles [60]. The fractured surface at room temperature were shown in Fig. 15.

The large dimples and quasi cleavage with sharp edge and various depth were found on the fractured tensile specimen surface of TIG welded joint with filler ER4043 and ER5356as shown in Fig. 15(a) and (b). Many large and equiaxed dimples were observed in TIG welded joint with filler ER4043 whereas the main fractured mode was quasi cleavage with lot of tiny dimples gathered around the large and quasi cleavage dimples found in filler ER5356 which shows the ductile fracture. The FSW processing parameters effect the temperature distribution and material flow across the TIG welded joint. Fig. 15(c) shows the SEM micrograph of factored surface of TIG+FSP welded joint, finer dimples are observed in TIG+FSP weldment as compare to TIG welded joint, which shows the better mechanical properties.

5. Conclusion

A new welding approach has been investigated to see the effect of FSP on TIG welding for dissimilar alloy AA6061 and AA7075 followed by the different filler wire ER4043 and ER5356. The mechanical characterization, finite element formulation and mathematical equations of heat transfer of TIG+FSP welded joints are investigated using ANSYS Fluent software by adjusting process parameters of FSP. The confidence interval has shown that tensile strength and hardness increased with increasing tool rotation, whereas residual stress decreased with increasing tool rotation. The maximum compressive residual stress 71 MPa were located at the fusion zone (FZ) of the TIG weldment with filler ER4043, whereas minimum compressive residual stress 37 MPa was obtained at stir zone (SZ) of the TIG+FSP with filler 5356. The large dimples and quasi cleavage with sharp edge and various depth were found on the fractured tensile specimen surface of TIG welded joint with filler ER4043 and ER5356 whereas fine dimples are observed in TIG+FSP weldment, which shows the better mechanical properties. TIG+FSP resulted in generated fine grains of 3.2−4 μm. The grain size decreases by increasing the tool rotational speed. The minimum grain size was obtained in TIG+FSP welded joint with filler ER5356 at 1600 rpm.

The asymmetry of the temperature distribution during TIG+FSP welding of AA7075 and AA6061 determined by numerical simulation corresponded with the experimental results. The peak temperatures on AS were higher than the RS of 20 K. The heat transfer is

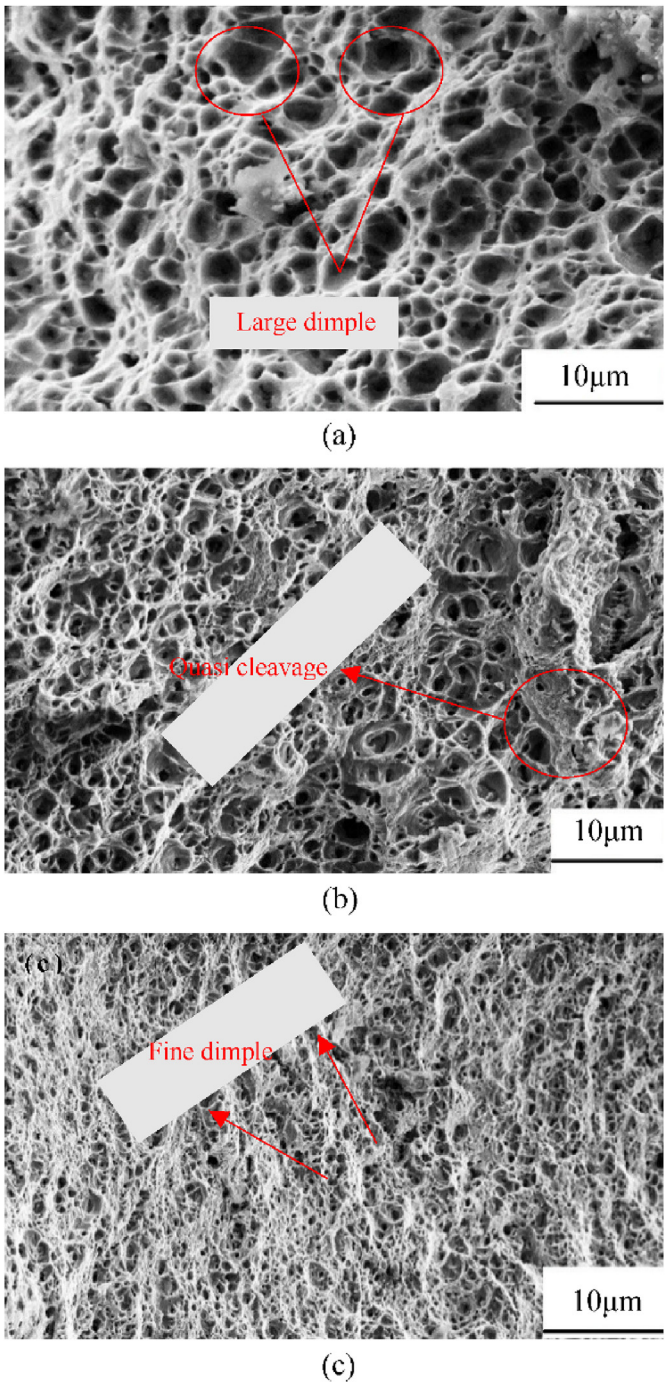


Fig. 15. SEM images of fractured tensile specimen, (a) TIG welded joint with filler ER4043, (b) TIG welded joint with filler ER4043, (c) TIG+FSP welded joint.

analyzed by the effect of different processing parameters of FSP tool. The temperature at advancing side is higher than the retreating side. The predicted peak values of temperature at the weld region was calculated by the ANSYS software and found the maximum temperature about 515 °C at tool rotation of 1600 rpm, whereas the maximum heat flux rate about 5.33×10^6 was observed at 1600 rpm.

Declaration of competing interest

The author declare that they have no conflict interest.

Please cite this article as: Mehdi H, Mishra RS, Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, Defence Technology, https://doi.org/10.1016/j.dt.2020.04.014

References

[1] Dialami N, Chiumenti M, Cervera M, de Saracibar CA. Challenges in thermo-mechanical analysis of friction stir welding processes. Arch Comput Methods Eng 2017b;24:189–225.
[2] Gemme F, Verreman Y, Dubourg L, Jahazi M. Numerical analysis of the dwell phase in friction stir welding and comparison with experimental data. Mater Sci Eng A 2010;527(16):4152–60.
[3] Mosneaga VA, Mizutani T, Kobayashi T, Toda H. Impact toughness of weldments in Al-Mg-Si alloys. Mater Trans 2002;43:1381–9.
[4] Karunakaran N. Effect of pulsed current on temperature distribution and characteristics of GTA welded Magnesium alloy, vol. 4; 2013. p. 1–8.
[5] Mishra RS, Ma ZY. Friction stir welding and processing. Mater Sci Eng R Rep 2005;50(1):1–78.
[6] Zhan M, Guo K, Yang H. Advances and trends in plastic forming technologies for welded tubes. Chin J Aeronaut 2015;29(2):305–15.
[7] Masubuchi K. Analysis of welded structures, residual stresses, distortion, and their consequences. New York: Pergamon Press; 1980.
[8] Chen G-Q, Shi Q-Y, Li Y-J, Sun Y-J, Dai Q-L, Jia J-Y, Zhu Y-C, Wu J-J. Computational fluid dynamics studies on heat generation during friction stir welding of aluminum alloy. Comput Mater Sci 2013;79:540–6.
[9] Su H, Wu CS, Pittner A, Rethmeier M. Thermal energy generation and distribution in friction stir welding of aluminum alloys. Energy 2014;77:720–31.
[10] Nandan R, Roy GG, Lienert TJ, Debroy T. Three-Dimensional heat and material flow during friction stir welding of mild steel. Acta Mater 2007;55(3):883–95.
[11] Ji S, Shi Q, Zhang L, Zou A, Gao S, Zan L. Numerical simulation of material flow behavior of friction stir welding influenced by rotational tool geometry. Comput Mater Sci 2012;63:218–26.
[12] Yu Z, Zhang W, Choo H, Feng Z. Transient heat and material flow modeling of friction stir processing of Magnesium alloy using threaded tool. Metall Mater Trans A 2012;43(2):724–37.
[13] Su H, Wu CS, Bachmann M, Rethmeier M. Numerical modeling for the effect of pin profiles on thermal and material flow characteristics in friction stir welding. Mater Des 2015;77:114–25.
[14] Dalle Donne C, Biallas G, Ghidini T, Raimbeaux G. Effect of weld imperfections and residual stresses on the fatigue crack propagation in friction stir welded joints. In: 2nd international conference on friction stir welding. Gothenburg, Sweden; 2000.
[15] Dalle Donne C, Lima E, Wegener E, Pizalla A, Buslaps T. Investigations on residual stresses in friction stir welds. In: 3rd international conference on friction stir welding, Kobe, Japan; 2001.
[16] Peel M, Steuwer A, Preuss M, Withers PJ. Withers, microstructure, mechanical properties and residual stresses as a function of welding speed in aluminum AA5083 friction stir welds. Acta Mater 2003;51(16):4791–801.
[17] Wang X-L, Feng Z, David SA, Spooner S, Hubbard CR. In: 6th int. Conf. Residual stresses, Oxford, UK, July 10-12; 2000.
[18] Cao YJ, Qi X. Thermal and thermo-mechanical modeling of friction stir welding of aluminium alloy 6001-T6. J Mater Process Manuf Sci 1998;7:215–33.
[19] Chen CM, Kovacevic R. Finite element modeling of friction stir welding-thermal and thermomechanical analysis. Int J Mach Tool Manufact 2003;43:1319–26.
[20] Zhu XK, Chao YJ. Numerical simulation of transient temperature and residual stresses in friction stir welding of 304L stainless steel. J Mater Process Technol 2004;146:263–72.
[21] Li T, Shi Q-Y, Li H-K, Wang W, Cai Z-P. In: Proc. Int. Weld. Join. Conf. Seoul, Korea: COEX Convention Center; May 2007.
[22] Sutton MA, Reynolds AP, Wang DQ, Hubbard CR. J Eng Mater Technol 2002;124(4):215–21.
[23] Mehdi Husain, Mishra RS. Mechanical properties and microstructure studies in friction stir welding (FSW) joints of dissimilar alloy-A review. J Achiev Mater Manuf Eng 2016;77(1):31–40.
[24] Mehdi Husain, Mishra RS. Influences of process parameter and microstructural studies in friction stir welding of diferent alloys: a review. Int J Adv Prod Ind Eng 2017;509:55–62.
[25] Mehdi Husain, Mishra RS. Mechanical and microstructure characterization of friction stir welding for dissimilar alloy—a review. Int J Res Eng Innov 2017;1(5):57–67.
[26] Mehdi Husain, Mishra RS. Analysis of material fow and heat transfer in reverse dual rotation friction stir welding: a review. Int J Steel Struct 2019;19:422–34.
[27] Husain Mehdi H, Mishra RS. Study of the influence of friction stir processing on tungsten inert gas welding of diferent aluminum alloy. SN Appl Sci 2018;1:712. 2019.
[28] Taheri-Behrooz Fathollah, Aliha Mohammad RM, Maroofi Mahmood, Hadizadeh Vahid. Residual stresses measurement in the butt joint welded metals using FSW and TIG methods. Steel Compos Struct 2018;28(No. 6):759–66.
[29] Zhang Changqing, Wang Weijie, Jin Xin, Rong Chen, Qin Zhuo. A study on microstructure and mechanical properties of micro friction stir welded ultra-thin al-1060 sheets by the shoulder less tool. Metals 2019;9(5):507.
[30] Devireddy Krishnaja, Devuri Venkateswarlu, Cheepu Murali Mohan, Kranthi Kumar Balina. Analysis of the influence of friction stir processing on gas tungsten arc welding of 2024 auminum alloy weld zone. Int J Mech Prod Eng Res Dev 2018;8(1):243–52.
[31] Mabuwa S, Msomi V. Effect of friction stir processing on gas tungsten arc-welded and friction stir-welded 5083-H111 aluminium alloy joints. Adv Mater Sci Eng 2019:1–14.
[32] Ishak Mahadzir, Noordin Nur Fakhriah Mohd, Shah Luqman Hakim Ahmad. Feasibility study on joining dissimilar aluminum alloys aa6061 and aa7075 by tungsten inert gas (TIG). J Teknol 2015;75(7):79–84.
[33] Mohammadi-pour M, Khodabandeh A, Mohammadi-pour S, et al. Microstructure and mechanical properties of joints welded by friction-stir welding in aluminum alloy 7075-T6 plates for aerospace application. Rare Met 2016. https://doi.org/10.1007/s12598-016-0692-9.
[34] Arora A, De A, DebRoy T. Toward optimum friction stir welding tool shoulder diameter. Scripta Mater 2011;64(1):9–12.
[35] Sasaki T, Hirose Y, Sasaki K, Yasukawa S. Influence of image processing conditions of debye Scherrer ring images in x-ray stress measurement using an imaging plate. Adv X Ray Anal 1997;40:588–94.
[36] Salleh MNM, Ishak M, Shah LH, Idris SRA. The effect of ER4043 and ER5356 filler metal on welded Al 7075 by metal inert gas welding. WIT Trans Built Environ 2016;166:213–24.
[37] Smith William F, Hashemi Javad. Foundations of materials science and engineering. fourth ed. McGraw-Hill; 2006, ISBN 978-0-07-295358-9.
[38] Xu W, Liu J, Luan G, Dong C. Temperature evolution, microstructure and mechanical properties of friction stir welded thick 2219-O aluminum alloy joints. Mater Des 2009;30:1886–93.
[39] Xu W, Liu J, Luan G, Dong C. Microstructure and mechanical properties of friction stir welded joints in 2219-T6 aluminum alloy. Mater Des 2009;30:3460–7.
[40] Luijendijk T. Welding of dissimilar aluminium alloys. J Mater Process Technol 2000;103:29–35. 2000.
[41] Jia Yonggui Qin, Ou Yiwen, Wang Kehong, Liu Jizi. The influence of microstructural heterogeneity on mechanical properties of friction stir welded joints of T6-treated Al-Zn-Mg alloy 7A52Yang. Metals 2018;8(7):527.
[42] Ishak Mahadzir, Noordin Nur Fakhriah Mohd, Shah Luqman Hakim Ahmad. Feasibility study on joining dissimilar aluminum alloys aa6061 and aa7075 by tungsten inert gas (TIG). J Teknol (Sci Eng) 2015;75(7):79–84.
[43] Taira S, Tanaka K, Yamazaki T. A method of x-ray microbeam measurement of local stress and its application to fatigue crack growth problems. J Soc Mater Sci Jpn 1978;27(294):251–6.
[44] Yeni C, Sayer S, Pakdil M. Comparison of mechanical and microstructural behaviour of TIG, MIG and friction stir welded 7075 aluminium alloy. Kovove Mater 2009;47:341–7.
[45] Rossini NS, Dassisti M, Benyounis KY, Olabi AG. Methods of measuring residual stresses in components. Mater Des 2012;35:572–88.
[46] James MN. Residual stress influences in mechanical engineering. In: XVIII congreso nacional de Ingeniería Mecánica; 2010. p. 1–13.
[47] Yang YP, Babu SS, Kikel JM, Brust FW. Investigation of weld crack mitigation techniques with advanced numerical modeling and experiment-Summary. In: Lippold J, Böllinghaus T, Cross CE, editors. Hot cracking phenomena in welds III; 2011. p. 353–65.
[48] Buchanan DJ, John R. Residual stress redistribution in shot peened samples subject to mechanical loading. Mater Sci Eng A 2014;615:70–8.
[49] Zhang Zhao, Zhang Hongwu. The simulation of residual stresses in friction stir welds. J Mech Mater Struct 2007;2(5):951–64.
[50] Hwang Yeong-Maw, Kang Zong-Wei, Chiou Yuang-Cherng, Hsu Hung-Hsiou. Experimental study on temperature distributions within the workpiece during friction stir welding of aluminum alloys. Int J Mach Tool Manufact 2008;48:778–87.
[51] Yaduwanshi DK, Bag S, Pal S. Numerical modeling and experimental investigation on plasma-assisted hybrid friction stir welding of dissimilar materials. Mater Des 2016;92:166–83.
[52] Nandan R, Roy GG, Debroy T. Numerical simulation of three-dimensional heat transfer and plastic flow during friction stir welding. Metall Mater Trans 2006;37(4):1247–59.
[53] Shi L, Wu CS, Chen MA. Numerical analysis of heat transfer and material flow in reverse dual-rotation friction stir welding. In: Proceedings of the 1st international joint symposium on joining and welding, Osaka, Japan, 6–8 November; 2013.
[54] Kou S. Welding metallurgy. second ed. New York: John Willy & Sons; 2002.
[55] Mahoney MW, Rhodes CG, Flintoff JG, et al. Properties of friction-stir-welded 7075 T651 aluminum. Metall Mater Trans A 1998;29:1955–64.
[56] Mohammadzadeh Jamalian H, Farahani M, Besharati Givi MK, et al. Study on the effects of friction stir welding process parameters on the microstructure and mechanical properties of 5086-H34 aluminum welded joints. Int J Adv Manuf Technol 2016;83:611–21.
[57] Kasman Ş, Yenier Z. Analyzing dissimilar friction stir welding of AA5754/AA7075. Int J Adv Manuf Technol 2014;70:145–56.
[58] Li Yupeng, Sun Daqian, Gong Wenbiao. Effect of tool rotational speed on the microstructure and mechanical properties of Bobbin tool friction stir welded 6082-T6 aluminum alloy. Metals 2019;9(8):894.
[59] Good SH, Brown IM. Acta Mater 1979;27:1.
[60] Tszeng TC. Interfacial stresses and void nucleation in discontinuously reinforced composites. J Eng Mater Technol 2000;122:86–92.

Please cite this article as: Mehdi H, Mishra RS, Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, Defence Technology, https://doi.org/10.1016/j.dt.2020.04.014

Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 06, June-2020

Effect of Mesh Size on Soil-Structure Interaction in Finite Element Analysis

Nasasira Derrick[1]
Department of Civil and Environmental Engineering,
Delhi Technological University,
New Delhi, India.

Amit Kumar Srivastava[2]
Department of Civil and Environmental Engineering,
Delhi Technological University,
New Delhi, India.

***Abstract*- Numerous consitutive models have been developed over the past years for modelling the stress-strain behaviour of soils, using finite element and/or finite difference calculations of soil structures and soil/structure interaction problems under axisymetric, plane strain, and/or general three dimensional conditions. Selection of the appropriate parameters and soil model can have a significant impact on the results of numerical analysis. In this study, the interaction of soil and reinforced cement concrete raft foundation in a sandy soil at a depth of 9m below the existing ground level is analysed by using with both analytical and finite element analysis (FEM). Analytical calculations were done using method available in literature and a finite element analysis was carried out using the hardening soil model available in PLAXIS 2D (2020, connect version). The results obtained from analytical calculations were compared with those obtained from the Hardening Soil (HS) Model. In the model sizes of mesh were changed and there effect on interaction behaviour is observed by studying the strength and deformation results obtained from FEM analysis. It was observed that very fine mesh size gave more accurate results compared to the other sizes.**

Keywords- Plaxis, sand, settlement, Raft

I. INTRODUCTION

Evaluation of correct strength and deformation behaviour of raft resting on the soil is important for safe and economical design. Strength and deformation behaviour of raft foundation can be analysed by different numerical methods. With the development of numerical methods such as finite element anaylsis and finite difference method, it has become feasible to analyse and predict the behaviour of complex soil structures and soil/structure interaction problems. Such anaylses depend on the representation of the relations between stresses and strains in a given material and are represented by a consitutive model, consisiting of mathematical expressions that model the behaviour of soil in a single element. Therefore the purpose of a consitutive model is to simulate the soil behaviour with sufficient accuracy under all conditions with sufficient accuracy under all loading conditions in the numerical computations.

II. LITERATURE REVIEW

The Hardening Soil model was established in 1999 [1]. In the framework of the theory of elasticity. The model involves friction hardening to model the plastic shear strain in deviatoric loading, and cap hardening to model the plastic volumetric strain in primary compression.because of these two types of hardening, the model is accurate for problems (like excavations) involving a reduction of mean effective stress and at the same time mobilization of shear strength. The verification and formulation of this model was discussed [1]. It is an advanced and latest model for the simulation and modelling of soil behaviour. The Hardening soil model requires the input of 10 parameters, i.e. three reference stifness parameters (E_{50}^{ref} for triaxial compression, E_{ur}^{ref} for triaxial unloading and E_{oed}^{ref} for oedometer loading) at a reference stress level P^{ref}, a power, *m*, for the stress dependent stifness formulation, poisson's ratio for unloading and reloading, μ_{ur}, the mohr coulomb strength parameters, Ø, and c, the dilatancy angle, ψ, the K_o value in primary one-dimensional compression (K_o^{nc}), a parameter called the failure ratio, R_f which determines the strain level at failure [2]. Fig. 1 shows the failure criterion of the harening soil model in principal stress space. The yield surface of the hardening soil model is not fixed in principal stress space, but it can expand due to plastic straining. The Plaxis hardening model can be used to accurately predict displacement and failure for general types of soils in various geotechnical applications [3]. The model does not include anisotropic strength or stiffness, nor time-dependent behaviour (creep). The hardening soil model is also known for stress-dependency of stiffness moduli [4]. This explains that all stiffnesses increases with the pressure.

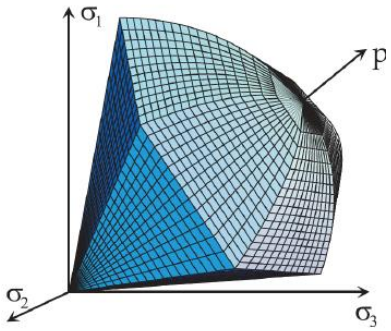


Figure 1: Failure criterion of the Hardening soil model in principal stress space [1].

The Hardening Soil model, also called as HS-Standard was explained by [1]. In order to modify the important phenomena which is exhibited by soils such as:

1. Densification, defines as decrease of the voids volume in soil due to plastic deformations, which also decreases the void ratio,

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 06, June-2020

2. Stress dependent stiffness, is basically an observed technique of increasing the stiffness modules with the increasing of confining stress,
3. Soil stress history, explains the accounting for preconsolidation process and its effects,
4. Plastic yielding, is very important which accounts for the development of irreversible strains with reaching a yield criteria,
5. Dilatancy, explains as the uses for an occurrence of negative volumetric strains during shearing.

This model is known to be one of the simplest and important in the class of latest models designed to handle the small strain stiffness as explained by [5]. It has two plastic mechanisms, shear and volumetric. Different to the other material models such as the Cap model or the Modified Cam Clay model, the magnitude of soil deformations, stresses, forces, etc. can be modelled more accurately by using three different input stiffness parameters which is related to triaxial loading stiffness (E_{50}), triaxial unloading-reloading stiffness (E_{ur}), and oedometer loading modulus (E_{oed}).
There are five parameters in the hardening soil model differ than the MC model is:

1. (E_{50ref}) secant modulus 50% strength[kN/m^2]
2. (E_{oedref}) Oedometric modulus [kN/m^2]
3. (E_{urref}) Unloading-reloading modulu[kN/m^2]
4. (ν_{ur}) Unloading-reloading Poisson's ratio [-]
5. (m) Exponent of the stress-stiffness function

Parameters previously used in MC model:

1. (Ø) Friction angle [°]
2. (c) Cohesion [kN/m^2]
3. (ψ) Dilatancy angle [°]

The hardening model can be used to predict displacement and failure for different types of soils in various geotechnical applications. This model does not include anisotrpic strength or stifness, nor time-dependent behaviour (creep). It's capabilities for dynamic applications are limited, but this model is, in principle, without any doubt the most accurate general soil model.

A. *Correlation between Angle of Friction and Standard Penetration Number*

The peak friction angle Ø', of granular soil has also been correlated with N_{60} by several investigators. A correlation between N_{60} and Ø' was given by [6] in a graphical form, which can be approximated as,

$$Ø'(deg) = 27.1 + 0.3N_{60} - 0.00054[N_{60}]^2 \quad (1)$$

B. *Correlation between Modulus of Elasticity and Standard Penetration Number*

The modulus of elasticity of granular soils (E_s) is an important parameter in estimating the
elastic settlement of foundations. A first order estimation for E_s was given by [7] as

$$\frac{E_S}{p_a} = \propto N_{60} \quad (2)$$

Where
α = 5 for sands with fines
10 for clean normally consolidated sands
15 for clean overconsolidated sands
p_a= atmospheric pressure in the same units as E_s
N_{60}= Field standard penetration number

III. PROBLEM STATEMENT

Bearing capacity requirement is a basic criteria to be satisfied in the analysis and design of shallow foundations. The criterion on bearing capacity ensures that the foundation does not undergo shear failure under loading. The use of standard penetration test in the analysis of bearing capacity has received numerous attentions [8]. Details of the field application of Standard Penetration Test are specified in IS 2131 [9]. This paper attempts to report interaction behavior of raft foundations resting on sand using Finite element software package provided by Plaxis 2D (connect editon). Steps for modeling is presented in fig. 2 from the model definition to the output results

IV. SELECTION OF PARAMETERS

Parameters used in this study were obtained through laboratory investigation of the undisturbed samples collected from site located at Noida in India, as most of the strata is having poorly graded sand, collection of undisturbed sample was difficult. Hence, the result of standard penetration test conducted as per IS 2131 [9] is used for analysis . Table 1 summarizes the selection of model paramters for analysis of the hardening soil model used in this study.

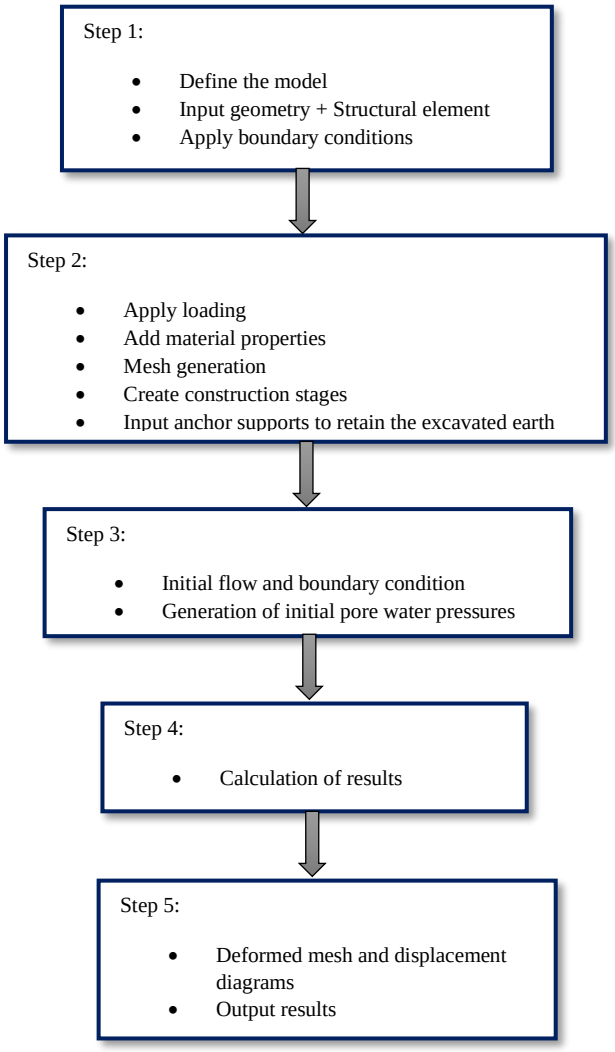


Figure 2: Steps for numerical modeling

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 06, June-2020

SOIL PROFILE AND BOREHOLE DETAILS

The samples were collected from two bore holes drilled at GH-01A/01, SECTOR-16, GREATER NOIDA WEST, U.P (India) and tested as per the relevant protocol of bureau of Indian Standard. The results of SPT are presented in Table2.

ANALYTICAL METHOD

The analytical calculations were done in accordance with Indian standards IS 6403 [10] and permissible settlement is restricted as per IS 1904 [11].

Based on the shear criterion IS code 6403 [10], net ultimate bearing capacity for local shear failure for a cohesionless sand

$$q_{nu} = q(N_q - 1)s_q d_q i_q + \frac{1}{2} B\gamma N_\gamma s_\gamma d_\gamma i_\gamma w' \quad (3)$$

Where s_q, s_γ= shape factors

d_q, d_γ= depth factors

i_q, i_γ= Inclination factors

N_q, N_γ= Bearing capacity factors

w', w''= Correction factor for location of water table.

Assuming there will be rise in water table during mansoon and water will reach to a level of bottom of raft.

$$\emptyset' = tan^{-1}(0.67 tan\emptyset) = 22.78^o$$

Nq=8.76, Nγ=8.44

$$from\ equation\ (3), q_{nu} = \gamma D_f (N_q - 1)s_q d_q i_q w' + \frac{1}{2} B\gamma N_\gamma s_\gamma d_\gamma i_\gamma w'' = 2209.88 KN/m^2$$

Assuming a factor of safety of 2.5, net safe bearing capacity(q_{ns}) $= 883.95 KN/m^2$

TABLE 1: SELECTION OF MODEL PARAMETERS FOR ANALYSIS USING THE HARDENING SOIL MODEL

General			
Material model	Model	Hardening soil model	-
Type of material behaviour	Type	Drained	-
Soil unit weight above phreatic level	γ_{unsat}	16.20 *(lab)*	KN/m^3
Soil unit weight below phreatic level	γ_{sat}	19.10 *(lab)*	KN/m^3
Parameters			
Young's modulus	E'	-	KN/m^2
Secant stiffness in standard drained triaxial test	E_{50}	$3.0x10^4$ *(Plaxis)*	KN/m^2
Tangent stiffness for primary oedometer loading	E_{oed}	$3.0x10^4$ *(Plaxis)*	KN/m^2
Unloading/reloading stiffness	E_{ur}	$1.00x10^5$ *(Plaxis)*	KN/m^2
Power for stress-level dependency of stiffness	m	0.5 *(Plaxis)*	-
Cohesion (constant)	C	0 *(lab)*	KN/m^2
Friction angle	Ø	32 *(lab)*	^{o}
Dilatancy angle	ψ	2.0 *(lab)*	^{o}
Poisson's ratio	μ	0.2 *(lab)*	-
K_0-value for normal consoilidation	K_0^{nc}	0.4701 *(plaxis)*	-
Interfaces			
Strength reduction factor Inter.	R_{inter}	0.67 *(plaxis)*	-
Initial			
K_0 determination	-	Automatic	-
Over-consolidation ratio	OCR	1.0 *(plaxis)*	-

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 06, June-2020

TABLE 2: SPT VALUES FOR DIFFERENT BOREHOLES

	B1		B2	
Depth (m)	Field N-Value	Corrected N-value	Field N-Value	Corrected N-value
1.5	10	17	7	12
3	12	17	9	13
4.5	18	21	13	15
6	20	20	17	17
7.5	24	22	19	17
9	25	21	24	20
10.5	26	20	27	21
12	34	24	29	21
13.5	37	25	30	18
15	39	20	34	19
18	44	21	39	20
21	51	22	46	21
24	56	23	52	22
27	58	23	54	22
30	59	23	59	23
33	55	21	62	23
36	60	22	66	23
39	68	23	71	24
42	70	23	76	25
45	75	24	81	25

TABLE 3: PARAMETER USED FOR CALCULATION BY ANALYTICAL METHOD

Bore hole No	1 & 2	Unit	Protocol
Depth of water table	14	m	
density above water table	16.2	KN/m^3	
density below water table	19.1	KN/m^3	
Type of foundation	raft		
width of foundation	40	m	
length of foundation	42	m	
depth of foundation	9	m	
Permissible settlement of foundation	0.075	m	
Cohesion (c)	0	kN/m^2	Direct Shear Test
Angle of internal friction (φ)	32	Degree	

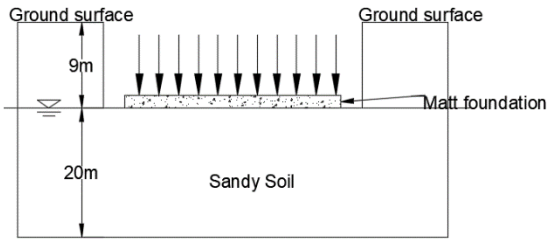


Figure 3: Prototype dimensions

Bearing capacity based on the settlement criterion

$N_{cor} = 22$
$\gamma_d = 16.2KN/m^2$
$\gamma_{sat} = 19.1KN/m^2$
$\mu = 0.2$
$C = 0$
$\emptyset = 32^o$

$$S_e = \frac{q_n B(1-\mu^2)I_F}{E_s} \quad (4)$$

Where q_n= Net safe bearing pressure

B = breadth of foundation

E_s= Young's modulus

I_F = Influence factor

s_e = allowable settlement (75mm)

at

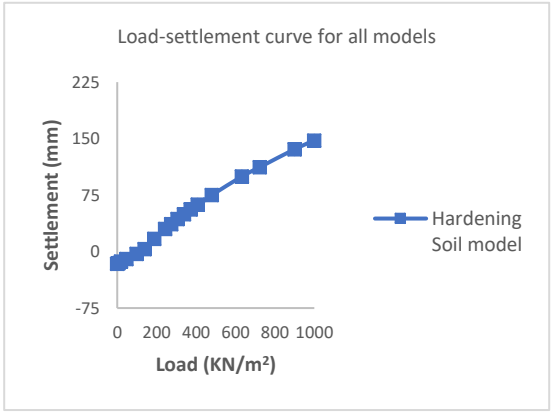


Figure 4: Load-settlement relationship for the Hardening soil model

75mm allowable settlement, from equation 4, q_n= 456.7KN/m^2.Therefore, considering the lower of the two values obtained, Allowable bearing capacity of the foundation was obtained as 456.7KN/m^2.

V. NUMERICAL ANALYSIS

Plaxis 2D was used to evaluate the performance of the Hardening Soil model model to estimate the bearing capacity of a raft foundation situated at a depth of 9m below the surface and to investigate the relationship between the results obtained. Fig 3 shows a sketch of the prototype dimensions showing the location of the water table. Simulations for different excavation stages were done. The distance between the boundary of the floor of the model and the lower boundary of the footing plays an important role in the pattern of bearing capacity value obtained.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 06, June-2020

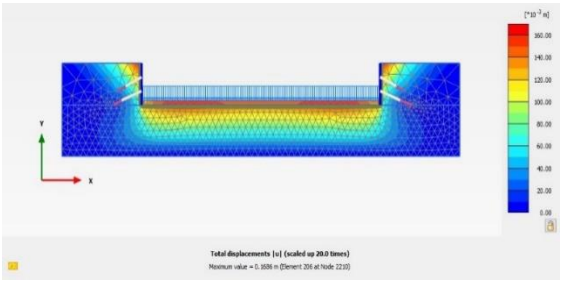


Figure 5: Output image showing total vertical displacement of foundation

To eliminate the effect of this factor, the bottom line intended at a distance greater than 2d.

It can be seen from fig 4, that Plaxis 2D software provides a linear load-settlement relationship for the hardening soil model, which is variant from the theoretical load-settlement curve for a foundation on sandy soil. The curve also indicates an initial negative settlement indicating a swelling in the soil before the footing is placed and load is applied. The reason for initial negative settlement is due to release of in situ soil pressure due to excavation.

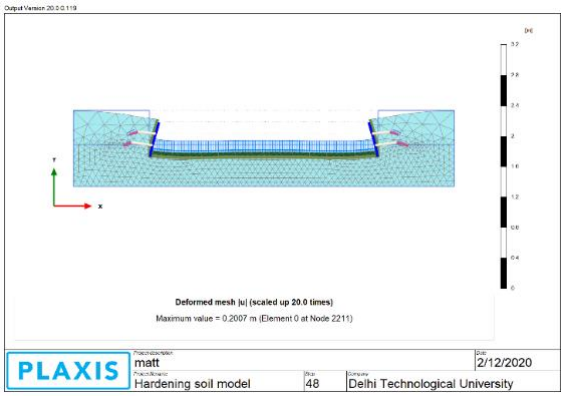


Figure 6: Deformed mesh of foundation.

It can be seen from the graph that settlement increases directly proportional to increase in stress applied on foundation. A value of 481KN/m^2 is obtained at 57mm settlement. It can be seen that the hardening soil model provides a result close to the value obtained from the IS code calculation (456.7KN/m^2) therefore making it more accurate to use in the field for design of foundations. Fig 5 shows the deformation of the raft footing after applying the design load. The contours are drawn for different displacements value below the raft.

Fig 6 shows the typical deformed mesh of the foundation. Three excavation stages were input into Plaxis software since the foundation is laid at a 9m depth. To support the dry excavation, anchors were used. The anchors were prestressed with 500KN force.

VI. EFFECT OF CHANGES OF MESH SIZE ON THE BEARING CAPACITY OF FOUNDATION

Here, the effect of mesh size changes on the model output results was analysed. Different mesh sizes were used in Plaxis 2D which are; very coarse, coarse, medium, fine and very fine. It should be noted that the mesh size of 15 nodal element, is a better distribution of nodes and more accurate results than to element with six nodes. Therefore, the analysis of 15 nodal element is used. Table 3 shows the effect of changes of mesh size on settlement pressure of foundation for 75mm settlement. Figure 7 shows the effect of change in mesh size on deformation of a raft foundation.

TABLE 3: EFFECT OF CHANGE IN MESH SIZE ON ALLOWABLE BEARING CAPACITY OF FOUNDATION

Allowable bearing capacity of foundation at 75mm settlement (KN/m^2)	Change in type of mesh size
494	Very coarse
488	Coarse
480	Medium
465	Fine
460	Very fine

From Fig 7, it can be seen that there is slightly or no difference in the pressure for 75mm of settlement. But at settlement more than 75 mm the effect of mesh size is to decrease the settlement for a given pressure if mesh size is increased from very fine to very coarse.

VII. CONCLUSION

Based on the findings, the following conclusions can be drawn;

1. The bearing pressure for a given settlement obtained by the hardening soil model was closer to the IS analytical method. However an initial settlement of 20mm should be deducted to cater for the swelling provided by plaxis FEM, because of removal of insitu stress because of excavation.
2. Mesh size with 15 nodal elements produces more accurate results as compared with element with six nodes.
3. The mesh sizes influences the bearing pressure for a given settlement. The influences are negligible at a settlement value of 75mm, but influences increase with increase in settlement.
4. Bearing capacity value increases with increase in refinement of mesh size in Finite element analysis. When a finer mesh is used, the value of bearing capacity obtained at a particular settlement is higher than that obtained when a coarse mesh is used. This therefore means researchers should ensure that a fine mesh is used when carrying out finite element analysis to obtain more accurate results.

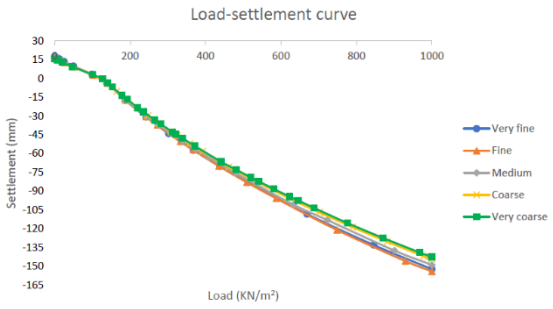


Figure 7: Load-settlement graph of the different mesh sizes

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 06, June-2020

ACKNOWLEDGMENT

The authors wish to thank the Department of Civil Engineering, Delhi Technological University. This work was supported in part by a grant from delhi Technological University, Department of Civil and Environmental Engineering.

REFERENCES

[1] Schanz T, Vermeer A, Bonnier P, (1999), "The hardening soil model: formulation and verification." Beyond 2000 Computional Geotechnics-10 years PLAXIS, Amsterdam, Netherlands, 281p.
[2] Brinkgreve RBJ, Al-Khoury R, (2020), "PLAXIS version 2020 reference manual".
[3] Braja, MD. (1999), "Principle of Foundation Engineering (4th edi.)". PWE Publishing Company, USA. pp243.
[4] Labuz, Joseph & Zang, Arno. (2015). "Mohr-coulomb failure criterion". The ISRM Suggested Methods for Rock Characterization, Testing and Monitoring: 2007-2014. 227-231.
[5] Hewitt, P. B. and Gue, S. S., (1994). "Piled Raft Foundation in a weathered sedimentary formation. Proc. Geotropica", Malaysia, 1-11
[6] Peck, Hanson and Thornburn (1974), "Modification of Teraghi and Peak Correlation by Peak, Hanson and Thornburn"
[7] Kulhawy, F.H. & Mayne, Paul, (1990). "Manual on Estimating Soil Properties for Foundation Design".
[8] Poulos, H.G., (2005). "Piled raft and compensated piled raft foundations for soft soil sites. Advances on designing and testing deep foundations". Geotech. Spec. Publ. (ASCE) 129: 214–235.
[9] Indian standards, IS:2131. (1981), "Method for standard penetration test for soils"
[10] Indian standards, IS:6403. (1981), "Code of Practice for Determination of Bearing Capacity of Shallow Foundations".
[11] Indian standards, IS:1904. (1986), "Code of practice for design and construction of foundations in soils : general requirements".
[12] Józsa, Vendel. (2011), "Effects of rarely analyzed soil parameters for FEM analysis of embedded retaining structures".
[13] Shield RT (1955)., "On Coulomb's law of failure in soils". J Mech Phys Sol 4:10–16
[14] Jerry A Yamamuro., Victor N,(2005), "Soil constitutive models, Evaluation, selection, and calibration". Geotechnical Special Publication No.128.
[15] Meyerhof, GG. (1956), "Penetration Test and Bearing Capacity of Cohesionless Soils". Journal of the Soil Mechanics and Foundation Division, ASCE. 82(1):1-19.
[16] Adachi, T., Oka, F., (1982). "Constitutive equation for normally consolidated clays based on elasto-viscoplasticity" Soils and foundations 22:57-70
[17] Broja, R.I., Kavaznjian, E, (1985). "A constitutive model for the σ-ε-t behaviour of wet clays". Geotechnique 35: 293-298.
[18] Russo, G., (1998). "Numerical Analysis of Piled Rafts. Int. J. Numer. Anal. Meth. Geomech", 22, 477-493.
[19] Prakoso, W. A. and Kulhawy, F. H., (2001). "Contribution To Piled Raft Foundation Design. Journal of Geotechnical and Geoenvironmental Engineering", ASCE, 127(1), 17-24.
[20] Reul, O. and Randolph, M. F., (2003). "Piled Rafts in Overconsolidated Clay: Comparison of in situ Measurements and Numerical Analysis". Géotechnique, 53(3), 301-315.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

Engineering Research Express

ACCEPTED MANUSCRIPT

Efficient strip to slot waveguide mode converter using fast quasi-adiabatic approach

To cite this article before publication: Nikhil Dhingra *et al* 2020 *Eng. Res. Express* in press https://doi.org/10.1088/2631-8695/aba819

Manuscript version: Accepted Manuscript

Accepted Manuscript is "the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an 'Accepted Manuscript' watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors"

This Accepted Manuscript is **© 2020 IOP Publishing Ltd**.

During the embargo period (the 12 month period from the publication of the Version of Record of this article), the Accepted Manuscript is fully protected by copyright and cannot be reused or reposted elsewhere.
As the Version of Record of this article is going to be / has been published on a subscription basis, this Accepted Manuscript is available for reuse under a CC BY-NC-ND 3.0 licence after the 12 month embargo period.

After the embargo period, everyone is permitted to use copy and redistribute this article for non-commercial purposes only, provided that they adhere to all the terms of the licence https://creativecommons.org/licences/by-nc-nd/3.0

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions will likely be required. All third party content is fully copyright protected, unless specifically stated otherwise in the figure caption in the Version of Record.

View the article online for updates and enhancements.

This content was downloaded from IP address 14.139.251.100 on 28/07/2020 at 07:00

IOP Publishing
Journal **XX** (XXXX) XXXXXX https://doi.org/XXXX/XXXX

Efficient strip to slot waveguide mode converter using fast quasi-adiabatic approach

Nikhil Dhingra[1], Geetika Jain Saxena[2], Kamal Kishor[3] and Enakshi Khular Sharma[1]

[1] Department of Electronic Science, University of Delhi South Campus, Benito Juarez Marg, South Moti Bagh, New Delhi - 110021, India
[2]Department of Electronics, Maharaja Agrasen College, University of Delhi, Vasundhara Enclave, New Delhi - 110096, India
[3]TIFAC-Centre of Relevance & Excellence in Fiber Optics & Optical Communications, Department of Applied Physics, Delhi Technological University, Bawana Road, Delhi – 110042, India

E-mail: ndhingra@south.du.ac.in

Received xxxxxx
Accepted for publication xxxxxx
Published xxxxxx

Abstract

We present an ultra-compact strip to slot waveguide mode converter to facilitate excellent fundamental mode coupling between the strip and slot waveguide over a broad wavelength range. The design utilizes tapered waveguide geometries for adiabatic mode conversion. The taper profiles are optimized using a fast quasi-adiabatic approach to achieve the adiabatic mode conversion within the smallest possible length. We numerically demonstrate a conversion efficiency of 99.4% at a wavelength 1.55 μm with a variation of <1.5% over a bandwidth of 100 nm and <5% over a bandwidth of 200 nm. In addition, the design is ultra-compact with a footprint of only 1.23×3.7 μm^2 and offers a high tolerance to the possible fabrication inaccuracies.

Keywords: adiabatic tapers, quasi-adiabatic dynamics, strip to slot converter, silicon photonics

1. Introduction

The slot waveguide can confine light within a nanoscale low index gap between two high index strips and hence facilitates strong light matter interaction. This property of slot waveguide has been utilized in a number of applications including optical modulation [1], all-optical signal processing [2,3], biosensing [4-6] and absorption-based sensing [7]. The use of polymer filled slot waveguide for modulation and all-optical signal processing allows efficient use of the non-linearity in polymers [1-3]. In biosensing applications, a strong interaction between the waveguide mode and the functionalized waveguide surface of the slot waveguide results in a better sensitivity with smaller device footprint [8,9]. However, scattering losses associated with slot waveguides are, in general, higher than that of the strip waveguides. Hence, the use of slot waveguides is better suited to the abovementioned applications while for most of the other photonic integrated circuits (PICs), the strip waveguide is a better choice due to its low loss and tight confinement. Also, the efficient designs of grating couplers and edge couplers, to allow coupling of light between the photonic chip and an optical fiber, are available for the strip waveguides [10]. To benefit from the advantages of both the waveguide geometries, there is a need for efficient coupling between the fundamental mode of the strip waveguide and the slot waveguide. Direct butt coupling results in a high loss due to a large mismatch between the fundamental mode of the strip and the slot waveguides and hence a mode converter is needed for efficient coupling of light between these waveguides. Few of the reported designs of the

 © xxxx IOP Publishing Ltd

Accepted Manuscript

strip to slot waveguide mode converters are solely based on adiabatic mode evolution [11-13]. For adiabatic mode evolution based V-shaped mode converter [11] and ultra-low loss converter [12], the coupling efficiency strongly depends upon the sharpness of the taper tip and hence poses challenges in terms of fabrication. The other design [13] based on adiabatic mode evolution is improved in terms of fabrication tolerance; however, the device footprint can be significantly improved by the appropriate choice of the taper profiles. The other category of strip to slot converter is based on split and combine technique. This technique requires the use of an addition splitter section prior to an adiabatic mode evolution based combiner section [14-17]. The technique utilizes either a y-splitter [14] or a MMI [15-17] to split the light into two strip waveguides followed by a tapered section for adiabatic evolution of the fundamental mode of these evanescently coupled strip waveguides into the fundamental mode of the slot waveguide. The performance of these designs is limited by the 3 dB response of the y-splitter or MMI. Also, the design with y-splitter has a very large device footprint and requires integrated heaters to compensate the phase error due to the differential length between the two arms [14].

We propose a compact and broadband strip to slot waveguide mode converter based on adiabatic mode evolution. In order to achieve adiabatic mode conversion within the smallest possible length, we optimize the taper profiles using a fast quasi-adiabatic (FAQUAD) [18,19] approach. Using the optimized tapers, a very high mode conversion efficiency is obtained over a broad wavelength range with a footprint of only 1.23×3.7 μm^2. The design also offers a very high tolerance to the possible fabrication inaccuracies.

2. Design criteria and optimization

The proposed strip to slot waveguide mode converter is designed for silicon-on-insulator platform with dimensions of the slot waveguide optimized specifically for biosensing applications [9]. Cytop is chosen as the cladding material as its refractive index ($n_c = 1.3335$ at $\lambda = 1.55$ μm [20]) matches well with the aqueous solutions used for biosensing and hence makes it an appropriate choice for cladding or microfluidic channel in integrated photonic biosensors [21-23]. Optimizing the mode converter with Cytop cladding also allows the freedom to use the mode converter within the sensing region with no significant change in the conversion efficiency. The cross-sectional view of the strip and slot waveguides, with corresponding dimensions and the field profile (E_x) of the fundamental quasi-TE mode, are shown in Fig. 1. The effective indices of the TE_0 mode of the strip waveguide and slot waveguide are obtained as 2.4277 and 1.6810, respectively.

It is clearly visible from the modal profile of the slot waveguide [Fig. 1], that the electric field is confined to the low index slot between the two silicon strips which is also evident from its low effective index. For direct butt coupling between strip waveguide and slot waveguide, with dimensions given in Fig. 1, a maximum coupling efficiency of only 70% is possible.

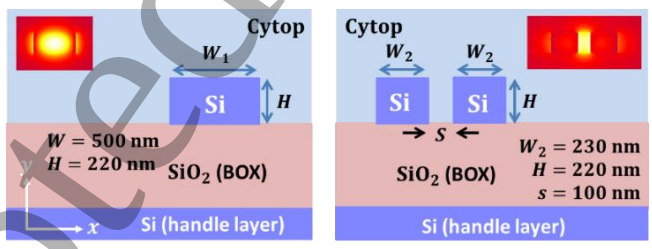


Fig. 1. Cross-sectional view of the cytop-clad silicon strip and slot waveguides with corresponding fundamental TE mode profile (E_x)

In waveguide based biosensing, the primary aim is to maximize the interaction between the waveguide mode and the functionalized waveguide surface. To find the optimum dimensions (W_2 and S) of the slot waveguide, we assume a biorecognition layer with refractive index 1.5 and thickness t directly above the waveguide surface (see inset of Fig. 2). As the sensitivity of a bio-sensor depends on the interaction between the propagating mode and the bio-recognition layer, the fill factor (FF), which represents the fraction of mode power confined in a given region, is calculated in the bio-recognition layer. For three values of S, equal to 80 nm, 100 nm, and 120 nm, the variation in the FF is shown as a function of W_2 in Fig. 2(a) and Fig. 2(b) for t = 5 nm and t = 10 nm, respectively. For each value of S, there exists an optimum value of W_2, for which the FF of the mode is maximum within the biorecognition layer. The FF increases with a decrease in S, however, the minimum value of S, which can be used is limited by the fabrication constraints. A spacing equal to or greater than 100 nm can be easily fabricated using

Accepted Manuscript

the 193 nm immersion photolithography and hence, we fix S = 100 nm. For S = 100 nm, the FF does not change much for values of W_2 between 210 nm and 230 nm. Hence, we choose W_2 = 230 nm for the ease of fabrication. From Fig. 2(a) and Fig. 2(b), it can be observed that the optimized dimensions of the slot waveguide do not change significantly with t.

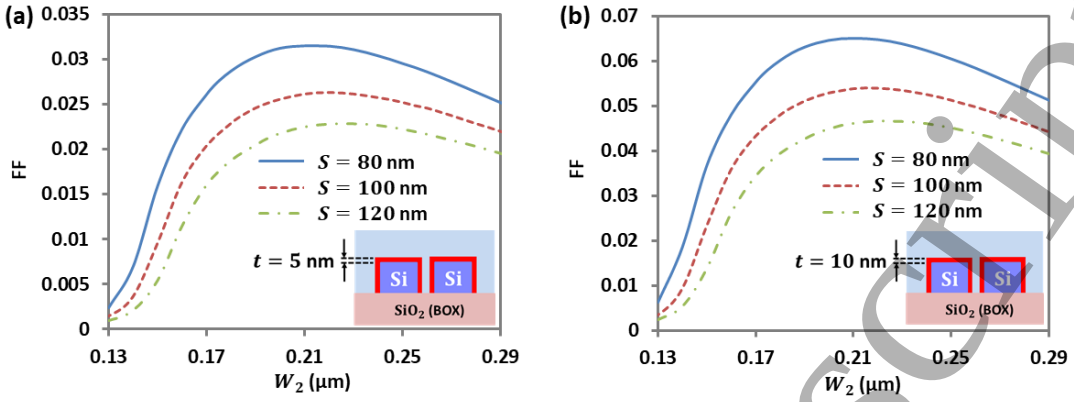


Fig. 2 Fill factor for the fundamental quasi-TE mode in the biorecognition layer of thickness (a) 5 nm and (b) 10 nm

The top view of the proposed strip to slot waveguide mode converter, consisting of two tapered sections (I and II), is shown in Fig. 3. In section I, a 230 nm strip is gradually coupled to the input strip waveguide by reducing the spacing from 500 nm to 100 nm. This ensures adiabatic coupling between the fundamental mode of the strip waveguide and the fundamental supermode of the evanescently coupled waveguides with a spacing of 100 nm at interface II. In section II, the width of the input strip waveguide is adiabatically reduced from 500 nm to 230 nm to form a slot waveguide.

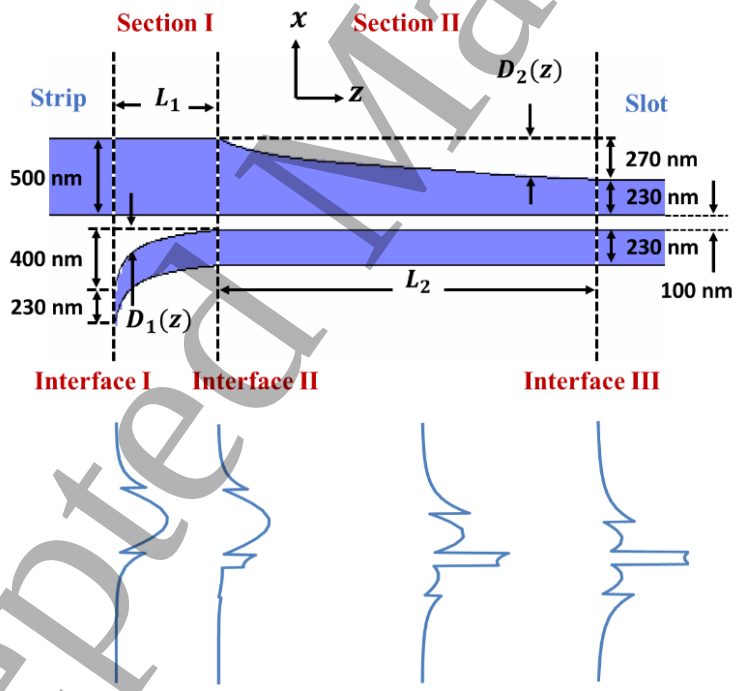


Fig. 3. Top view of the proposed design of mode converter with two tapered sections (I and II) along with the modal profiles of the fundamental quasi-TE mode at different positions along the length of the converter

The field profile (E_x), shown at different positions along the converter, shows the evolution of the fundamental mode as it propagates through the device. It is also possible to directly connect the 500 nm strip at interface II without the need for section I; however, this will result in the power loss at the interface II due to a small difference in the modal profiles of the fundamental mode of the strip waveguide and the fundamental supermode on the right hand side of interface II. Also, without the section I, the fabrication imperfections at the front edge of the 230 nm waveguide may affect the performance. Hence, we choose to design the mode converter with both tapered sections

Accepted Manuscript

I and II.

The tapering in both the sections should be slow enough to support adiabatic mode evolution, i.e., if the fundamental mode is excited at the input interface I, the power should remain in the fundamental mode as the geometry changes from a strip waveguide to a slot waveguide. A linear taper, in general, requires large length to ensure such an adiabaticity and hence results in a large device footprint. Different taper shapes [24,25] and techniques [26,27] to obtain taper profiles to support adiabatic mode evolution with shorter length of tapered sections have been suggested. However, none of these techniques provide a simple recipe to obtain the optimum taper profile for a specific application. In quantum mechanics, where the Hamiltonian changes with time, the adiabatic theorem defines an adiabaticity condition [28] in which a particle in the n^{th} eigenstate of the initial Hamiltonian remains in the n^{th} eigenstate of the changing Hamiltonian. An approach called fast quasi-adiabatic (FAQUAD) [29] can be used to optimize a single control parameter in the changing Hamiltonian of a quantum system to perform the process as fast as possible and still keeping it as adiabatic as possible. The optimization of the control parameter is done by maintaining the adiabaticity parameter to a small constant value over the entire process duration. The FAQUAD approach has been successfully translated to waveguide geometries, which vary along the propagation direction. It has been used in designing a variety of integrated photonic devices, based on z-varying optical waveguide structures, including a mode-sorting asymmetric Y-junction [30], a polarization splitter-rotator [31] and a 3-dB coupler [32]. For optical waveguides, the adiabaticity parameter (c_{nm}) for controlling the power transfer from m^{th} mode to n^{th} mode along a z-dependent waveguide geometry is given by [18,19]

$$c_{nm}(z) = \left| \frac{\iint \left[\frac{\mathrm{d}\boldsymbol{E_m}}{\mathrm{d}z} \times \boldsymbol{H_n} \right] \cdot \hat{z}\, \mathrm{d}x\mathrm{d}y}{(\beta_m - \beta_n)} \right| \tag{1}$$

where, $\mathrm{d}\boldsymbol{E_m}/\mathrm{d}z$ represents the derivative of the electric field of the m^{th} mode at a position z along the propagation length, $\boldsymbol{H_n}$ represents the magnetic field of the n^{th} mode with β_m and β_n the propagation constants of m^{th} mode and n^{th} mode, respectively. To ensure the adiabatic mode evolution, the adiabaticity parameter $c_{nm}(z)$ should be less than a desired small value ε throughout the length. For the proposed design, the FAQUAD protocol is used to distribute the adiabaticity uniformly over the entire taper length by controlling a single taper parameter. In our design, this taper parameter $D(z)$ is equals to $D_1(z)$ in section I and $D_2(z)$ in section II, as shown in Fig. 3. The Eq. (1) can also be represented in terms of $D(z)$ as

$$c_{nm}(z) = \left| \frac{\iint \left[\frac{\mathrm{d}\boldsymbol{E_m}}{\mathrm{d}D} \times \boldsymbol{H_n} \right] \cdot \frac{\mathrm{d}D}{\mathrm{d}z} \hat{z}\, \mathrm{d}x\mathrm{d}y}{(\beta_m - \beta_n)} \right| = F_{nm}(D) \frac{\mathrm{d}D}{\mathrm{d}z} \tag{2}$$

where, $F_{nm}(D)$ is given by

$$F_{nm}(z) = \left| \frac{\iint \left[\frac{\mathrm{d}\boldsymbol{E_m}}{\mathrm{d}D} \times \boldsymbol{H_n} \right] \cdot \hat{z}\, \mathrm{d}x\mathrm{d}y}{(\beta_m - \beta_n)} \right| \tag{3}$$

For ideal strip to slot mode conversion, the fundamental mode excited at interface I should propagate adiabatically throughout the length of the tapers to transform into the fundamental mode of the slot waveguide. In principle, a waveguide supports a few guided modes and infinite number of radiation modes. However, the radiation modes can be discretized to a finite number of modes using appropriate computational boundaries. The mode solvers in Photon Design [33] use reflecting boundaries to obtain the complete basis set in terms of finite number of guided and radiation modes. Hence, to ensure the adiabatic mode evolution of the fundamental mode, the sum of the adiabaticity parameters over all the N higher order modes, i.e., $c(z) = \sum_{n=2}^{N} c_{n1} = (dD/dz)F(D)$, where the parameter $F(D)$ is given by $F(D) = \sum_{n=2}^{N} F_{n1}(D)$, should be below a desired small value ε. The finite element method (FEM) and eigenmode expansion (EME) method [34], in Photon Design, are used for the calculation of $F(D)$. The variation in the parameter $F(D)$, represented by $F(D_1)$ for section I and $F(D_2)$ for sections

Accepted Manuscript

II, is shown as a function of D_1 and D_2 in Fig. 4(a) and Fig. 4(b), respectively. For a higher value of $F(D)$, a lower value of dD/dz is required to maintain constant value of $c(z)$ throughout the length of taper. Hence, the taper profile can be chosen in such a way that the slope of the taper is high where $F(D)$ is small and vice versa.

To find the optimum taper profiles $D(z)$, the adiabaticity parameter should be equal to a constant value ε throughout the length of the taper, i.e.,

$$\varepsilon = F(D)\frac{dD}{dz} \tag{4}$$

Integrating from 0 to z along the length of the taper results in

$$\varepsilon\int_0^z dz = \int_0^D F(D)\,dD \tag{5}$$

At $z = L$, where L is the length of the taper, $D = D_f$ and ε can be calculated as

$$\varepsilon = \frac{1}{L}\int_0^{D_f} F(D)\,dD \tag{6}$$

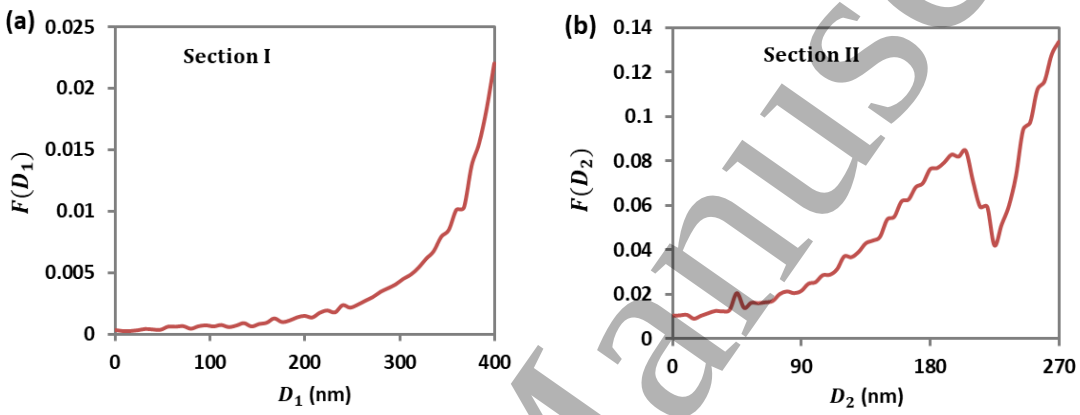


Fig. 4. (a) $F(D_1)$ as a function of D_1 [section I] and (b) $F(D_2)$ as a function of D_2 [section II]

Hence, by substituting the expression for ε from Eq. (6) into Eq. (5), the taper profile $z(D)$, which ensures a constant value of $c(z)$, can be obtained as

$$z(D) = L\frac{\int_0^D F(D)\,dD}{\int_0^{D_f} F(D)\,dD} \tag{7}$$

The optimized taper profiles for tapered sections I and II, obtained using Eq. (7), are shown in Fig. 5(a) and Fig. 5(b), respectively, along with the corresponding linear profiles for comparison. The parameter $F(D_1)$ and $F(D_2)$ and hence the taper profiles $D_1(z)$ and $D_2(z)$ are calculated at wavelength 1.55 μm. The use of these optimized taper profiles with constant $c(z)$ are expected to give rise to a significant reduction in the required length of the taper while maintain a low sensitivity towards wavelength and high tolerance towards the fabrication inaccuracies. The results shown in the section 4 validate these points.

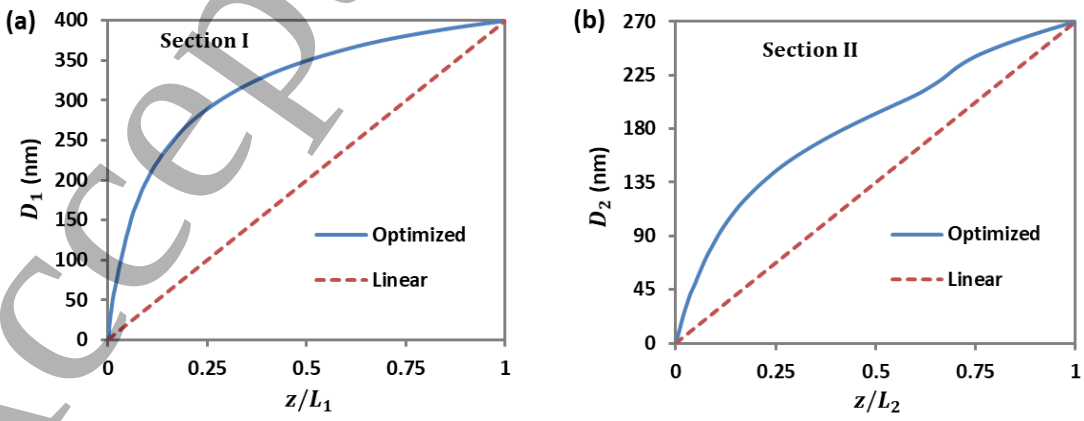


Fig. 5. Optimized taper profiles and corresponding linear profiles as a function of normalized position along the

Accepted Manuscript

length of the taper for (a) section I and (b) section II

3. Propagation through the optimized design

In order to estimate appropriate values of the length of the optimized tapers, we study the fractional power retained in the fundamental mode in propagating from interface I to interface II in section I and from interface II to interface III in section II as a function of lengths L_1 and L_2, respectively. For comparison, the corresponding study for the linear tapers is also carried out. The results shown in Fig. 6(a) and Fig. 6(b) for both linear and optimized tapers clearly show that optimized tapers perform better than the linear tapers at smaller lengths. At higher values of taper length, the conversion efficiency for both linear and optimized taper starts to converge to a maximum value. As the requirement is to obtain an efficient design with footprint as small as possible, the optimization of the taper profiles is important. For section I, although the first maximum in the Fig. 6(a) is observed at a taper length of 1.9 μm, we choose $L_1 = 0.79\,\mu\text{m}$, which provide sufficiently high value of fractional power (99.9%) in the fundamental mode at the output of section I. For section II, the length of the taper is chosen as $L_2 = 2.91\,\mu\text{m}$, corresponding to the first maximum (99.6%) in Fig. 6(b).

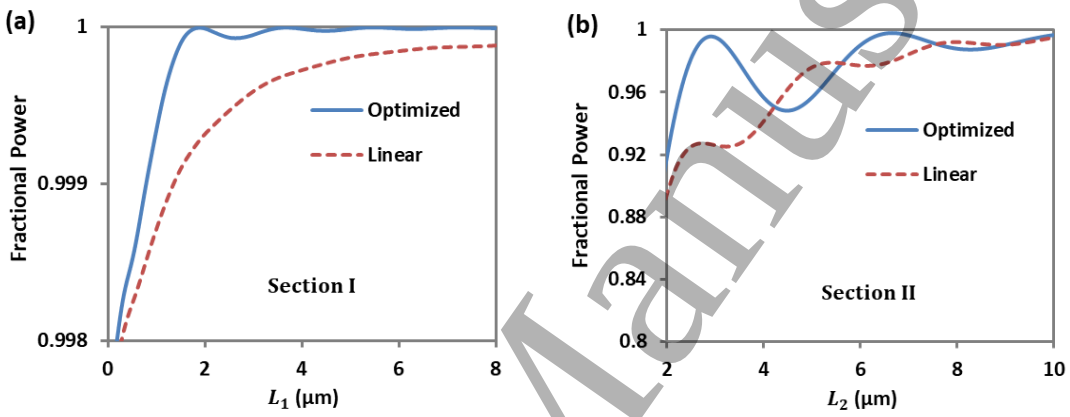


Fig. 6. Fractional output power for both optimized and linear tapers as a function of length of tapered section for individual (a) section I and (b) section II

Further, we look at the distribution of the adiabaticity parameter $c(z)$ along the length of the optimized tapers and compare it with the corresponding distribution for linear tapers of same length. The variation is shown in Fig. 7(a) and Fig. 7(b) with corresponding lengths $L_1 = 0.79\,\mu\text{m}$ and $L_2 = 2.91\,\mu\text{m}$. It can be clearly seen that the adiabaticity parameter $c(z)$ has a small value ε and is uniformly distributed over the entire length of the optimized tapers, as expected from the earlier discussion. A small noise in the $c(z)$ is due to the limited accuracy of the FEM solver. To maintain the value of $c(z)$ for the linear taper below the value obtained for the optimized taper, the slope (dD/dz) of the linear taper has to be reduced by an increase in the length of the linear taper by at least five times for section I and three times for section II.

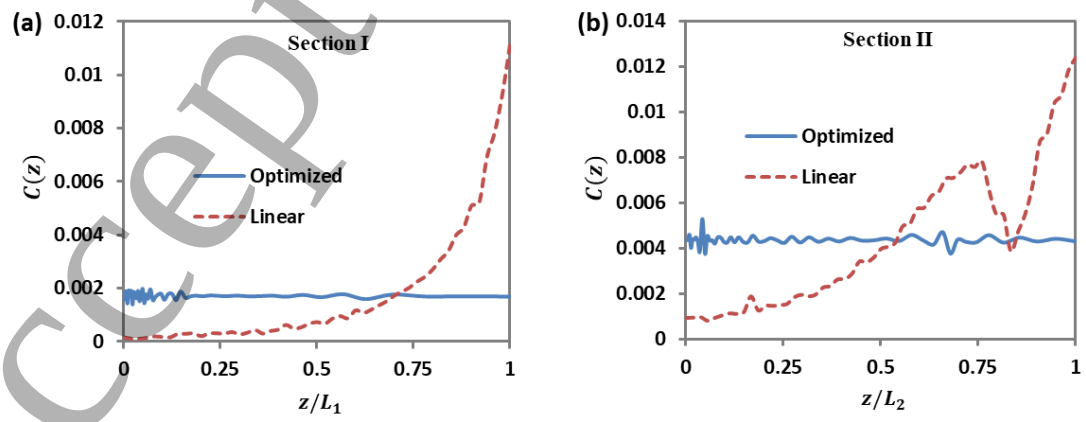


Fig. 7. (a) Distribution of parameter $c(z)$ throughout the length of taper for linear and optimized designs for (a) section I and (b) section II

Accepted Manuscript

The propagating field (E_x) through the proposed design, with $L_1 = 0.79\,\mu m$ and $L_2 = 2.91\,\mu m$, is shown in Fig. 8 for the input excitation by the fundamental strip mode at $\lambda = 1.55$ µm. It can be seen that the fundamental mode of the strip waveguide, excited at the input end, evolves adiabatically into the fundamental mode of the slot waveguide at the end of the converter. The conversion efficiency of the complete design is obtained as 99.4% at the design wavelength of 1.55 µm with a device footprint of only 1.23×3.7 µm^2.

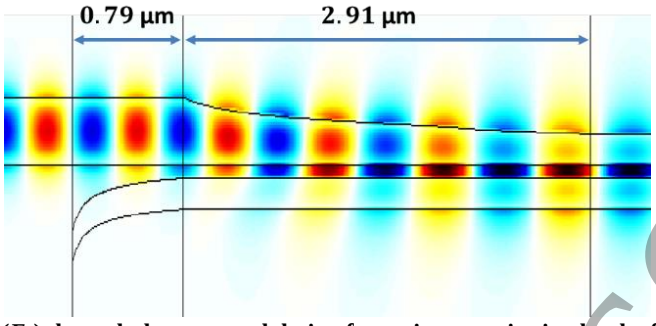


Fig. 8. Propagating field (E_x) through the proposed design for an input excitation by the fundamental mode of the strip waveguide

4. Wavelength response and fabrication tolerance

As the design utilizes the adiabatic mode evolution instead of conventional evanescent coupling, the design can be expected to be less sensitive to wavelength and fabrication variations. The mode conversion efficiency is plotted as a function of wavelength in Fig. 9. A variation of <1.5% in the conversion efficiency for a bandwidth of 100 nm and <5% for a bandwidth of 200 nm is observed which shows the broadband nature of the proposed design.

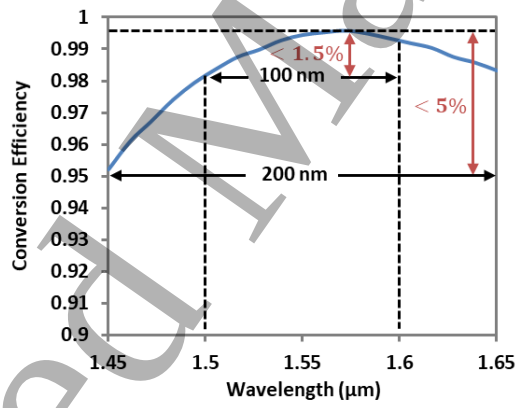


Fig. 9. Strip-to-slot waveguide fundamental mode conversion efficiency as a function of wavelength

The fabrication of the proposed design involves a single lithography and etching step which can be easily realized using the 193-nm immersion lithography followed by reactive ion etching [35]. The variation in the thickness of the device layer is entirely dependent on the SOI wafer. On the other hand, the width variation primarily depends on the lithography and etching process. Using 193-nm deep ultraviolet (DUV) lithography process, a linewidth accuracy of ±20 nm, and uniformity in the device layer thickness within ±10 nm can be easily achieved. The statistical results for a large-scale manufacturing shows the standard deviation in both the linewidth and thickness under 5 nm [36].

To study the fabrication tolerance of the proposed design, we consider a ±20 nm variation in the width and the thickness of the silicon device layer. These limits are more than sufficient to find the design tolerance of the proposed design to fabrication inaccuracies. First, we look at the conversion efficiency as a function of variation (ΔW) in width with $\Delta H = 0$ and then as a function of variation (ΔH) in height with $\Delta W = 0$ as shown in Fig. 10(a). Next, to calculate the worst-case conversion efficiency, we first vary the height and width by same magnitude simultaneously, as shown in Fig. 10(b), and then with different signs, as shown in Fig. 10(c). For the worst case, the conversion efficiency is found to be >96.5%. The results show the highly tolerant nature of the proposed strip

Accepted Manuscript

to slot waveguide mode converter to the possible fabrication inaccuracies.

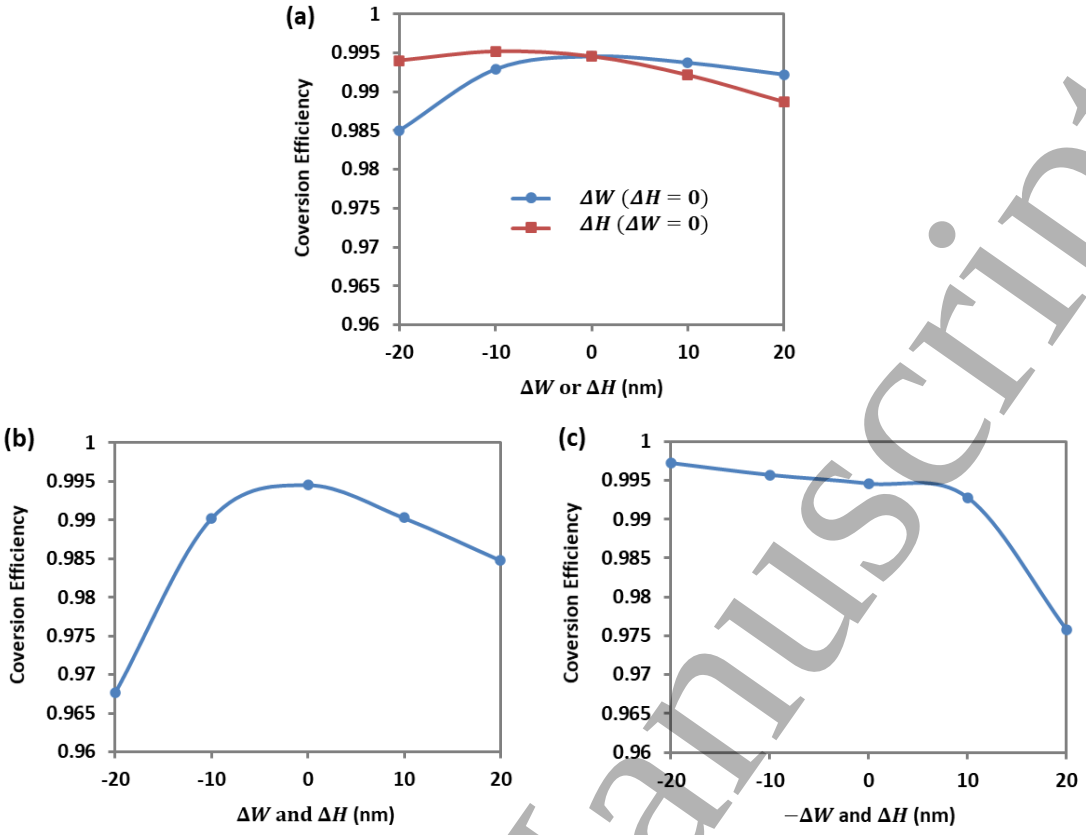


Fig. 10. Variation in the conversion efficiency as a function of deviation from the optimized parameters (a) deviation in width with optimized value of height and deviation in height with optimized value of width (b) deviation in width and height by same magnitude (c) deviation in width and height by same magnitude but with opposite signs

For the characterization of the proposed design, the grating couplers [37] need to be fabricated to allow in/out coupling from/to a single mode fibre (SMF). As the efficient designs of the grating couplers are not available for slot waveguide, an additional slot to strip converter needs to be fabricated as shown in Fig. 11(a) in order to allow coupling to the output SMF. As the loss associated with the proposed mode converter is very low, the inaccuracies in the measurement can arise due to a slight variation in the coupling efficiency of the grating couplers. Hence, to reduce the influence of such errors in the measurement, it is better to use a number of these sections in sequence [Fig. 11(b)] and to calculate the average response. To subtract the response of the grating couplers with respect to the fabricated design, a calibration structure consisting of two grating couplers connected through a strip waveguide is also need to be fabricated.

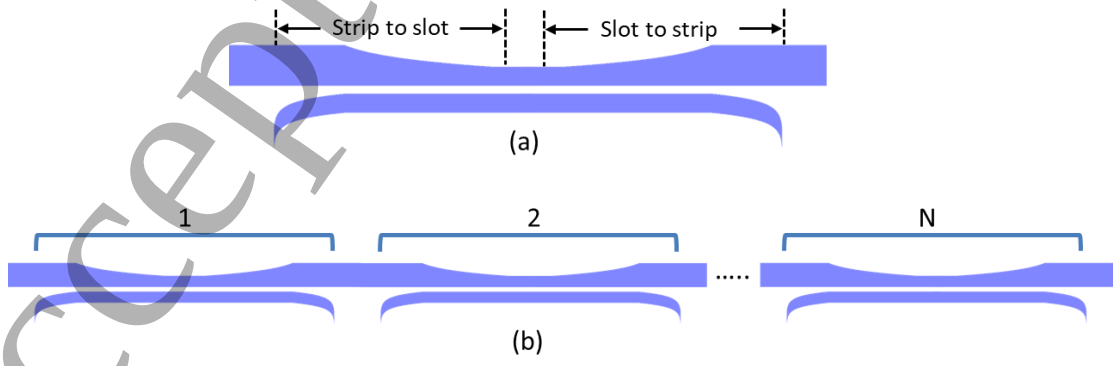


Fig. 11 (a) Mode converter pair consisting of a strip to slot converter and a slot to strip converter (b) Several cascaded pairs of mode converters

Accepted Manuscript

A tunable laser source, with operating wavelength range 1450 nm to 1650 nm, can be used as an input source. A polarization controller is required to control the polarization of the input light to allow efficient coupling from input SMF to the fundamental quasi-TE mode of the strip waveguide. To measure the wavelength response, the power can be coupled through the output grating coupler to an output SMF to observe the wavelength response using an optical spectrum analyzer.

The proposed design allows coupling between a standard strip waveguide and a symmetric slot waveguide, which confines the light within the narrow slot between the two silicon strips. However, there also exist other slot waveguide geometries including asymmetric slot waveguide [38], strip-loaded slot waveguide [1-3,39], plasmonic slot waveguide [40], and a slotted photonic crystal waveguide [41]. The mode converter for the asymmetric slot waveguide can be designed using two tapered sections by following a procedure similar to the one used for the symmetric slot waveguide. To facilitate the coupling between the strip waveguide and the other geometries of slot waveguides, the designs based on adiabatic mode evolution has been reported in the literature [39-41]. Each of these designs require certain number of tapers to facilitate the required mode coupling. The taper profiles in these designs can be optimized using the FAQUAD approach to allow efficient mode conversion with smaller device footprint.

5. Conclusion and discussion

Strip to slot waveguide mode converters are pivotal components in slot waveguide based integrated photonic devices [1-8,42]. The overall performance of these devices also depends on the performance metrics of the mode converters used to facilitate efficient fundamental mode coupling between the strip and slot waveguides. Hence, a design based on adiabatic mode evolution is proposed to obtain efficient fundamental mode coupling from a strip waveguide to a slot waveguide. A fast quasi-adiabatic approach is utilized to obtain optimized taper profiles which maintain a low constant value of adiabaticity parameter throughout the taper length. Using the optimized taper profiles, a very high (99.4%) conversion efficiency is obtained with small lengths of tapers, resulting in an ultra-compact footprint of only $1.23\times3.7\ \mu m^2$. In addition, the design is ultra-broadband with a variation of <1.5% in the conversion efficiency over a bandwidth of 100 nm and <5% over a bandwidth of 200 nm. The design also offers a very high tolerance to the fabrication inaccuracies.

Acknowledgement

This work was supported in part by the University Grant Commission, India, and in part by the University of Delhi South Campus, India.

References

1. Palmer, R., Koeber, S., Elder, D.L., Woessner, M., Heni, W., Korn, D., Lauermann, M., Bogaerts, W., Dalton, L., Freude, W. and Leuthold, J., 2014. High-speed, low drive-voltage silicon-organic hybrid modulator based on a binary-chromophore electro-optic material. *Journal of Lightwave Technology*, *32*(16), pp.2726-2734.
2. Koos, C., Jacome, L., Poulton, C., Leuthold, J. and Freude, W., 2007. Nonlinear silicon-on-insulator waveguides for all-optical signal processing. *Optics Express*, *15*(10), pp.5976-5990.
3. Koos, C., Vorreau, P., Vallaitis, T., Dumon, P., Bogaerts, W., Baets, R., Esembeson, B., Biaggio, I., Michinobu, T., Diederich, F. and Freude, W., 2009. All-optical high-speed signal processing with silicon–organic hybrid slot waveguides. *Nature photonics*, *3*(4), p.216.
4. Carlborg, C.F., Gylfason, K.B., Kaźmierczak, A., Dortu, F., Polo, M.B., Catala, A.M., Kresbach, G.M., Sohlström, H., Moh, T., Vivien, L. and Popplewell, J., 2010. A packaged optical slot-waveguide ring resonator sensor array for multiplex label-free assays in labs-on-chips. *Lab on a Chip*, *10*(3), pp.281-290.
5. Liu, Q., Tu, X., Kim, K.W., Kee, J.S., Shin, Y., Han, K., Yoon, Y.J., Lo, G.Q. and Park, M.K., 2013. Highly sensitive Mach–Zehnder interferometer biosensor based on silicon nitride slot waveguide. *Sensors and Actuators B: Chemical*, *188*, pp.681-688.
6. Luan, E., Shoman, H., Ratner, D.M., Cheung, K.C. and Chrostowski, L., 2018. Silicon photonic biosensors using label-free detection. *Sensors*, *18*(10), p.3519.
7. Butt, M.A., Khonina, S.N. and Kazanskiy, N.L., 2018. Silicon on silicon dioxide slot waveguide evanescent field gas absorption sensor. *Journal of Modern optics*, *65*(2), pp.174-178.
8. Dell'Olio, F. and Passaro, V.M., 2007. Optical sensing by optimized silicon slot waveguides. *Optics Express*, *15*(8), pp.4977-4993.
9. Milvich, J., Kohler, D., Freude, W. and Koos, C., 2018. Surface sensing with integrated optical waveguides: a design guideline. *Optics express*, *26*(16), pp.19885-19906.
10. Marchetti, R., Lacava, C., Carroll, L., Gradkowski, K. and Minzioni, P., 2019. Coupling strategies for silicon photonics integrated chips. *Photonics Research*, *7*(2), pp.201-239.
11. Wang, Z., Zhu, N., Tang, Y., Wosinski, L., Dai, D. and He, S., 2009. Ultracompact low-loss coupler between strip and slot waveguides. Optics letters, 34(10), pp.1498-1500.
12. Feng, N.N., Sun, R., Kimerling, L.C. and Michel, J., 2007. Lossless strip-to-slot waveguide transformer. *Optics letters*, *32*(10), pp.1250-1252.

Accepted Manuscript

13. Palmer, R., Alloatti, L., Korn, D., Heni, W., Schindler, P.C., Bolten, J., Karl, M., Waldow, M., Wahlbrink, T., Freude, W. and Koos, C., 2013. Low-loss silicon strip-to-slot mode converters. *IEEE Photonics Journal*, *5*(1), pp.2200409-2200409.
14. Mere, V., Kallega, R. and Selvaraja, S.K., 2018. Efficient and tunable strip-to-slot fundamental mode coupling. *Optics express*, *26*(1), pp.438-444.
15. Deng, Q., Liu, L., Li, X. and Zhou, Z., 2014. Strip-slot waveguide mode converter based on symmetric multimode interference. *Optics letters*, *39*(19), pp.5665-5668.
16. Xiao, J. and Xu, Y., 2016. Ultracompact and broadband silicon-based strip-to-slot mode converter. IEEE Photonics Technology Letters, 28(13), pp.1414-1417.
17. Deng, Q., Yan, Q., Liu, L., Li, X., Michel, J. and Zhou, Z., 2016. Robust polarization-insensitive strip-slot waveguide mode converter based on symmetric multimode interference. *Optics express*, *24*(7), pp.7347-7355.
18. Martínez-Garaot, S., Muga, J.G. and Tseng, S.Y., 2017. Shortcuts to adiabaticity in optical waveguides using fast quasiadiabatic dynamics. *Optics express*, *25*(1), pp.159-167.
19. Chung, H.C., Martínez-Garaot, S., Chen, X., Muga, J.G. and Tseng, S.Y., 2019. Shortcuts to adiabaticity in optical waveguides. *EPL (Europhysics Letters)*, *127*(3), p.34001.
20. Cytop technical brochure, AGC Inc., Tokyo, 2018, https://www.agcchem.com/documentation/technical-data/cytop/
21. Wong, W.R., Fan, H., Adikan, F.R.M. and Berini, P., 2018. Multichannel long-range surface plasmon waveguides for parallel biosensing. Journal of Lightwave Technology, 36(23), pp.5536-5546.
22. Leosson, K. and Agnarsson, B., 2012. Integrated biophotonics with CYTOP. *Micromachines*, *3*(1), pp.114-125.
23. Krupin, O., Wong, W.R., Béland, P., Adikan, F.R.M. and Berini, P., 2016. Long-range surface plasmon-polariton waveguide biosensors for disease detection. Journal of Lightwave Technology, 34(20), pp.4673-4681.
24. Ren, G., Chen, S., Cheng, Y. and Zhai, Y., 2011. Study on inverse taper based mode transformer for low loss coupling between silicon wire waveguide and lensed fiber. *Optics Communications*, *284*(19), pp.4782-4788.
25. Khoo, E.H., Liu, A.Q. and Wu, J.H., 2005. Nonuniform photonic crystal taper for high-efficiency mode coupling. Optics express, 13(20), pp.7748-7759.
26. Sun, X., Liu, H.C. and Yariv, A., 2009. Adiabaticity criterion and the shortest adiabatic mode transformer in a coupled-waveguide system. *Optics letters*, *34*(3), pp.280-282.
27. Horth, A., Cheben, P., Schmid, J.H., Kashyap, R. and Quitoriano, N.J., 2016. Ideal, constant-loss nanophotonic mode converter using a Lagrangian approach. *Optics express*, *24*(6), pp.6680-6688.
28. Schiff, L.I., 1998. Quantum Mechanics, 3rd ed. (McGraw- Hill).
29. Martínez-Garaot, S., Ruschhaupt, A., Gillet, J., Busch, T. and Muga, J.G., 2015. Fast quasiadiabatic dynamics. *Physical Review A*, *92*(4), p.043406.
30. Chung, H.C., Lee, K.S. and Tseng, S.Y., 2017. Short and broadband silicon asymmetric Y-junction two-mode (de) multiplexer using fast quasiadiabatic dynamics. *Optics express*, *25*(12), pp.13626-13634.
31. Chung, H.C. and Tseng, S.Y., 2018. Ultrashort and broadband silicon polarization splitter-rotator using fast quasiadiabatic dynamics. *Optics express*, *26*(8), pp.9655-9665.
32. Hung, Y.J., Li, Z.Y., Chung, H.C., Liang, F.C., Jung, M.Y., Yen, T.H. and Tseng, S.Y., 2019. Mode-evolution-based silicon-on-insulator 3 dB coupler using fast quasiadiabatic dynamics. *Optics letters*, *44*(4), pp.815-818.
33. FIMMWAVE/FIMMPROP, Photon Design Ltd, http://www.photond.com.
34. Gallagher, D.F. and Felici, T.P., 2003, June. Eigenmode expansion methods for simulation of optical propagation in photonics: pros and cons. In *Integrated optics: devices, materials, and technologies VII* (Vol. 4987, pp. 69-82). International Society for Optics and Photonics.
35. Selvaraja, S.K., Winroth, G., Locorotondo, S., Murdoch, G., Milenin, A., Delvaux, C., Ong, P., Pathak, S., Xie, W., Sterckx, G. and Lepage, G., 2014, April. 193 nm immersion lithography for high-performance silicon photonic circuits. In Optical Microlithography XXVII (Vol. 9052, p. 90520F). International Society for Optics and Photonics.
36. Lu, Z., Jhoja, J., Klein, J., Wang, X., Liu, A., Flueckiger, J., Pond, J. and Chrostowski, L., 2017. Performance prediction for silicon photonics integrated circuits with layout-dependent correlated manufacturing variability. *Optics express*, *25*(9), pp.9712-9733.
37. Marchetti, R., Lacava, C., Carroll, L., Gradkowski, K. and Minzioni, P., 2019. Coupling strategies for silicon photonics integrated chips. *Photonics Research*, 7(2), pp.201-239.
38. Spott, A., Baehr-Jones, T., Ding, R., Liu, Y., Bojko, R., O'Malley, T., Pomerene, A., Hill, C., Reinhardt, W. and Hochberg, M., 2011. Photolithographically fabricated low-loss asymmetric silicon slot waveguides. *Optics express*, 19(11), pp.10950-10958.
39. Liu, Y., Baehr-Jones, T., Li, J., Pomerene, A. and Hochberg, M., 2011. Efficient strip to strip-loaded slot mode converter in silicon-on-insulator. *IEEE Photonics Technology Letters*, 23(20), pp.1496-1498.
40. Chen, C.T., Xu, X., Hosseini, A., Pan, Z., Subbaraman, H., Zhang, X. and Chen, R.T., 2015. Design of highly efficient hybrid Si-Au taper for dielectric strip waveguide to plasmonic slot waveguide mode converter. *Journal of Lightwave Technology*, 33(2), pp.535-540.
41. Zhang, X., Subbaraman, H., Hosseini, A. and Chen, R.T., 2014. Highly efficient mode converter for coupling light into wide slot photonic crystal waveguide. *Optics express*, 22(17), pp.20678-20690.
42. Subbaraman, H., Xu, X., Hosseini, A., Zhang, X., Zhang, Y., Kwong, D. and Chen, R.T., 2015. Recent advances in silicon-based passive and active optical interconnects. *Optics express*, 23(3), pp.2487-2511.

Accepted Manuscript

Journal Pre-proof



Energy efficient transmission trends towards future green cognitive radio networks (5G): Progress, taxonomy and open challenges

Akanksha Srivastava, Mani Shekhar Gupta, Gurjit Kaur

PII: S1084-8045(20)30234-4

DOI: https://doi.org/10.1016/j.jnca.2020.102760

Reference: YJNCA 102760

To appear in: *Journal of Network and Computer Applications*

Received Date: 7 January 2020

Revised Date: 21 April 2020

Accepted Date: 23 June 2020

Please cite this article as: Srivastava, A., Gupta, M.S., Kaur, G., Energy efficient transmission trends towards future green cognitive radio networks (5G): Progress, taxonomy and open challenges, *Journal of Network and Computer Applications* (2020), doi: https://doi.org/10.1016/j.jnca.2020.102760.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier Ltd.

Energy Efficient Transmission Trends Towards Future Green Cognitive Radio Networks (5G): Progress, Taxonomy and Open Challenges

Akanksha Srivastava, Mani Shekhar Gupta and Gurjit Kaur

***Abstract*—With the growing cognizance about environmental concern and global warming-related to communication technologies, the researchers have been seeking some solutions to diminish the consumption of energy in the telecommunication industry. There is a remarkable advancement in mobile communication from simple voice-based devices to ubiquitous data-hungry smartphones. Nearly, 8 billion mobile subscribers are expected to be added by 2030 and also required an extra spectrum to serve. The existing static spectrum allocation-based technologies are not in a position to fulfill this extra spectrum requirement and handle this future traffic load. This volatile evolution of global traffic data urges research attention globally and can be handled by future cognitive radio networks. There is also a demand for providing fast speed and seamless services for which operators need to deploy more base stations continuously and increase the transmitting antenna power. As a result, a drastic hike in emissions of carbon dioxide into the environment, and exposure to harmful radiations in large amounts which is also authentically harmful to humans, animals, and birds. This increased energy consumption and scarcity of resources have made the energy as a geopolitical issue for 21 century. In this direction, this work contributes by introducing cognitive-based green communication technology to ensure the environmental and health concerns caused due to hike in the CO_2 level. In the second phase, an enlightening survey on futuristic approaches for making the future wireless networks green is covered with pros and cons. In the third phase, this paper provides a framework of research challenges with ongoing project activities on green communication which further requires attention from the research community.**

***Index Terms*—Green communication, Carbon dioxide emissions, Carrier aggregation, Massive MIMO, D2D communications, Green IoT, Cognitive radio network.**

I. Introduction

In the last few years, the mobile telecommunication industry has increased exponentially. The call for wireless applications is growing because of the astronomical boom of mobile subscribers, multimedia applications, and data applications. In this work, the word green cognitive radio (CR) communication is used to express an efficient method of the congregation of energy-efficient technologies at different levels to reduce the unfavorable effects of technology on the environment. Globally, by the year 2022, the internet user population will be increased from 45.3% to 59.7% [1]. In addition to this, Fig. 1 depicts the growth rate of different parameters from 2018 to 2025 [2]. The nuts and bolts of a telecommunication network are a mobile node, access network and a backbone core network. Base Stations (BSs) are the key component of the radio access network that connects mobile phone devices to a core network. So, with a growth rate in the number of mobile subscribers per day, the growth rate of BSs is also increasing. According to [3] it is expected that during period $2017-2021$, the BS market of Long Term Evaluation (LTE) system will increase by more than 17 percent compound annual growth rate (CAGR). Available literature work centering on improving both quality of service (QoS) and quality of experience (QoE), while ignoring the increased demand for energy for the mobile communication system. This increasing energy demand has motivated us to work on the subject of cognitive-based green communication with the objective of energy-efficient wireless communication. Based on our studies, we found two major goals of the green communications network. The first is to reduce the CO_2 emission from the system and second is to save energy by reducing transmitting power which will lead to reducing the operational expenditures (OPEX). The off-grid sites (to serve in remote areas) are the main cause of CO_2 emission. Mostly diesel-power generators are required for working on such sites. Presently, as the number of mobile subscribers is increasing, nearly 43% of the total CO_2 emission is contributed by the mobile communication system and this value will increase by 51% of the total in 2020 [4]. Due to the different segments of the mobile communication system, the CO_2 level is continuously increasing in the atmosphere. Fig. 2 represents the level of CO_2 emissions by a different segment of the communication system during 2010, 2015 and in 2020. The measure of data traffic related to a mobile node is expected to touch nearly 621.90 million Tbytes/year by 2025 while in the year 2012, it was 45 million Tbytes/year [5]. The electrical energy consumption by the ICT will be increased from 611 to $1,752$ Tera-watt hours by 2030 [6]. This rapid energy consumption rate by wireless communication networks will cause a serious problem in the future if no steps are taken to change the existing trend.

A. Motivation for Green Communication

The telecommunication industry expressions a serious energy consumption challenge. This industry is growing rapidly to serve the smartphone user anywhere, anytime and anyone. The simple reason is that every human is using dual a SIM card enabled phone. At the end of 2019, in the Asia Pacific there were 2.9 billion (nearly the world's half population) mobile subscribers [7]. India and China are the most populous

Journal Pre-proof

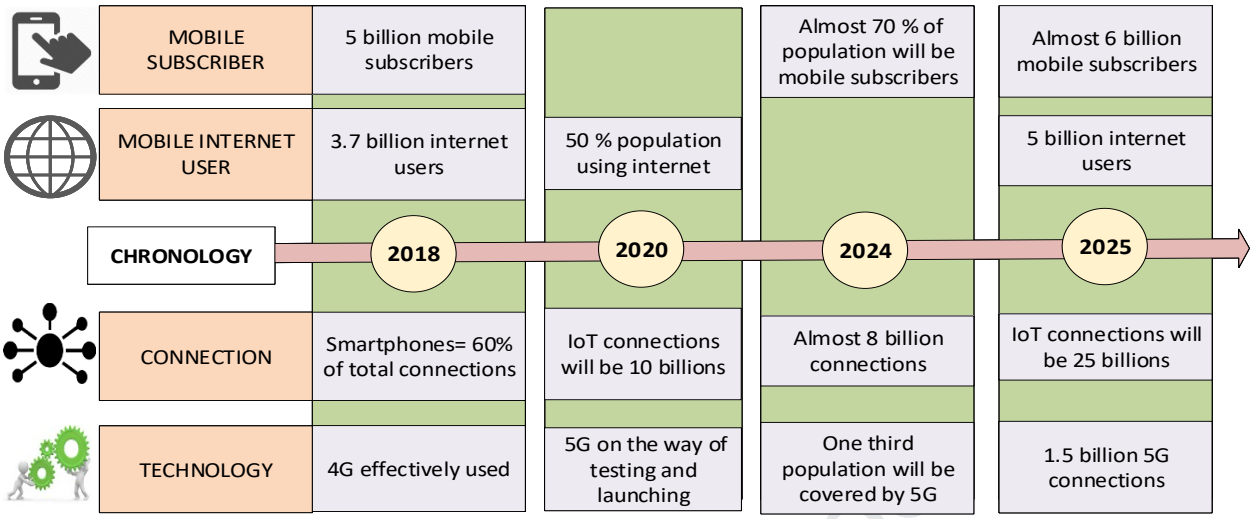


Fig. 1: Breakthrough in the field of telecommunication industry

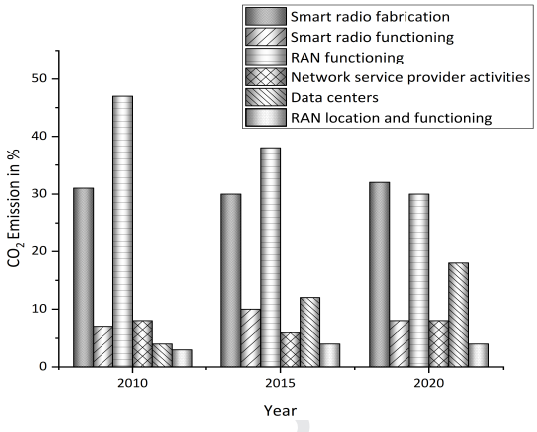


Fig. 2: Variation in carbon footprint produced by various parts of mobile communication system

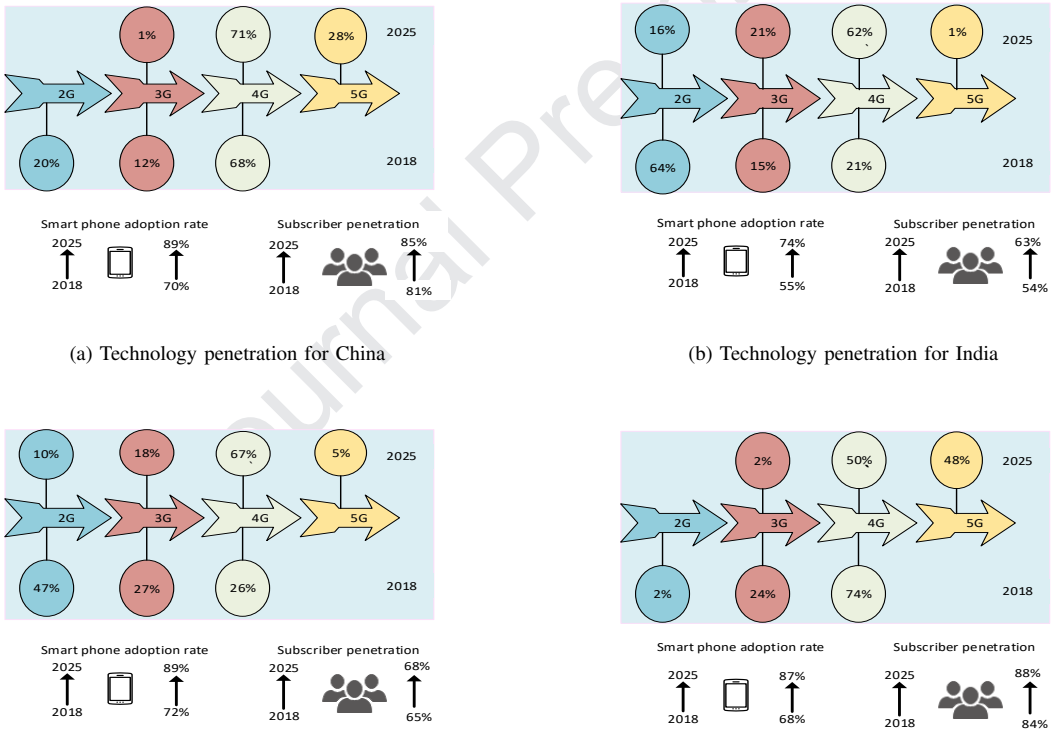
countries, which jointly covers nearly two-thirds of total mobile users. Nearly, 71.9 million mobile users are expected to be added in China by 2025, which is nearly 1.2 billion (85%) of the total population [8]. On the other hand, in India, it is expected to be nearly 207.9 million (22% of total Asia Pacific) by 2025. So, this makes a remarkable improvement in digital literacy and advanced mobile phones. In 2018, Indonesia was developed as a third-largest mobile phone market in the world (behind India and China) [9]. The smartphone adoption rate is to be expected very high (nearly 90% of total connections) for Indonesia by the end of 2025. Fig. 3 depicts the technology penetration of the four fastest-growing countries. Fig shows the percentage of the population which was using the different wireless generation (2G, 3G, 4G and 5G) in 2018 and how much this percentage value will be changed when we will reach 2025. In China and Japan, the 2G population will completely remove till 2025, on the other hand, some populations will present in India and Indonesia. In addition to this, Fig. 4 represents the addition of new mobile users globally till June 2019 in million [10] and Fig. 5 shows the mobile broadband adoption rate for different countries from 2015 to 2020.

The wireless communication networks have experienced a prodigious evolution from 1G to 5G networks. Nowadays 4G is effectively in use and 5G on the way of testing and launching. Recently, during the testing of the 5G network, 100s of birds dropped from the sky at Hague in the Netherlands [11]. The 5G antenna was situated at the Dutch railway station, to examine how large the 5G range was. At once, birds fell dead from the sky and the ducks that were swimming in nearby ponds seemed to respond very strangely. To escape from radiation, they were instantaneously inserting their heads inside water whereas some flew away. If they all had a healthy body, healthy blood, no effect of any virus and bacterial infection then the reasonable explanation behind this incident is using of microwaves in 5G having a serious impact on birds' hearts. It highly resonates with erratic pulsed microwaves (millions per second) which harm body organs [12]. Table II represents the upgradation of the various generation of wireless networks. This table reflects different characteristics, to discuss a tangible contrast among the generations of wireless networks. In 5G a drastic growth in the CO_2 emissions is presented.

To overcome this problem a pathway is designed in Fig. 6 which represents greenhouse gas (GHG) deductions across different sectors to reach 2030. Fig. shows that the amount of CO_2 was present almost 440 MMT in 2015 and the target is to reduce this value to 261 MMT till 2030. To curb such enormous emissions, various technologies have been discussed in this paper. This paper focuses on the mobile sector and provides comprehensive data on changing trends of information and communication technology (ICT) on the

TABLE I: Table of abbreviations

Abbreviation	Description	Abbreviation	Description
ASIC	Application-specific integrated circuit	AMPS	Advanced Mobile Phone Service
ATIS	Alliance for Telecommunications Industry Solutions	BDMA	Beam Division Multiple Access
CPLD	Complex Programmable Logic Device	CA	Carrier Aggregation
CAGR	Compound Annual Growth Rate	CRN	Cognitive Radio Network
D2D	Device-to-Device	EMR	Electromagnetic Radiation
ETSI	European Technical Standards Institute	FBMC	Filter Bank Multi Carrie
FPGA	Field Programmable Gate Array	GHG	Greenhouse Gas
IARC	International Agency for Research on Cancer	ITU	International Telecommunication Union
LTE-A	Long Term Evolution Advanced	MIMO	Multiple-Input Multiple-Output
OFDM	Orthogonal Frequency Division Multiplexing	PA	Power Amplifier
OPEX	Operational Expenditures	SAR	Specific Absorption Rate
SIM	Subscriber Identification Module	SM	Spatial Multiplexing
UDN	Ultra Dense Network	UMTS	Universal Mobile Telecommunications System
V2X	Vehicle to Everything	WHO	World Health Organization



(a) Technology penetration for China

(b) Technology penetration for India

(c) Technology penetration for Indonesia

(d) Technology penetration for Japan

Fig. 3: Technology penetration of four fastest growing countries in Asia Pacific from 2018 to 2025.

global platform. Emission of CO_2 is required to be declined.

B. Key contribution of survey

This work coveres, an up-to-date survey and technical aspects of advances which is providing future research directions to the foremost study on energy-efficient green communication in cognitive radio networks (CRNs). The key contributions of this work can be outlined as follows:

- A complete survey on green wireless communication networks have been covered in a single paper, with consideration of energy-efficient transmission.
- The futuristic technologies like cognitive radio, carrier aggregation, Terahertz communication, Internet of Things (IoT), massive MIMO (multiple-input multiple-output) and mm wavelength are briefly reviewed to prepare for advancing recent research contributions. These future

TABLE II: Upgradation and contrast among various generations of wireless networks in green communication aspect

Technology	1G	2G	3G	4G	5G
Frequency band	800 Mhz	850/900/1800/1900 Mhz	1900 /2100 Mhz	1.8/2.6 Ghz	30-300 Ghz
Power density ($Watt/meter^2$)	4	4.5-9	4.5-10	10	10
Power level of MS	Low	GSM 850/900 :35 dBm GSM 1800/1900 :22-37 dBm	20-32 dBm	23 dBm	High
Power level of BS	Low	Macro 46 dBm, Micro 13-30 dBm	32-38 dBm	44-49 dBm	High
CO_2 emission	23 Mt CO_2	30 Mt CO_2	85 Mt CO_2	171 Mt CO_2	236 Mt CO_2
Carbon footprint per mobile	10kg	15kg	20kg	24kg	30kg
Power optimization methods	Modulation techniques such as SSB-FM used to save power.	Moved towards digital communication, Sleep mode at BS and small powered radio signals used.	Sleep mode at BS and relay technique is used.	In OFDMA and SC-FDMA, PAPR minimization used small cells concept used.	D2D, UDN, C-RAN, IoT used.
Date rate	2.4 Kbps	100 Kbps	2.4-30 Mbps	100-200 Mbps-1.5-3 Gbps	10-50 Gbps
Access technology	AMPS	GSM-GPRS	WCDMA, UMTS, CDMA 2000, HSPDA	LTE-A, WiMAX	BDMA, FBMC
Spectral efficiency (bps/Hz)	0.46	1.4	2.7	4.27	10.6
MIMO/Massive MIMO	SISO	SISO	SIMO, MISO	MIMO	Massive MIMO
WLAN Technology	802.11	802.11b	802.11ag, 802.11n	802.11ac, 802.11an	802.11ad
BS density per square kilometer	Very low	Low	4-5 BSs	8-10 BSs	40-50 BSs
OPEX	-	-	Low	High	High
SAR value	-	-	High	Higher	Expected to reduce
Used antenna	Low frequency antenna	High frequency antenna	High gain and clip antenna	Slot and Patch antenna	Phased array antennas

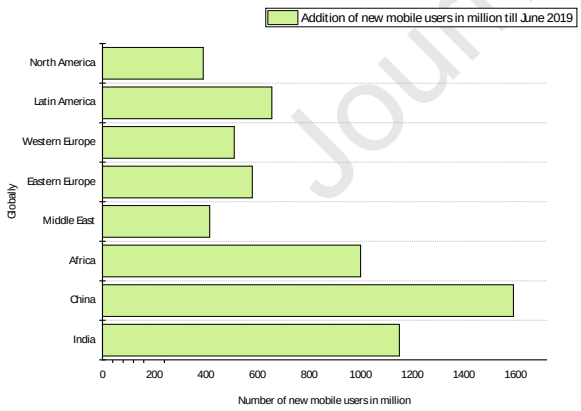


Fig. 4: Addition of new mobile users globally in million till June 2019

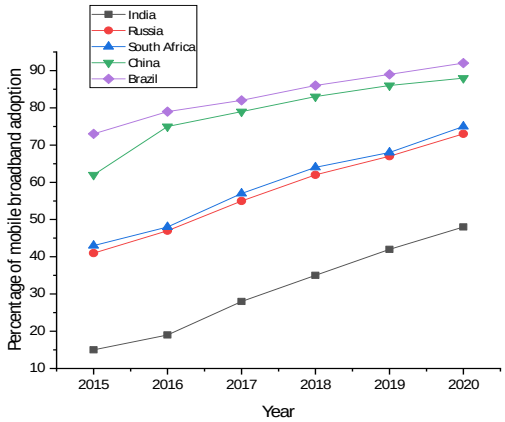


Fig. 5: Mobile broadband acquisition rate

technologies for 5G are proficient for improving the energy efficiency of the next generation wireless networks with high QoS and low GHG emissions.

- Starting with the impact of the increasing number of mobile users globally, their demands for different services and their adverse effects on the environment are discussed in detail.
- Several energy efficiency metrics used for the calculation of different parameters at different levels of communication system have been illustrated.
- Various future research directions and challenges in the pathway of green network are also covered to handle the implementation challenges.

C. Paper Organization

The progression of the work is maintained as in Fig. 7. In Section II, the recent facts and figures are briefly discussed

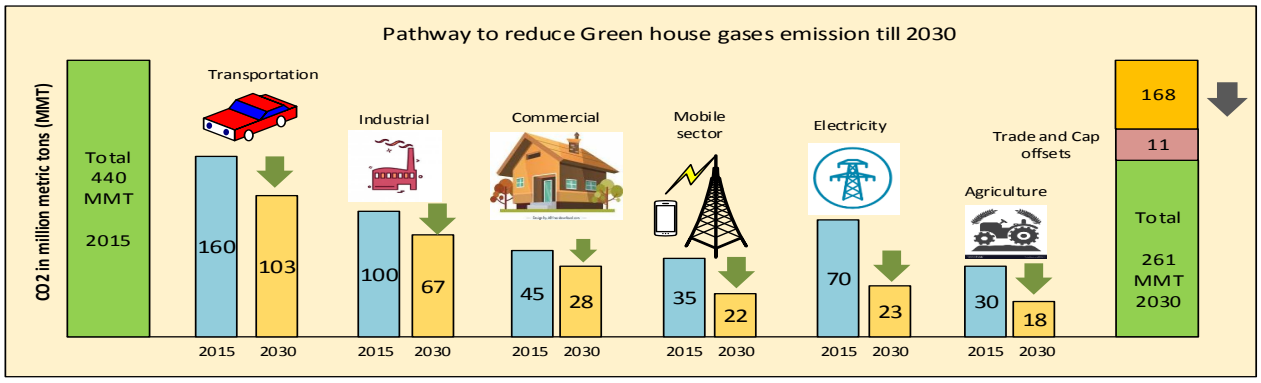


Fig. 6: Aim to reduce GHG across various sectors to reach 2030

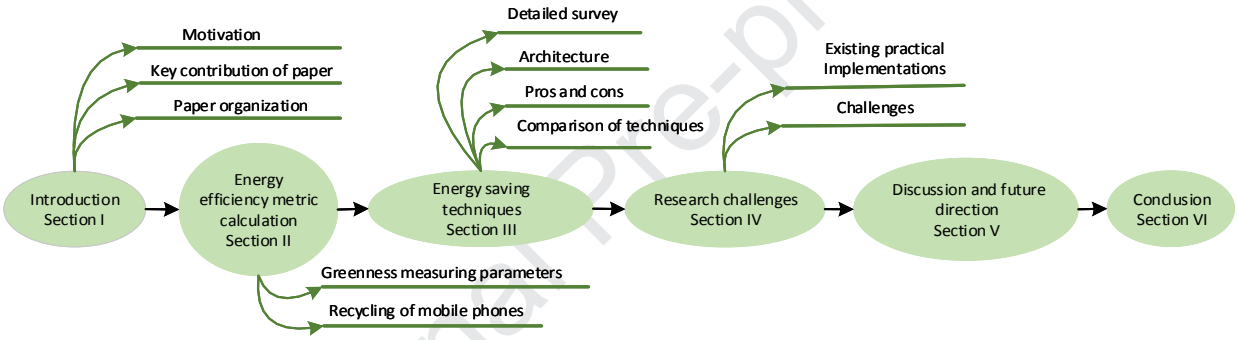


Fig. 7: Paper organization

with energy efficiency metrics for greenness calculation. The main purpose of this section is to familiarize with the vital facts and the expressive concern of an energy aware green CRN research which is the subject of attention. In Section III, various energy aware techniques have been discussed with their concept, detailed survey, merits and demerits and also discussed how they are capable of supporting green networking for next generation (5G) networks. A comparison is performed to highlight the role of each techniques to make communication energy-efficient. Various research challenges associated with green communication have been discussed in Section IV. Future research directions have been explored in Section V. Finally, conclusion is presented in Section VI.

II. Energy Efficiency Metrics for Greenness Calculation

A. Energy Efficiency Measurement (EEM)

As we know, the cellular system is the largest component of ICT sector. So, EEM statistics have encouraged both academia and industry researchers to evolve methods to cut down the energy expenditure of cellular system and make the wireless network system greener. The primary goals of energy-aware green CRNs are:

- Improving EE.
- Improvement in the tradeoff between energy-saving and seamless connectivity.
- Reducing unnecessary GHG emission.
- Proper utilization of resources.

Fig. 8 represents that in a wireless communication network BS consumes the maximum amount of energy as compared to the other component of the network [13]. The number of BSs are continuously increasing to provide better QoE expected by mobile users. Thus, energy consumption will also increase proportionally. Nowadays this is a very crucial subject for research to find out some effective methods to minimize energy consumption by fixed and dynamic components of BS while maintaining the QoS. The internal components like digital signal processors, power amplifier (PA), feeder cables and cooling system of BS are considered under the fixed components while network planning, deployment and management based on proper load balancing are under the dynamic components. For the conversion of a base station into a green base station, several schemes have been suggested in

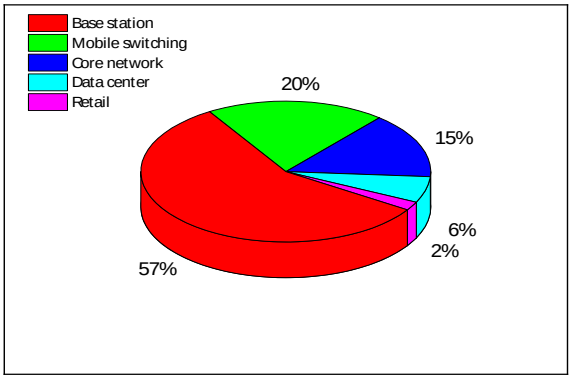


Fig. 8: Energy consumption w.r.t. different components of communication system

[14]. A game theory-based scheme considers heterogeneous networks for the reduction of power consumption at the base station level is discussed in [15]. A novel network architecture having energy-efficient techniques based on the physical layer and cross-layer of cellular networks is represented in [16]. Different research activities and projects based on energy-efficient communication has been considered in [17].

The various components in wireless networks consume different amounts of energy. So, for calculating and comparing the amount of energy which is consumed by the various components and a communication system as a whole, EE metrics are used. Researchers have been proposed different levels (facility, component, and system) for measuring EE metrics [18]. The high-level systems such as data centers are considered under the facility level [19]. Presently, with the upsurge in research activities on green energy efficient wireless networks, some standard organizations like European Technical Standards Institute (ETSI) and Alliance for Telecommunications Industry Solutions (ATIS) are making efforts to describe diverse EE metrics for wireless networks. The EE metrics consist of three major metrics as follows:

1) Energy Efficiency: EE is a very essential metric to measure in green wireless CRN [20]. This can be obtained by dividing network output energy to network input energy consumption [21] and expressed as

$$\eta = \frac{\text{Network output energy}}{\text{Network input energy consumption}} \tag{1}$$

here η represent EE. The best EE can be achieved by minimizing total power consumption and maximizing the data rate.

2) Secrecy Energy Efficiency: This is another important metric consider under the system level. The high value of this metric provides better security in the CRN [22] [23]. This is given by

$$S_{EE} = \frac{Secrecy\ rate}{Power\ consumption\ at\ \text{network}} \tag{2}$$

here secrecy rate (SR) is defined as the difference of data transmission rate of the transmitter-receiver link to the transmitter-eavesdropper link and S_{EE} represents secrecy energy efficiency.

3) Area Energy Efficiency: The A_{EE} metric is considering for a dense deployment [15]. This can be obtained by dividing overall energy efficiency to the macro cell area and given by

$$A_{EE} = \frac{Energy\ efficiency}{Total\ area\ of\ macrocell} \tag{3}$$

here A_{EE} is area energy efficiency.

The readers can discover different taxonomy of green communication metrics which are summarized in Table III.

B. Recycling of Mobile Phones

Usage of recyclable stuff for manufacturing of next generation devices can help to make the environment eco-friendly. For example cell phones are manufactured by some scanty natural resources likes plastic, copper, etc and also contain some non-biodegradable materials. This can surge the GHG effect if it did not correctly dispense when the cell phones are in a dead state. In Australia, almost 24 million cell phones are existing in cupboards and drawers which are not currently used [35]. Presently, the whole world is fighting against vast energy consumption and GHG effect. So, the trends for recycling cell phones are ever-increasing because 90% material of cell phones is recyclable. Fig. 9 represents the cell phone recycling process with a percentage of CO_2 emitted at each stage.

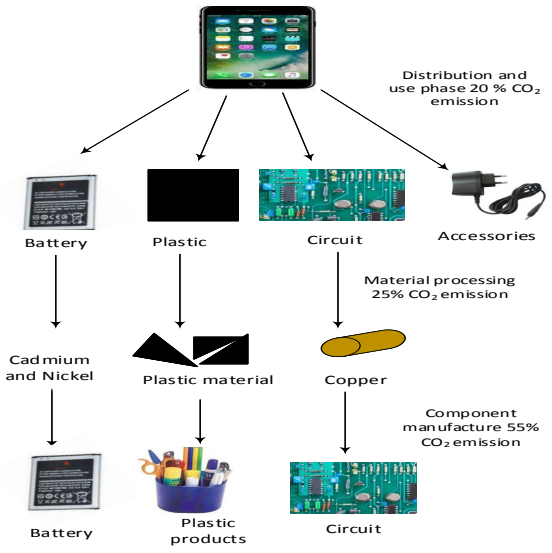


Fig. 9: Cell phone recycling process with CO_2 emission

III. Energy Efficient Techniques

In this section, several traditional and recent advanced energy efficient communication approaches for future CRNs (5G) are considered with their detailed overview in an informational manner. Various energy-efficient communication approaches are highlighted in a classification tree as shown in Fig. 10. These different approaches are introduced by researches to handle various inspiring topics allied with green CRNs.

TABLE III: Summary for several EE metrics evaluation in green communication networks

Level	Metric	Remark	Unit	Reference
Facility level	Power usage efficiency	It gives the ratio of total power consumption at facility level to the total power consumption at equipment level.	PUE ≥1	[24], [25]
Facility level	Data center efficiency	Reciprocal of PUE	DCE in percentage	[24]
Element level	Telecommunications energy efficiency ratio (TEER)	Ratio of maximum throughput to total power consumption.	$Gbps/Watt$	[26]
Element level	Telecommunication equipment energy efficiency rating (TEEER)	Calculated by summing the amount of energy consumed by the equipment for various traffic conditions.	$-\log(\frac{G_{bps}}{Watt})$	[27]
Element level	Energy consumption rating (ECR)	Ratio of energy consumed to the adequate system capacity.	$Watt/Gbps$	[18]
Element level	ECR-weighted (ECRW)	Evaluated in the similar manner as ECR but energy consumption is in ECRW is calculated as $0.25E_i$+$0.4E_h$+$0.35E_f$ here E_i, E_h and E_f represents Energy consumption in idle mode, half load and full load.	$Watt/Gbps$	[18]
Access point level	Energy consumption rate	Energy used to transmit a piece of information.	J/bit	[28] [15]
Access point level	Spectral efficiency	For a communication system, the transmitted information rate to a given frequency	$b/s/Hz$	[29] [30]
Access point level	Power efficiency	Ratio of output power to input power	-	[18]
Access point level	Energy consumption index	Ratio of (BS/AP) input power to the key performance indicator. ECI evaluate the power utilization efficiency of a BS. ECI inversely proportional to EE.	-	[31] [32]
System level	Power per bit per area	The average power used in terms of the average transmission rate and the coverage area.	$W/bps/m^2$	[33] [32]
System level	Energy per bit per unit area	The energy consumed for transferring the bits for a particular coverage area.	$J/bit/m^2$	[25]
System level	The power consumed per unit area	Ratio of power consumed to unit area	W/km^2	[34]
System level	Performance Indicator in urban areas (PI urban)	Ratio of total number of mobile users to the total power consumed by the base station/ access point.	$Users/Watt$	[34] [18]
System level	Performance Indicator in rural areas (PI rural)	Ratio of coverage area to the total power consumption by the base station/ access point.	$km^2/Watt$	[34]
System level	Energy reduction gain	Defined as the derivative of energy consumption gain metric.	ERG in percentage	[26]

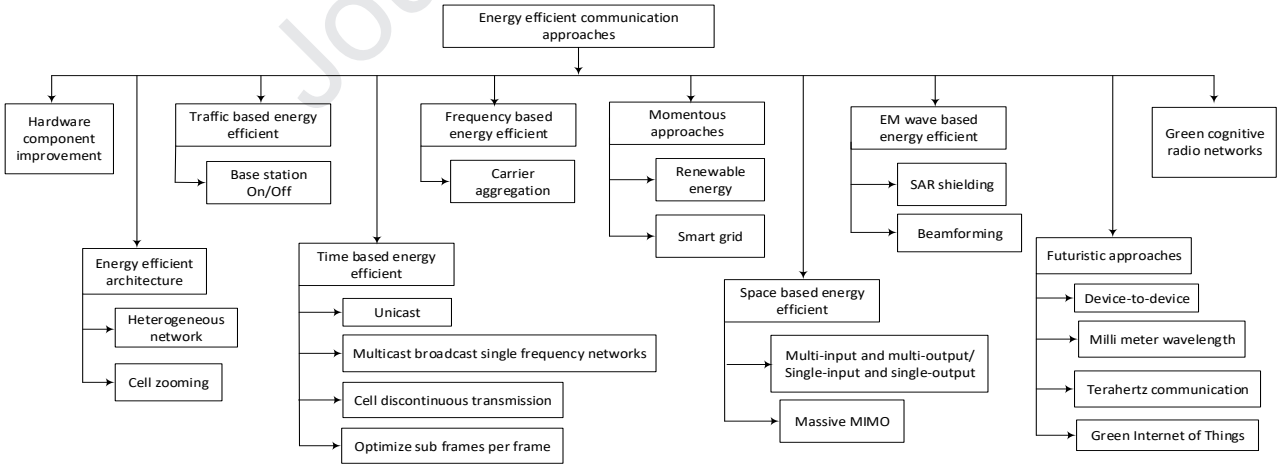


Fig. 10: A classification tree of different energy efficient communication approaches

A. *Hardware Component Improvement*

In the telecommunication system, the PA consumes maximum amount of energy (nearly 60%) in the BS. Therefore, for OPEX cost reduction a highly proficient PA is necessary for green CRNs. In [41], various techniques for improving EE related to PAs and cables used in the feeder and cooling components are discussed. An operating point of the most efficient PA is adjacent to the saturation region. In [42], the author has

TABLE IV: Approaches to enhance the EE of power amplifiers

Approaches	Description	Energy saved	References
Doherty designs	Unique design method which is able to offer significant enhancements in EE.	30 to 35%	[18], [36]
Switched mode power amplifier (SMPA)	SMPA works on the principle that the active device works in saturation, therefore it works as a switch, so either current or voltage waveforms are alternatively reduced to decrease overlapping. This leads to minimizing power dissipation. If a transistor is considered as an ideal switch, 100% of efficiency can be attained by appropriate designing of the network.	80-90%	[19]
Enhancement in digital pre-distorted Doherty architecture and gallium nitride (GaN) amplifiers	GaN is a special material that is characterized by large breakdown voltage, high current density, and high frequency. Excellent power dissipation and thermal conductivity are used in the construction of power amplifier transistors, which enhance the efficiency of the system.	50%	[37]
Inverse Class F	Considered for WiMAX network having carrier frequency approximately 2.45 GHz.	74%	[38]
Envelope tracking designs	To track the input signal magnitude envelop this technique uses amplifier's drain bias.	Up to 60%	[37]
Class-AB with digital pre-distortion	Class AB with digital pre-distortion amplifiers, display gain roll-off at small input power and at saturation this property helps reduce energy.	50%	[39]
Light Radio Cubes	Alcatel-Lucent has been designed as an innovative BSs in 2012 which reduces a large amount of energy consumption as compared to the existing radio access network devices.	50%	[40]
Class J Amplifier	Due to the presence of nonlinear capacitance, the Class-J amplifier represents high efficiency. This nonlinear capacitance generates a harmonic voltage which reduces the phase difference between current components and fundamental voltage.	70 to 90%	[19]

discussed that PA operates in the linear region due to non-constant envelope signals of orthogonal frequency-division multiplexing (OFDM) modulation and nonlinear effects ranged from 6 to 12 dB below saturation region. For designing of PA, the following factors are having major importance:

- Wider bandwidth for communication.
- Follow linearity property, to provide higher-order modulation technique.
- Improved BS energy efficiency to achieve OPEX.
- Higher average output power levels.

The higher value of DC power supply is required for the operation of PAs. During a day, the traffic load is not constant like in-office hours traffic load is high and in the night it is low. So, a large amount of power is wasted at night because the working of PAs does not depend on traffic load. To overcome this challenge two techniques can be used. In first, the PAs can be deactivated in some time slots with no data transmission. In second, an operating point of PA can adjusted to reduce devouring of power and also optimize the power efficiency for different traffic load scenarios (low, medium and high). The major advantage of this technique is that it saves the maximum amount of energy but high replacement cost is its limitation. Table IV summarizes a survey of various approaches that have been used to improve the hardware component of the communication system and also represents how much amount of energy can be saved by these approaches.

Instead of several endeavors, an improvement in hardware approach for saving energy at BS is not in a position to achieve the desired result. Ineffective utilization of resources also results in a large amount of energy wasted. So, for designing energy-efficient green CRNs, exploitation of both network-level and equipment level approaches are required. In the succeeding subsection of the paper different approaches are discussed to improve EE.

B. Energy-Efficient Architecture

Energy-efficient architecture considers an efficient approach for both energy efficiency improvement and energy saving. This approach is based on the cooperation among BSs to offer the best QoS to mobile subscribers and save energy at low traffic. In this subsection, two techniques such as heterogeneous networks (HetNets) and cell zooming are discussed.

1) Heterogeneous Networks (HetNets): HetNets represents a layered architecture that has different networks for providing the service to the same mobile subscriber. Considering the small size cell (like femto, pico, and microcells) to reduce consumption of energy and enhance the throughput of the network. These cells differ in data rate, size, capacity, output power, coverage and also emit a different amount of CO_2 as indicated in Fig. 11 (the size of the cloud). As the diagram represents that macrocells are having the largest coverage area and least CO_2 emission while femtocell is having less area but emits the maximum amount CO_2. The principle of the HetNet approach is the deployment of macrocells for providing complete network coverage, while small cells will come in an operational state when the traffic demand increases in that area. In [48], [49] author considers both scenarios heterogeneous and homogeneous (microcells and macrocells) networks having a different number of microsites, based on data traffic situations for calculating EE. The results indicated that better EE can achieve in HetNets deployment. Fig. 12 represent architecture in which different networks are deployed to form HetNets. The deployment of more microsites in network increases the throughput of the network. In fig., various links like (macro-macro, macro-micro, macro-pico, macro-femto etc.) are present.

Table V summarizes the various schemes and algorithms used in the available literature to improve EE in HetNets. The table also showing the description and percentage of

TABLE V: Algorithms/Schemes for improving energy efficiency using Hetnet

References	Objective	Description	Energy saved	Scheme/Algorithm
[43]	Multi-layer nature of HetNets is analyzed which shows energy consumption depends on the traffic requirement and BS hardware components.	Consider the classical 0-1 Knapsack issue. The scheme is analyzed to reduce consumption of energy by adapting the BS on/off situation.	35%	On/off scheme
[44]	Micro BSs have been Deployed in the area where macro BSs were situated and calculated saved power at the various radius of micro BSs.	Analyze the placement of various BSs based on a new power-saving metric.	Maximum energy is saved in Greedy BSs deployment algorithm as compared to others	Region-based, Grid-based and Greedy BSs deployment algorithm
[45]	Deployment of the heterogeneous network which is having different size cells (micro, pico, and femto) for achieving green efficient system.	An energy-saving method is proposed that permits femtocell BS to completely switch off its communication and processing when not involved in an active call.	37.5%	Voice traffic model
[46]	To calculate the EE gains	Implementation of femtocells within a microcell, while maintaining QoS. The results show that the usage of femto cells minimize the power consumption of the integrated network.	70%	Fast fading and path loss model
[37]	To analyze the impact of energy efficiency.	Deployment of different picocell in a macrocell.	60%	-
[47]	Investigate the area power consumption in terms of network performance metric	Discuss the effects of many deployment policies on the power consumption of mobile system.	55%	Power consumption models

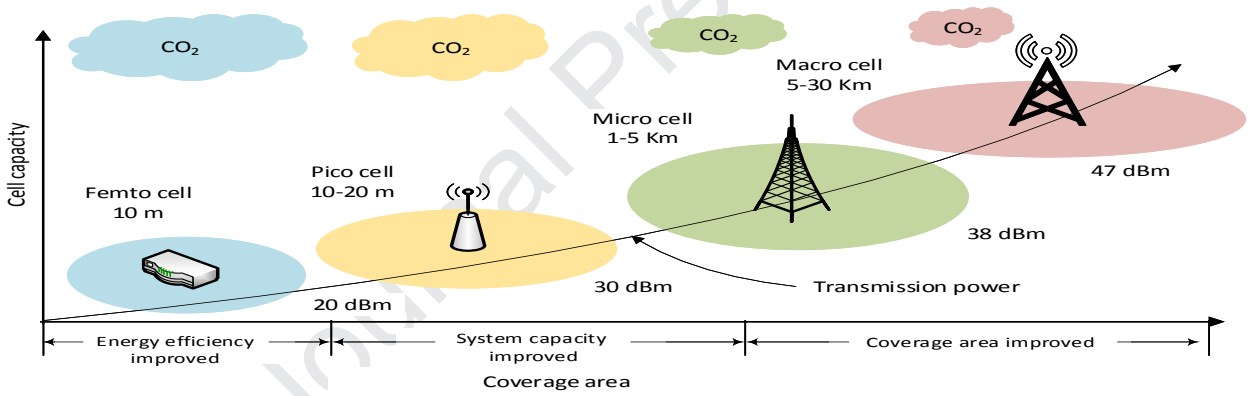


Fig. 11: Power consumption by different cell size for HetNet

energy saved by these schemes. A combination of different networks and smaller size cells provides an improvement in EE, coverage, and capacity because of the short transmitter-receiver distances. There are also some challenges in this technique, including the organization of the HetNets interfaces and the dead zone problem.

2) Cell Zooming: Cell zooming technique not only saves energy but also improve network performance. In this technique, the size of the cell varies with variation in traffic load, state of channel and requirement of mobile subscribers. In the condition of high traffic load, the congested cell will be in the zoom-in mode and low traffic cells will be in zoom out mode to provide the service to the mobile subscribers of the congested cell. In cell zooming algorithm the time is represented by $'T'$ (called cell zooming period). There are three stages in each period such as coordination, transition and serving phase as shown in Fig. 13. A cell zooming server (also known as the brain of the system) will sense the traffic situation and then decides the occurrence of cell zooming operation. A specific control message is released after the sensing process. After coordination with neighbor cells, the required cell will zoom in/zoom out by cell zooming server under the serving phase. The working of the cell zooming approach is represented in Fig. 14 having a seven cells cluster cellular network. Six nearby cells are covering one central cell as shown in Fig. 14a. The position of BSs is at the center of each cell. In a cell, mobile nodes are located arbitrarily. The central cell will perform a zoom-in operation when the cell is congested by the movement of mobile nodes into this cell. Thus, the size of the cell is minimized as shown in Fig. 14b. On the other hand, The central cell will perform the zoom-out operation, in the case when the movement of the mobile node will out from the central cell. As from Fig. 14c. congestion can be avoided by the zoom-in operation of neighboring cells.

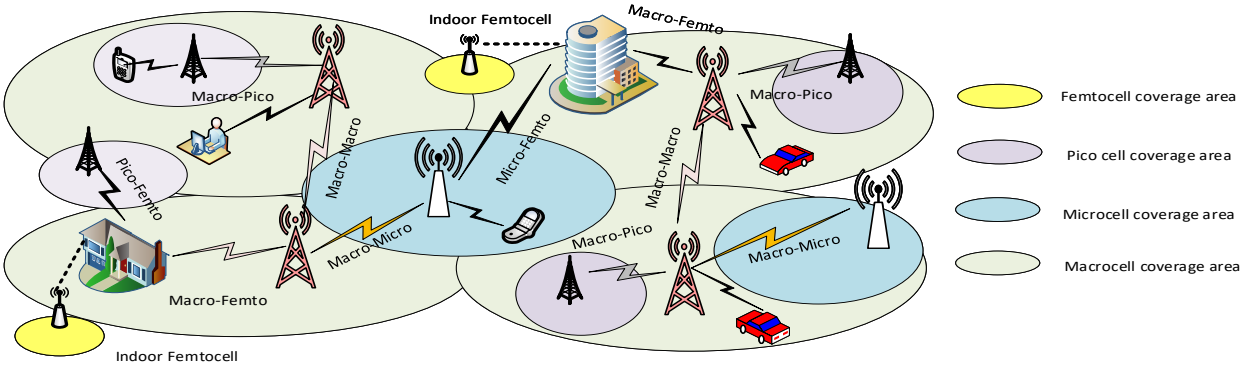


Fig. 12: Architecture for energy efficient HetNets

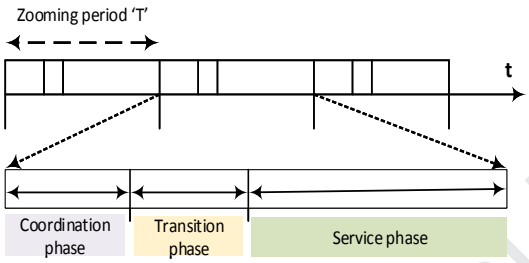


Fig. 13: The framework of cell zooming process

The mobile users will get service by nearby cells when the central cell is less congested and goes in sleep mode as shown in Fig. 14d. Table VI represents the literature based on many algorithms which are used for cell zooming technique. Almost 50% of energy can be saved by cell zooming technique.

There are various benefits of the cell zooming technique. It provides proper load balancing and reduces the blocking probability. This technique also enhances the battery life of the handset and throughput but intercell interference is the main disadvantage of this technique when some active nearby cells perform zooms out operation simultaneously. In cell zooming operation, coverage holes (no network coverage at some places) may occur. The newly arrived mobile user can be served by performing zoom in operation by nearby cells to cover these unreachable areas.

C. Traffic based Energy Efficient Approaches

The traffic based energy-efficient approache is discussed in the following subsections.

1) Base Station ON/OFF Technique: In wireless networks, energy can be saved while maintaining QoS by turning off idle wireless resources and BSs. The traffic load condition varies with space and time in the coverage region. So, underutilized BSs can be dynamically turned off. The user can be served by remaining BSs which are in an active mode of operation [54].

To ensure global coverage, the telecom operators are expanding to almost remote areas. Fig. 15a shows the data of off-grid and poor-grid BS from 2014 to 2020. Due to the absence of electrical grids, the off-grid BSs are expensive. In 2009, a telecom company named China Mobile has minimized the expenditure of energy (nearly 36 million kWh/year) by applying BS sleep control. In [55], an optimization approach related to reducing the consumption of energy is suggested with QoS constraints. The model is a combination of signal propagation models and a set of actual sized universal mobile telecommunications system (UMTS). The outcome consists of very useful constraints like guaranteed service rates, full coverage, minimum delay and accomplishes a significant reduction in network energy almost 35 and 57%. The amount of consumed energy depends on the BS type and application scenario. From Fig. 15b, 15c, and 15d, it is cleared that how much energy percentage is consumed by the separate parts of different BSs.

In a few cases, the power utilized by the BS is considered as zero in sleep mode but in actual it consumes very less power as compared to the active state of BS. However, the BS never goes in completely off state, it may still utilize some power like detection power [56]. This type of power consumption can not be ignored, especially in BSs that do not have on-site cooling systems and installed in a small location. Sleep mode is further divided into many sleep levels. In [57], an opportunistic sleep concept is discussed with an idea of BS sleep only for a certain period to improve the system capacity. The range of energy saving depends on various traffic load conditions and lies from 13% to 41%. In [58], an author describes four status of the BS and their respective power consumption ratio as represented in Table VII.

The BS with sleep/wake-up strategy is discussed in [59]. According to the traffic density, BS switches among active, sleep and off mode. Presently, a new module named as sleep/wake-up is augmented in each BS. Nearly 72.9% of energy can be saved by using this module. In [60], the distance between BS and the mobile node are considered as criteria

Journal Pre-proof

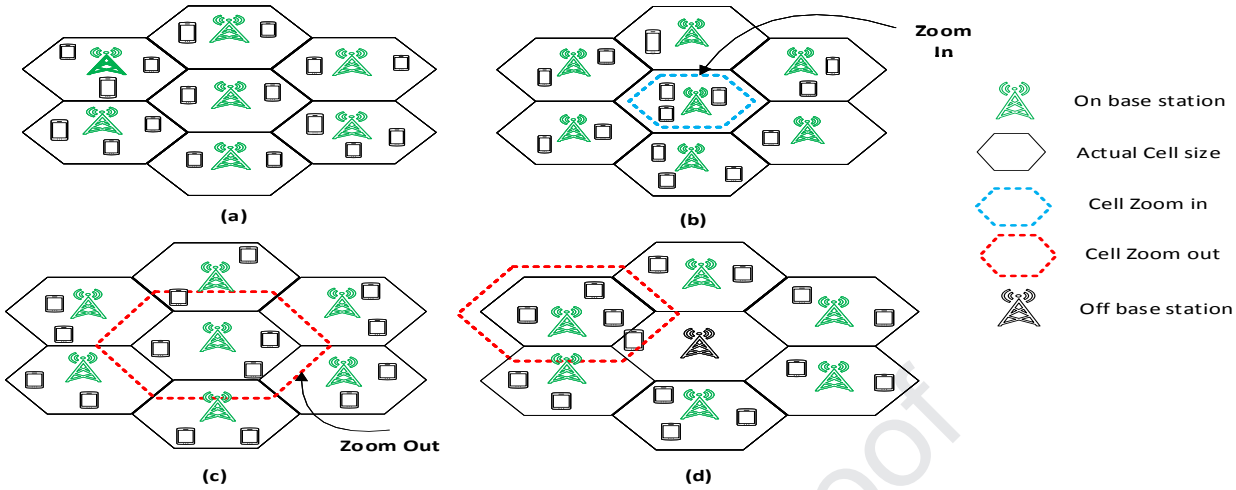


Fig. 14: Concept of cell zooming (a) Typical layout of cell (b) Cell zoom in because of high congestion in network traffic (c) Cell zoom out to serve mobile users (d) Data transfer in sleep mode via another nearby cell.

TABLE VI: A literature survey on Cell Zooming Technique

Objective	Description	Energy saved	Scheme/Algorithm	References
To achieve user traffic request expands network EE and optimize user service by adjusting the coverage of cell	For cell coverage author describes Cell Zooming Factor (CZF), which gives the relationship between small cell and large cell densities. To optimize the CZF and minimize area power consumption game theory is applied.	41%	Adaptive cell zooming scheme, cell sleeping scheme	[50]
To reduce the consumption of energy at BS and maintain proper load balancing of traffic	Analyze the adaptive network traffic load balancing technique to overcome the energy consumption issue.	55.75%	Adaptive traffic load balancing	[51]
Divide the area into more small cells to minimize energy consumption	The author expressed that if the number of cells increases then more cells can be switched off, therefore blocking probability will reduce and EE can improve.	40%	cell switch off scheme	[52]
To reduce the extreme traffic of cellular network	The author considers Relaying, Physical Adjustment, BS Cooperation and BS Sleeping for implementing cell zooming concept.	55%	Adaptive cell zooming scheme	[53]

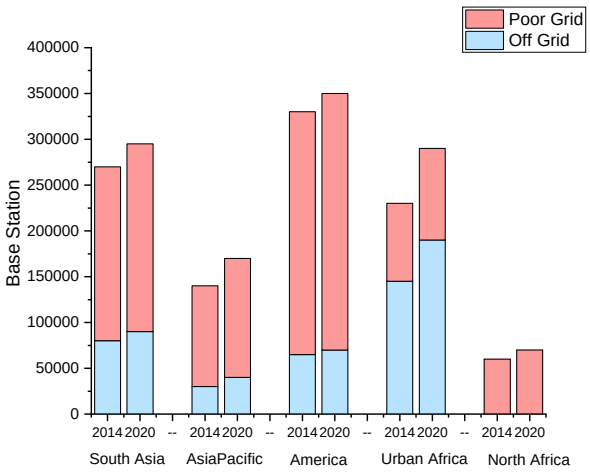
TABLE VII: Power consumption ratios status in several mode

S.no.	Base station status	Power consumption ratios
1	On	100
2	Standby	50
3	Sleep	15
4	Off	0

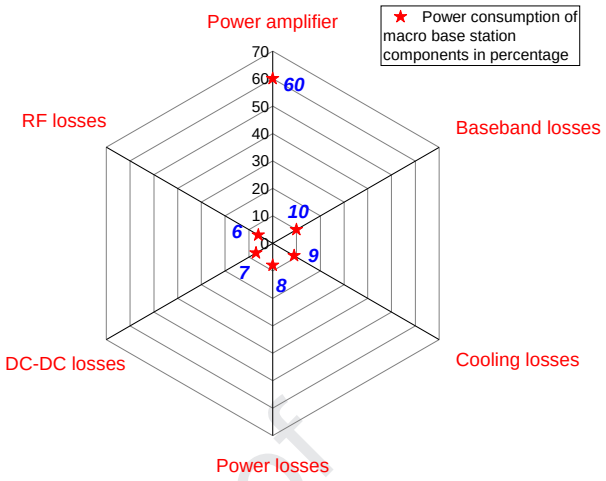
for decision making about making BS in on/off state. This technique reduces the energy consumption of nearly 29%. The switching period of the BS (switching from on to off and off to on state) is associated with QoS and plays an important role [61]. In [62] the author investigates that the processing time, required for handoff is ranged from 50 ms and 100 ms. The cooperation among BSs is necessary to serve the mobile users belonging from switched-off cites. A flowchart is proposed, related to BS on/off operation to attain an energy-efficient solution as represented in Fig. 16. The BS on/off technique achieves a remarkable trade-off between QoS and energy saving. To make the energ-efficient system, the uncovered region related to inactive BSs is covered by increase the transmission power of active BSs. Since the BS power is limited so an issue related to network coverage occurs. The battery life of the mobile users is also affected by this approach because of the requirement of high power at the receiver end to maintain connectivity with nearby BS.

D. Time based Energy Efficient Approach

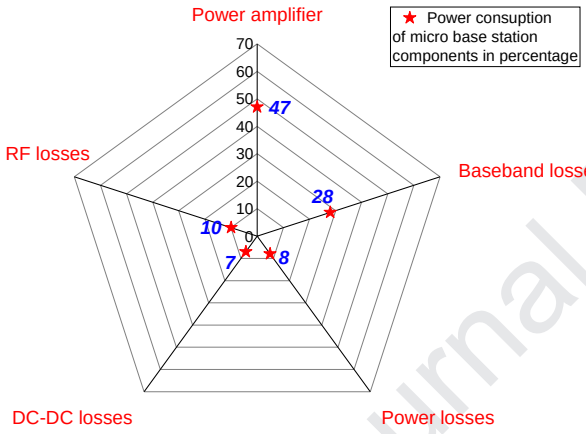
In the time domain approach, the PA of LTE base station is temporally turned off during low network traffic in the downlink and idle channel state. This power-saving solution is tightly associated with the LTE system frame structure. The concept of the frame structure, reference signals (RS) and control signals have investigated in [63]. LTE downlink uses orthogonal frequency division multiple access (OFDMA)



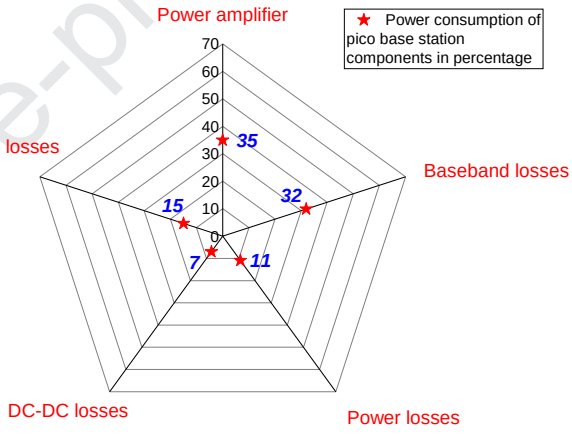
(a) Increment of base station from year 2014-2020



(b) Energy consumed by the different part of Macro base station



(c) Energy absorbed by various parts of Micro base station



(d) Energy absorbed by various part of Pico base station

Fig. 15: Energy consumed by different parts of different BS

approach. In this, an RS signal performs two main tasks in the network such as channel estimation for the mobile node and search neighbor cell for handover [64]. There are mainly three techniques to minimize the number of RSs.

1- Normal unicast:- The LTE frame consists of 7 symbols per slot. The RSs, applies in 0, 4, 7 and $11th$ OFDM symbols at every subframe. This pattern of RSs keeps on slot $1-4$ and $6-9$ of subframe and does not apply on slot 0 and 5 of the subframe. Slots 0 and 5 are used to transmit control signals [65].

2- Multicast Broadcast Single-frequency Networks (MBSFN):- The services like mobile TV and DTH use the MBSFN technique in the LTE system. In this, the pattern of RSs keeps only in slot 1. So, the number of RSs is reduced.

3- Cell Discontinuous Transmission (DTX)/Microsleep:- In DTX approach, the RSs signals are not transmitted in slot $1-4$ and $6-9$ of subframe if no mobile node is found in the active state [63].

Fig. 17 represents that each frame of the LTE system includes 10 subframes (from 0 to 9) and there are 7 OFDM symbols in one slot. In a single slot, four RSs are used for communication purposes. Some amount of energy is required for the transmission of RSs. In cell discontinuous transmission technique no RS is sent for idle and low traffic load condition so the maximum amount of energy can be saved in this technique as compared to the normal unicast and MBSFN technique in which RSs are sent at low traffic conditions.

In [66], a dynamic subframe allocation technique is introduced for multimedia broadcast multicast services with MBSFN. In [67], an upcoming network traffic load is predicted by MBSFN. The BS turning off periods are calculated with the help of MBSFN based energy-saving technique [68]. In this technique, the needful resources are pre-calculated based on traffic load. Thus, more energy is saved by keeping off underutilized resources. This technique also has features

Journal Pre-proof

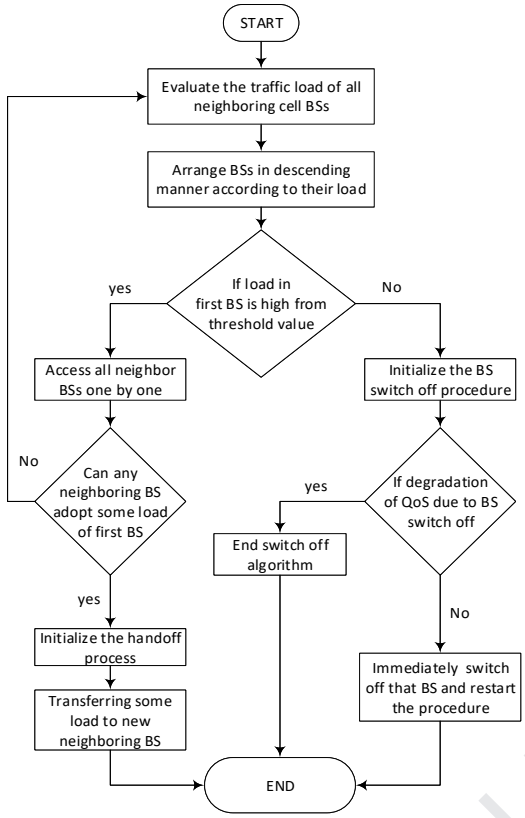


Fig. 16: Flow diagram of proposed BS sleeping operation

like backward compatibility, delay and process time [69]. In addition to this, Fig. 18 shows the trade-off between on period of PA in a frame to the energy-savings for various time-domain approach. The maximum amount of energy can be saved by the extended cell DTX technique. This technique gives the best results for rural areas where traffic is low and idle case scenario. This technique also saves energy at components levels like cooling systems, DSP units and ADC/DAC devices in turn-off time duration. Some mobile node faces troubles due to the large number of RSs during synchronization operation with BS. Thus, the mobile node may fail to admit in DTX mode. This reduces the battery life of the mobile node and causes a negative effect. There is also difficulty arises during control signal decoding.

E. Frequency based Energy Efficient Approaches

Energy-efficient approache based on frequency is discussed in the following subsection.

1) Carrier Aggregation: The LTE-A (Long Term Evolution Advanced) system uses the bandwidth of up to 100MHz in different frequency bands [70]. Such frequency bands are 450–470 MHz, 698 – 960 MHz, 1710 – 2025 MHz, 2110 – 2200 MHz, 2300 – 2400 MHz, 2500 – 2690 MHz and 3400 – 3600 MHz define by the International Telecommunication Union (ITU) [71]. The carrier aggregation (CA) technique used by the LTE-A, exploits and supports various frequency bands. The concept of the CA technique is to amalgamate two or more component carriers (carrier frequency) into one frequency to amplify the channel capacity and data rate. The frequencies belonging to the same or different frequency bands can be combined in this technique [72]. The concept is represented in Fig. 19. In [73] author considers multi-cell and single cell scenario to analyze the impact of CA on overall power, send by eNodeB. A relation among assigned bandwidth, cell radius, and the total send power by the eNodeB is expressed as:

$$eNB_{TP} = \frac{2R^{\alpha}(2^{\frac{D\times\sigma}{BW}-1})\mu \times BW \times N_0}{(\alpha+2)kd_0(-\ln[1-P_{r(out)}]^{\frac{1}{M}})} \quad (4)$$

Where eNB_{TP} represents the total send power of eNodeB and R is the cell radius. α gives the value of channel path loss exponent, D is the minimum data rate for better QoS. The total number of active users is represented by σ. The term BW, μ, T and K represent transmission bandwidth, modulation coefficient, noise temperature and Boltzman constant respectively. Here $K = 1.38X10^{-23}m^2kgs^{-2}K^{-1}$. The transmission power in both scenarios decreases by nearly 65% by applying the CA technique and EE of the network improves. According to [74], the carriers are combined into groups in CA technique and separate PAs serve every group individually. The limitation of this approach is that the BS with total carriers must have individual PA connected to every group of carriers.

CA supports aggregation of licensed carrier and unlicensed frequency spectrum. Hence unlicensed bands are also utilized. Scaling of the carrier also facilitates the expanded coverage. This technique also helps in enhancing network efficiency and optimizes user experience. Multiplexers are required at transmitter end to combine the carrier components and RF filters are required at the receiver end to isolate carrier components. PAs and switches are specially designed to fulfill the need of CA-based RF system. These additional components raise the cost of the apparatus and device manufacturers and also increases system complexity.

F. Momentous Approach

1) Renewable Energy Resources: Usage of renewable energy resources (RER) in the mobile communication system for making the system greener is a very remarkable approach. With growing wireless communication, infrastructure needs a high amount of electricity. These RER is used for powering cellular system and having many advantages. The major advantage and popular justification behind using RER are to make the environment eco friendly and reduce the emission of GHG. Currently, India has more than 320000 mobile phone towers, which consume almost 3 billion liters of diesel per year. This switching from diesel to solar and other alternative energy sources will result in deduction of nearly 6 million tons of CO_2 emissions and save nearly 1.5 billion in operating expenses for wireless communication tower companies. Switch to RER can produce millions of carbon credits that could balance the OPEX on telecommunication towers. Areas like islands and deserts where the installation of a smart power grid is difficult, the RER can be considered as a problem

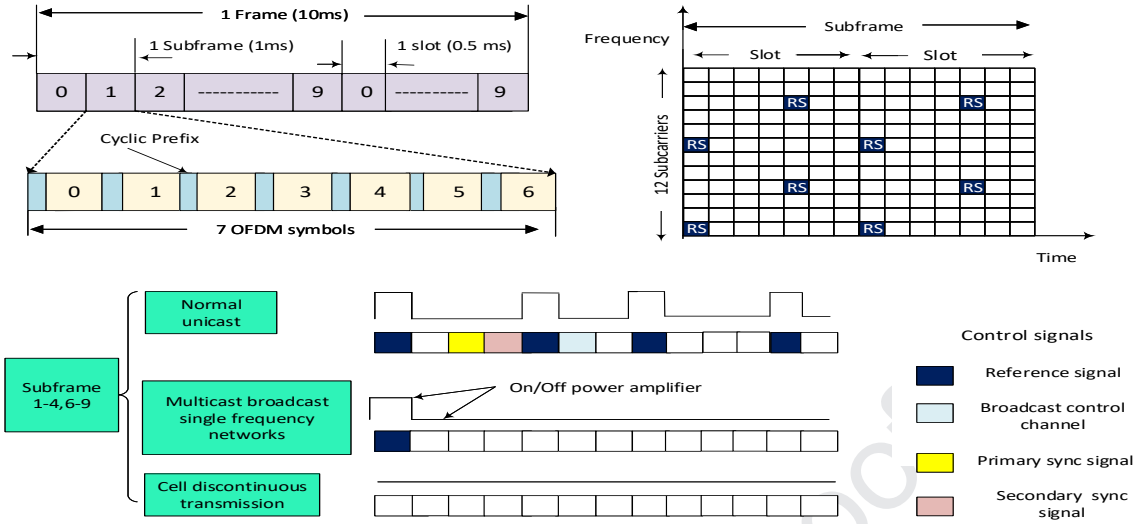


Fig. 17: Reduction of RSs in LTE frame for idle and low traffic load condition

TABLE VIII: The value of power consumption for various cellular BSs

Base Station Type	NTRX (number of transceiver chains)	PA (W) power consumed at power amplifier system	BB (W) power consumed at baseband system	Cooling (%)	Mains Power(%)	DC-DC (%)	RF (W)	Total per TRx (W)	Total Power for NTRX (W)
Macro	6	128.2	29.6	10	9.0	7.6	12.8	225.0	1350
Micro	2	6.2	2.6	0.0	9.0	39.0	54	72.3	745.8
Pico	2	0.13	4.0	0.0	11.0	4.4	6.8	7.2	14.8
Femto	2	1.1	2.5	0.0	11.0	9.0	0.6	5.3	10.5

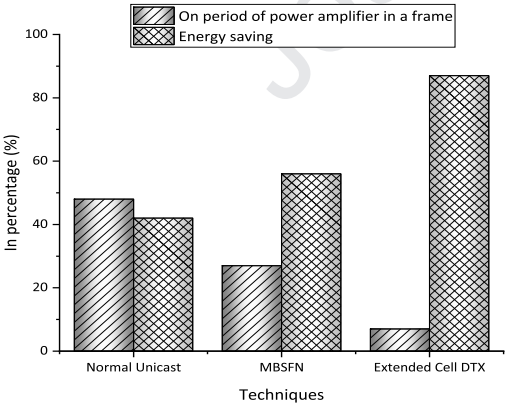


Fig. 18: Tradeoff between active time of power amplifier and energy saving

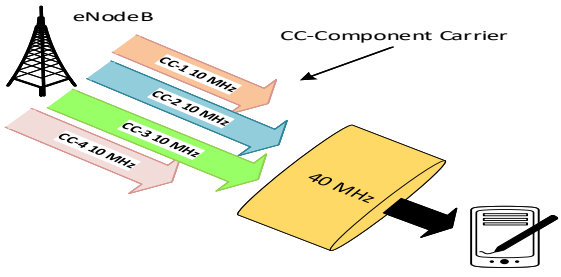


Fig. 19: Concept of carrier aggregation

solving solution [75]. Until now, in such types of areas, mobile operators depend on diesel generators for working of BS which is one of the most pollutant origins of energy and also very expensive because of its price and transportation [76]. The several RER likes sunlight, wind, biomass energy are not only sustainable but also clean in nature and can be used in future networks. The GSM Association has been initiated a program named as "Green Power for Mobile" to support industries related to telecommunications and also to establish several RER for proper operation of 640000 new and prevailing off-grid BSs [77]. With this initiative, nearly 0.347% of the global diesel consumption saved.

In [78] the author describes how RER can reduce the practical realization and working costs of upcoming future

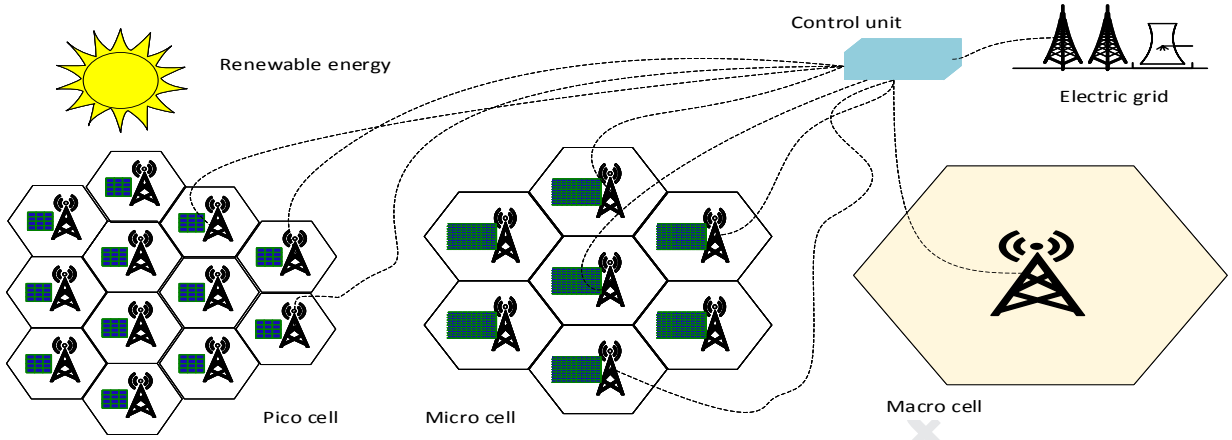


Fig. 20: Addition of renewable energy resource with cellular system

wireless communication systems. Sunlight is an excellent RER used for electricity generation. Another RER used for powering the BSs are wind power, fuel cell, hydro-power and hybrid (solar and wind) system, etc. Fig. 20 depicts the utilization of RER with the wireless communication system. Here solar panels are connected with BSs of the macrocell and microcell to convert solar energy into electrical energy [76]. The major components responsible for power consumption in BS are PA, baseband unit, RF module, cooling system, DC power supplies and losses incurred in feeder [79]. The power consumption in BS can be determined as

$$P_{consumed\ at\ BS} = N_{TRX}\frac{P_{BB}+P_{PA}+P_{RF}}{(1-\xi_{DC})(1-\xi_{MS})(1-\xi_{cool})} \quad (5)$$

here N_{TRX} is the number of transceiver chains. The P_{BB}, P_{PA} and P_{RF} denote power consumed by baseband system, PA and RF system, respectively. The ξ_{cool}, ξ_{MS} and ξ_{DC} represent losses associated with the cooling system, main power supply and DC power supply, respectively. The P_{PA} can be given by the ratio of antenna feeder power output to the addition of PA efficiency (η_{PA}) and feeder loss (ξ_{feeder}) and expressed as

$$P_{PA} = \frac{P_{output}}{\eta_{PA}(1-\xi_{feeder})} \quad (6)$$

Table VIII represents the consumption of power (as a percentage of the overall consumption of power) associated with various parts of BS. The consumption of BS power is divided into two types named as static and dynamic. In first, the power consumption remains constant. In the second, power consumption depends on the number of transceivers deployed and traffic-load. In BS the PA consumes maximum power about 50 to 60% of the total power, and linearly depends on traffic-load. This can be expressed as

$$P_{Consumed} = N_{TRX}(P_{static} + \tau_p P_{Dynamic}) \quad (7)$$

where $0 < P_{Dynamic} \leq P_{\max}$

Here τ_p is the slope of traffic load dependent power consumption. The term $P_{Dynamic}$ and P_{static} represents static and dynamic power respectively.

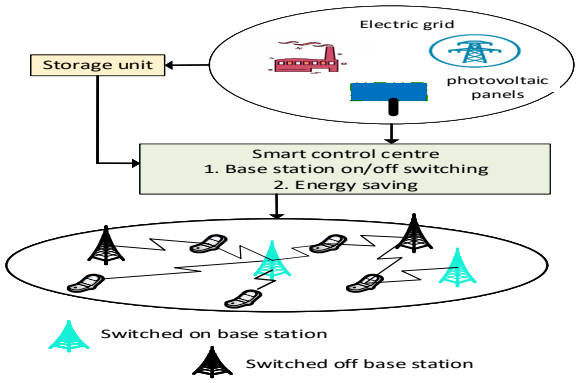


Fig. 21: Future wireless network powered by smart grid

2) Smart Grid: This is an integrated communication system which has various major subsystems such as wired/wireless communication channels, the smart control center (also work as a transmitter/receiver) and software-based management systems. Fig. 21 represents the architecture of future green wireless communication networks concurrently powered with the help of a smart grid and locally deployed RER. The key component of the architecture is the smart control center which sets up the connection between operator and smart grid to optimize wireless network operations like resources management, taking decisions about turning on/off the BSs periodically and storage of auto-generated energy amount etc. For a power supply, the BSs cluster is linked with the electric grid and also connected to the locally generated RE, which generates energy by the array of photovoltaic panels. Due to the irregular production of the RE, the system is connected

TABLE IX: Comparison among various techniques in terms of energy saving

	SISO	SIMO	MISO	MIMO
Merits	Simple circuitry	Known as receiver diversity	Transmitter diversity	Improvements channel throughput
	At short distance SISO system is more energy efficient than others	Under low traffic SIMO system is more energy efficient than others	Less battery consumption	Improve EE for long distance communication
	No additional processing required	Overcome the effects of ionospheric fading and interference	Multiple antennas are combined to decrease errors and increase data speed	Reduce co-channel interference
	Diversity not required	Easy to implement	Reduce the problems of intermittent reception	Improve channel capacity and channel robustness
	$C_{SISO} = B\log_2(1+\frac{S}{N})$	$C_{SIMO} = A_r B\log_2(1+\frac{S}{N})$	$C_{MISO} = A_t B\log_2(1+\frac{S}{N})$	$C_{MIMO} = A_t A_r B\log_2(1+\frac{S}{N})$
	here C, B and S/N are respectively capacity, bandwidth and signal to noise ratio	here C, B, S/N and A_r are respectively capacity, bandwidth, signal to noise ratio and no of receiving antenna	here C, B, S/N and A_t are respectively capacity, bandwidth, signal to noise ratio and no of transmitting antenna	here C, B, S/N, A_t and A_r are capacity, bandwidth, signal to noise ratio, no. of transmitting and receiving antennas respectively
Demerits	Limited performance	Additional processing required at receiver	Additional processing required at transmitter	Requires high processing
	Channel bandwidth is limited	Problem of battery drain	Channel capacity is not Improve	Complex circuitry
	Low throughput	High cost than SISO	High cost than SISO	High cost

with a storage unit, which stores the extra amount of produced RE that can use when RE is not generated, so that the uneven production of RE can be compensated. There is an additional connection between the smart grid and the storage unit shown in the figure. Hence, this storage unit may be used to harvest energy produced from the grid. The arrows represent the possible direction of energy flow.

There are several benefits of using a smart grid in next-generation wireless networks. For mobile operators, the smart grid can be work as a tool for the optimization of energy acquirement, bound their CO_2 emission and reduce their energy bills. The main functions performed by the smart grid are i) To integrate control and information technologies for increasing the generation. ii) Efficient allocation and utilization of energy. iii) To serve consumer's mobile with safe and secure functionality of the smart grid. iv) To design demand-based management that provides controlling choices and secure information to consumer's mobile. In [80], the cost of power consumption with maintaining consumers is minimized by the integration of the smart grid and RER. In [81] a novel combined architecture of RER and electrical grid for wireless networks is designed to utilize the advantages of network operators and smart grid.

In [82], a real-time energy management system based on the smart grid is suggested with the beamforming technique to limit the cost of energy consumption by using locally generated RER. In cooperation with a smart grid, the network state depends on several factors like mobile subscriber density, day of the week (weekend, holiday, working), day time (noon, office, busy hour, midnight) and user's behavior profile, etc.

Smart grid often encourages cost-effective energy saving and high penetration of local renewable energy generation to support future green communication network and also, minimizes the overall emission of total harmful CO_2 with improved QoS. The smart grid system is complex and has number of problems that required to be handled in order to set up a consistent and efficient communication system.

G. Space based Energy Efficient Approach

The various space based energy efficient approaches are discussed as:

1) Multiple-input and Multiple-output (MIMO): This is the most promising approach considers for designing of future networks. Without increasing the transmission bandwidth and power, this technique has overcome the problem of fading and improves the data rate, capacity, and throughput as compared to other antenna systems.The fading effect at the receiver end can be minimized using diversity coding concepts in the MIMO system. Available literature indicates that EE of MIMO system is highest than the other antenna transmission like SISO (Single-input/Single-output), MISO (Multiple-input/Single-output) and SIMO (Single-input/Multiple-output) due to the existence of a larger number of antennas. Table IX shows a comparison among these techniques in energy-efficient communication perceptive.

In [83], three important parameters such as capacity, path loss and delay are considered to make energy-efficient green BS. A trade-off between diversity gain (transmitting similar signals through multiple paths) and multiplexing gain (sending independent signals) is discussed in [84]. In [85] author stated that energy can save at BS by employing steerable directional antennas work on frequency 28 GHz and 38 GHz. In [86], a resource allocation scheme is introduced to decrease the amount of data transmission energy. According to [87], [88]

energy saving MIMO technology depends on traffic demand in that area, so by proper traffic load adaptation, the performance of the MIMO system can improve. In [89], the number of reference signals in MIMO system is decreased in frequency and time domain to make transmission energy-efficient. In [90], a switch-off RF and PA units based novel approach is introduced to reduce energy consumption. It means, at low traffic system, switching from M-antennas to 1 antenna reduces the consumption of energy by the factor 1 by M.

2) Massive MIMO (M-MIMO): M-MIMO is one of the advanced technology of MIMO having several antennas at BS. This technique requires shorter wavelengths (higher frequencies) because the system needs to physically pack more antennas into a small area than the other mobile network [91]. This approach is not only energy efficient but also spectrum efficient. The main advantage is that a BS can serve multiple subscribers simultaneously within the same spectrum. This results in better multiplexing and array gain at the same time. Now the question is for green massive MIMO system, how many antennas are required? In [92] author suggested that in the M-MIMO system, a BS can have X antennas and subscribers have Y single-antenna. A subscriber can scale down uplink power to X and $\sqrt{X}$. In [93], a power consumption model is investigated for obtaining an energy-efficient optimal solution by installing $150 - 250$ antennas to serve a large number of mobile subscribers. Fig. 22 depicts the concept of M-MIMO where an array of antennas is situated in BS and serves multi users simultaneously. Table X summarized a literature survey on M-MIMO techniques and table XI provides a comparison between MIMO and massive MIMO considering energy saving prospect.

H. Electromagnetic Radiation based Energy Efficient Approach

Now a day's mobile communication networks have become an important part of human life. Enormous efforts are dedicated to developing mobile communication services without thinking about the adverse effects of its on human. It is always a topic of discussions among people about the installation of BSs near the residential areas. The Electromagnetic Radiation (EMR) from these BSs causes long-term effects on animals, birds, and humans. The literature stated that BS's EMR causes low risk if antennas are not directly in contact. But the long-term effects of this radiation are still concerning. Paradoxically, BS is considered as the main source of EMR exposure, even though there is a greater health risk from the emissions by mobile phones because those antennas are directly in contact with the human body usually head and ear when it operates. This EMR can interpose the working of our body organs particularly that of our brain and also the reason for some serious health problems such as depression, insomnia and a shortened attention span [102]. A report of the World Health Organization (WHO) shows the classification of RF emissions as a carcinogen of the group 2B category, which is a category shared by lead and pesticides. This study suggests that excess usage of cellphone leads to the development of malignant tumors, especially of the brain. Some research also indicates that exposure to non-ionizing radiation can lead to skin problems, fatigue, headaches, allergies, Parkinson's disease, Alzheimer's disease and more [103]. To overcome this problem, Specific Absorption Rate (SAR) are explored to minimize EMR. Thus, these techniques provide the direction for the designing of next generation low-EM exposure to wireless networks.

EMR Metrics– For finding the direction of EMR of any antenna, two factors named as directivity of an antenna and transmit power are used. The antenna which radiates uniformly in all direction is known as an omnidirectional antenna on the other hand directional antennas radiates in a particular direction. The high value of antenna gain causes stronger the signal in a particular direction. According to the electric and magnetic field theory, depending upon the distance from the transmitting antenna two regions exist such as Far-field and Near-field regions. The two main metrics are used in this field for calculating the EMR, first, is SAR considered for near-field and the second is power density for the far-field. Fig. 23 shows EMR fields. Characterizing the EMR fields in the reactive area is a little bit complicated. The human body easily absorbs the energy transmitted within the reactive near-field region. In the reactive near-field region, the field strength changes inversely with the cube of separation from transmitting antenna i.e., $1/(distance^3)$.

SAR– SAR is evaluated when the body is influenced by EMR. Since the phone is generally near to the body, so EMR from communicating node is typically in the near-field. The value of EMR exposure in the near-field region is calculated for a given period t -

EMR exposure $(J/kg) = SAR.t$

$$SAR = \frac{\delta}{\delta t}\left(\frac{\delta E_A}{\delta M}\right) = \frac{\delta}{\delta x}\left(\frac{\delta E_A}{\Psi \delta V}\right) \tag{8}$$

Where E_A gives the value of energy absorbed by the body, V, M and Ψ represents volume, mass and density of the body, receptively [104]. Different body parts have different SAR value because of different absorption rate as represented in Table XII. *EM Radiation Reduction Techniques–*The target of the 5th generation network is to increase the spectral efficiency, throughput and data rate as well as it also focuses on criteria like improving latency, energy efficiency and reducing EMR exposure from the wireless networks. In this paper some approaches are discussed that are using to decrease the transmit power and EMR exposure levels and SAR of wireless networks because these parameters are directly related to each other. These EMR reduction techniques are:

1) SAR Shielding: Environment and human health [111] have been unpleasantly affected by the increased EMR in the surrounding. The absorption of this EMR by the human body is defined as SAR. For human health higher SAR is very injurious. Different manufacturers have a different value of SAR and this value is governed by several environmental bodies around the world. SAR shielding is used to overcome this problem. In this technique, a metamaterial/ferrite is sandwiched between the human head and mobile, to minimize SAR [112]. The conductivity of ferrite materials is very small therefore there is a small current-induced [113]. The

Journal Pre-proof

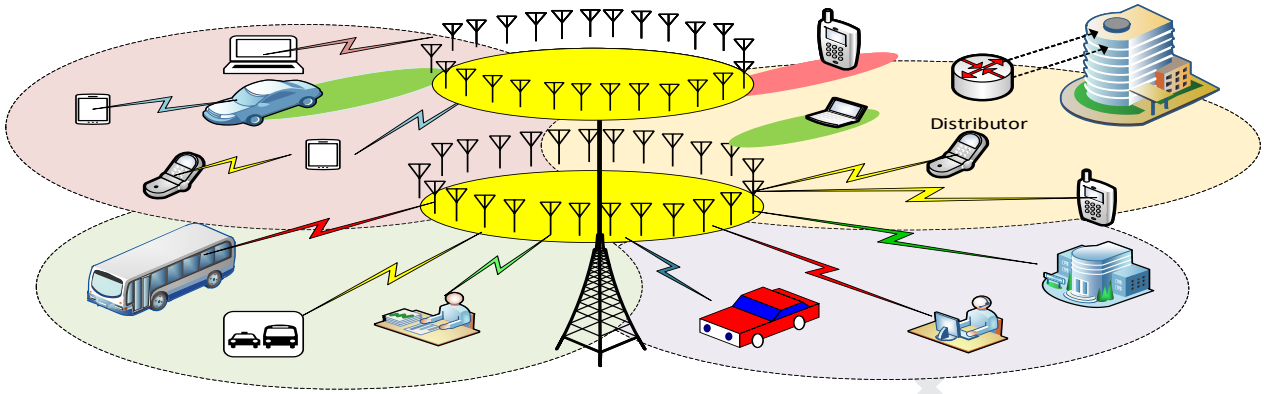


Fig. 22: Massive MIMO antenna array layout

TABLE X: A summary on Massive MIMO techniques

Proposal	Methodology	Scheme/Algorithm	Route	Remark	Reference
To design cell-free MIMO downlink, which provides the service of various single antenna users in the same frequency-time resource.	Work on spectral efficiency with the consideration of the consequence of channel approximation errors, power control, non-orthogonality of pilot channels and backhaul power consumption.	Conjugate beamforming scheme with consideration of local channel state information.	Downlink channel	Design optimal power control algorithm which increase EE as well as access point power constraint and user spectral efficiency.	[94]
Energy utilization in massive MIMO system is optimized.	For a massive MIMO network consider single-cell and multi-cell scenarios.	Bisection searching algorithm	Uplink channel	Optimizing pilot sequences length and find optimal parameters for energy efficient massive MIMO system.	[95]
Work on joint power allocation in source, channels and relay to maximize the EE.	Massive MIMO relaying technique has been used to enhance wireless PHY security, even with no instantaneous legitimate channel state information.	Energy-efficient power allocation scheme	Downlink channel	Energy-efficient joint source and relay power allocation schemes maximize the EE.	[96]
This paper presents the performance of MIMO cellular systems with pilot contamination.	Consider uniformly distributed user location scheme, find out the area EE in a cell.	Power consumption model	Uplink channel	Investigate closed-form estimation of the area spectral efficiency for multi cell multiuser massive MIMO systems.	[97]
Investigate the zero-forcing beamforming, maximum ratio transmission and scaling laws of EE.	Circuit power, transmit power, pilot contamination, and channel estimation errors are considered for improving the EE.	Realistic channel models' scheme	Downlink channel	-	[98]
Design cell free architecture with the consideration of user centric approach to improve coverage.	An interaction between alternating and sequential optimization is considered.	Hybrid analog/ digital beamformers	Uplink/ Downlink channel	To maximize global EE downlink power control algorithms are used.	[99]
Multi-user, multi-cell, state is considered in which every user work to improve their EE.	Integration of antenna arrays are considered	Distributed power control algorithm	Uplink/ Downlink channel	Evolutionary game theory (EGT) is analyzed to maximize the fairness between users in the network.	[100]
Enhancement in system EE, spectral efficiency and reduce energy consumption for massive MIMO system.	Use full duplex method to improve spectral efficiency and EE.	Self-backhauling and energy harvesting jointly used	Uplink/ Downlink channel.	System EE improved by reducing multi-tier interference	[101]

magnetic field region of an incident EM wave is revoked by this metamaterial/ferrite. This metamaterial has negative permittivity. When metamaterial/ferrite is layered between the human head and mobile model, the value of SAR reduces to 47.68% over 1 gram of tissue on the other hand 27% and 52% over 1 gram of tissue at frequency 900 and 1800 MHz.

TABLE XI: Comparison between MIMO and Massive MIMO

Parameters	MIMO	Massive MIMO
Energy Efficiency	Improved EE	100 times improved than MIMO because of the focus of radiated energy on UE.
Capacity	Improved Capacity	Five times improved than MIMO and coverage of lower frequency bands.
System Hardware	Less complex using up to 10 antennas	More complex using upto 100 antennas
Channel modeling	Easy channel modeling	Complex channel modeling
Battery lifetime	Reduces battery lifetime of MIMO based system because of faster processing of complex and computationally intensive signal processing algorithms.	More reduce battery lifetime as compared to conventional MIMO.
Battery spectrum efficiency	Improved SE	Improved SE because of large multiplexing gain as well as antenna array gain
MIMO based system	WiMAX (IEEE 802.16), WLAN (IEEE 802.11), 4G systems, LTE	Advanced WLAN (802.11ac, 802.11ad), LTE Advanced, 5G system.
Operation Mode	Operate in frequency division duplex (FDD) and time division duplex (TDD) both.	Massive MIMO was originally invented for the TDD operation, but can potentially be applied also in FDD operation.
Throughput	Having high downlink and uplink throughput because of multiple antennas and spatial multiplexing technique.	Large improvement in throughput

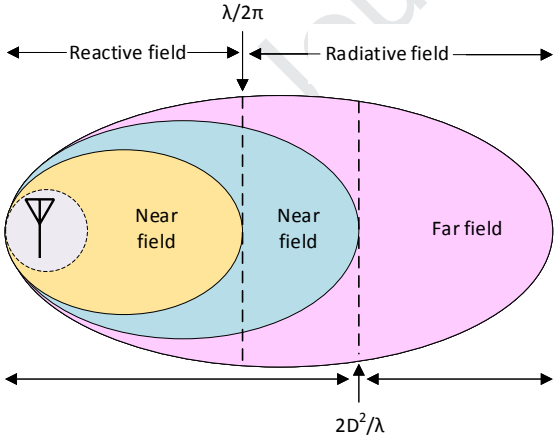


Fig. 23: Effect of EM radiation fields

Recently, the German Federal Office has published a list of top 16 smartphones emitting the most radiation with a SAR. The Xiaomi Mi A1 releases the most radiation with a SAR of 1.74 watts per kilogram while the ZTE Axon Elite and Samsung Galaxy Note 8 have emerged as the lowest radiation-emitting

TABLE XII: Different SAR calculation for different body parts

Metric	Description	Calculative Formula
Whole Body Averaged SAR (WBA-SAR)	Defined as the total power take in by the body to the mass of the body.	$WBA_{SAR} = \frac{1}{M_B}\int_{R_B} SAR(\delta m)$ Where M_B, R_B shows the mass and region of the body
Specific Organ Averaged SAR (SOA-SAR)	This depicts the absorption of EM fields by the tissue or specific organ of the body.	$OSA_{SAR} = \frac{1}{Mo}\int_{Organ} SAR(\delta m)$ Here $Mo, Organ$ shows the organ mass and organ under consideration.
Spatial Peak Averaged SAR (SPA-SAR)	Discussed as maximum local SAR averaged over organ specific mass.	-

phones with a SAR of 0.17 watts per kilogram.

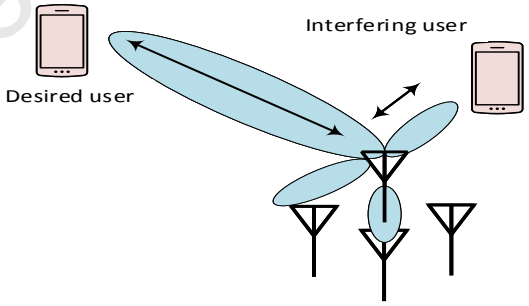


Fig. 24: Principle of Beamforming.

2) Beamforming: A direct approach to reduce power consumption from the mobile communication system is beamforming. An array of antennas is used in this technique to focus EMR towards every desired mobile phone user direction. In the beamforming approach, each transmitting signal of the mobile subscriber is multiplied by complex weights. These weights are adjustable and have a signal's phase and magnitude in a specific direction. By using the DSP algorithms, the exact location of the user can be tracked in the beamforming technique. So, the signal strength in a specific direction maximizes while interference, EMR exposure and transmit power reduces [114]. Fig. 24 shows the principle of beamforming. In figure, the beam formation is in the direction of the desired user. Table XIII represents a detailed survey of the beamforming techniques in terms of energy saving.

I. Futuristic Approach

In this subsection, various important futuristic approaches like Device-to-device (D2D) communications, Mm wave and Green IoT is discussed in an informational fashion.

TABLE XIII: Survey of Beamforming technique

Proposal	Remark	Reference
Authors investigate the effect of beamforming for SAR improvement.	Reducing the absorption of EM radiation by placing nulls in the head direction. Results indicate that a SAR reduction can be achieved at least 10 dB in this method.	[105]
To overcome the problem of EM radiation exposure radiated from mobile equipment, a less complex, SAR-aware optimum beamformer is investigated.	The SAR-aware beamformer for multi-antenna terminals is used to improve transmission power constraints and SNR. Results indicate that the SAR-aware beamformer performs better than the traditional method and provide SNR gain up to 5 dB.	[106]
Propose hybrid beamformers to fulfill the objective of reducing power while maintaining users' QoS requirements.	The author presents a phase array switching framework to comprehend the eigenmode selection beamforming. Results show that the investigated eigenmode selection method can attain a significant amount of power savings.	[107]
An energy-efficient multi-cell multi-cluster coordinated joint beamforming scheme is investigated.	The author works on antenna specific higher power constraints to overcome the problem of non-linear distortion in power amplifiers. It represents the relation between beamformers and binary antenna selection variables and numerical results illustrate the gains achieved in the form of EE.	[108]
In this paper user scheduling problem and joint multicast beamforming was discussed and another goal is to minimize the total transmitting power of multiple channels.	An algorithm based on Convex relaxation is proposed. Extensive mathematical researches illustrate the benefits of the suggested co-design model over random scheduling and fixed scheduling in terms of power ingesting.	[109]
At the transmitter end, a digital-beamforming method depends on DSP was investigate	The solution presented high throughput and improved data rate for a cost-effective 5G wireless access network which operates at frequency 28 GHz and also represents an improvement in EE.	[110]

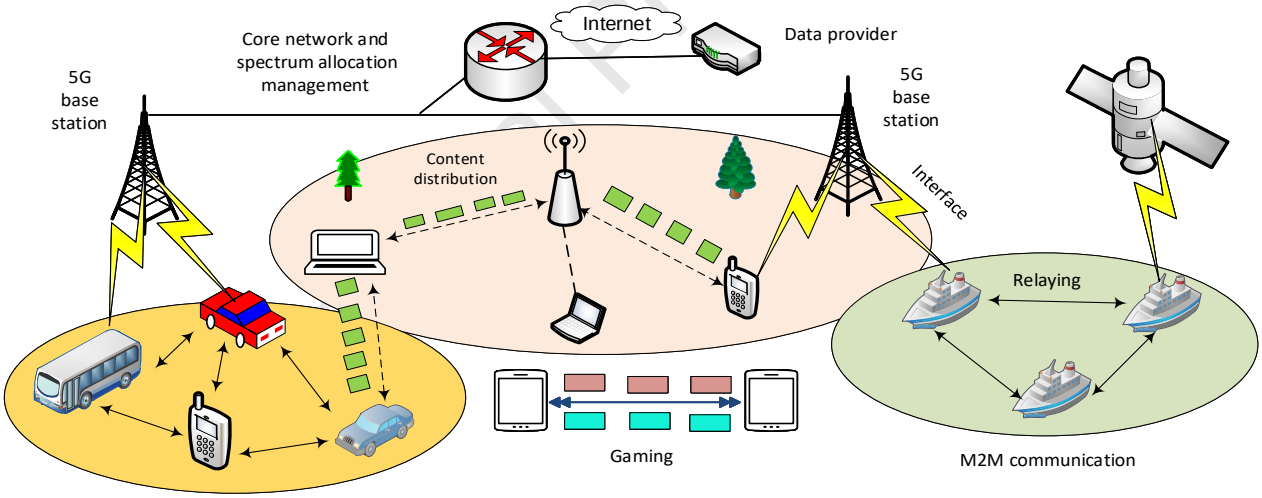


Fig. 25: Energy efficient D2D communications model

1) D2D Communications: D2D transmission is one of the effective technical approaches to reduce the consumption of power and improve the data transmission rate. In this technique two physically separated nearby located nodes can directly communicate with each other with low transmit power [115]. Thus, a remarkable amount of energy can be saved at BSs. In the case of wireless network unavailability and network congestion, D2D communications offer alternative communication. The D2D approach is recognized as a public safety network for future wireless communication by FCC (Federal Communications Commission) because of low cost and high data rates.

An energy-efficient high-level D2D communications model is represented in Fig. 25. Here different users are linked with BS via a cellular interface. These users are also communicating directly via D2D links between nodes. Several Radio Access Technologies (RATs) including 3GPP evolution like Bluetooth and Wi-Fi standards are used to establish direct links between communicating nodes such as mobile phones and laptops. In Vehicular to Infrastructure (V2X) communication, a dedicated 802.11p standard operates in the frequency band of 5.9 GHz is used by cars in an intelligent transport system (ITS).

TABLE XIV: Comparison of existing surveys of D2D in terms of energy efficiency

Proposal	Methodology	Technology	Route	Evaluation	Achieved Performance	Reference
Enhance power efficiency	Exhaustive search	LTE-A	Downlink	Simulation on system level	20% energy can be saved	[116]
Improve power efficiency	Heuristic algorithm	LTE	Uplink	Simulation on system level	30% power efficiency improved	[117]
Increase power efficiency	Linear programming	OFDMA	Uplink/Downlink	Simulation on system level	40% to 80% energy saved	[118]
Enhance EE	Linear programming	OFDMA, LTE-A	Uplink/Downlink	Simulation on system level	50% energy saved	[119]
Improve EE	-	LTE	Uplink	Simulation on numerical based	Improved EE from 0.9 bps/Hz/mW to 22 bps/Hz/mW	[120]
EE management	Inducement mechanism based on open market and sealed market model for D2D (auction algorithm)	OFDM	Downlink	Framework for cooperative and D2D communications	37% energy saved	[121]
Analysis of EE	-	LTE	Uplink	-	Energy saved bps/Joule	[122]
Analysis of EE	Distributed resource allocation algorithm	LTE	Uplink	Tradeoff between EE and SE	Energy saved in bits/Hz/J	[123]

Depends on the spectrum usage, this approach is divided into two categories named inband D2D and outband D2D. In first, the identical licensed band is exploited by both D2D and cellular networks. The major drawback of inband D2D communications is interference prompt by D2D to the cellular link. The unlicensed spectrum is utilized by D2D links in outband D2D communications. The reason behind the usage of the unlicensed spectrum is to limit the interference between cellular links and D2D network. But, more interface requires to connect different wireless technologies like ZigBee, Bluetooth and, Wi-Fi [124].

In [125], joint optimization of adaptive power allocation and uplink subcarrier assignment-based approach is investigated for cellular and D2D network to achieve improved EE. Various power controlling algorithms used for D2D communications have been explained in [126] [17]. Power and resource block allocation using game theory has been discussed in [127]. The author represents a cross-layer optimization scheme in conjunction with nash equilibrium under the accurate availability of channel state information (CSI) [128]. In [129], the author suggests a joint scheme of power control and resource allocation. The adoption of a two-layer method is involved, to achieve an optimal solution for energy efficient architecture. Table XIV provides a complete survey of D2D in terms of energy efficiency.

2) Mm Wave: For next-generation networks, Millimeter waves are the most promising approach. For delivering fast multimedia services, high-quality audio-video, and real-time services, a large amount of bandwidth is required. Due to a large number of consumers, the traffic congestion problem in the network will be arisen by 2030. To solve this problem the fifth generation network uses millimeter wavelength. The signals are operating between 30 and 300 GHz and being shifted to a higher spectrum. On the other hand, this technology has a large penetration loss, high signal attenuation and extremely susceptible to blockage effects which lead to the limited coverage area. Spatial multiplexing technique is used

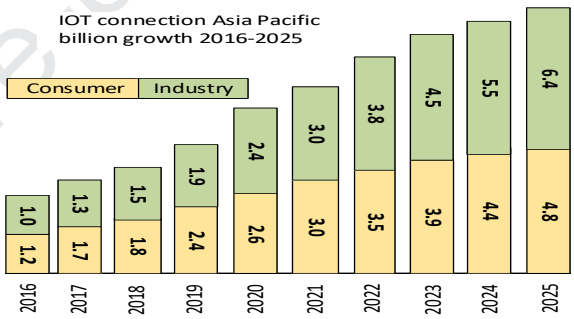


Fig. 26: Growth rate of IoT in Asia Pacific

for transmission and reception because the wavelength is very small. To obtain the high capacity, fast data rate and reduction in energy consumption, an array with directional transmission between MS and BS is used in this technique [130] . Table XV provides a complete survey of mm wave technique in terms of energy efficiency.

3) Terahertz communication band: In the upcoming decade, it is expected that the data rates will be increased up to 12 Gbps or more. This requirement of a high data rate cannot be fulfilled by a licensed band spectrum [137]. In addition to this mmWave technology alone cannot realize the need for a high data rate since there is only 9 GHz bandwidth for data communication. Free Space Optical (FSO) communication systems may be one of the solutions with minimum operating power for this demand but it requires direct line of sight (LOS) communication between the transmitter and receiver, which makes it impractical for mobile communication. So the most promising solution to achieve the demand of high data rate is Terahertz (THz) communication band which is least considered frequency region, where the data can be transmitted over non-line-of-sight (NLOS) links. Terahertz Band com-

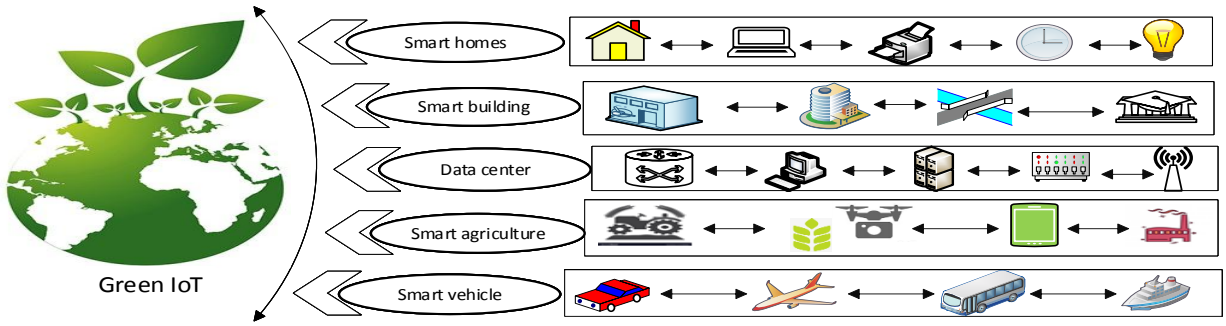


Fig. 27: Green IoT Architecture

TABLE XV: Comparison of existing surveys of Mm wave in terms of energy efficiency

Proposal	Key Points	Consideration	Model	Reference
Deployment of optimal BS is investigated	To reduce interference of two BSs, precoding between multi-BS and single BS coordination technique and different antenna angles are considered.	Attenuation coefficients and packet loss ratio	Path loss model	[83]
Non-line of sight and Line of sight link model investigated.	Green metrics, power efficiency, network latency,and area energy efficiency are calculated.	73 GHz frequency band has better SE and 28 GHz frequency band has better EE.	LOS and NLOS model	[131]
Downlink power control algorithms are investigated to enhance the EE	Path loss is calculated for different scenario.	licensed and unlicensed both applications	Large-scale, small scale path loss models and LOS probability model	[132]
Adaptive power control method is investigated	Interference between cellular and D2D user is calculated.	Radio resource allocation and management in wireless networks.	Power allocation mechanism	[133]
Implementation of full duplex relaying	Circuit power consumption and relay interference elimination.	Energy efficiency oriented solutions.	Joint beamforming	[134]
Beam-alignment protocol based energy-efficient architecture deployment.	For higher range, fractional search decreases power expenditure by 2dB with respect to bisection.	optimality in fractional search method	Markov decision model	[135]
Dynamic BS sleep control and RF chain activation are investigated.	Considering centralized and distributed both schemes for designing traffic aware dynamic scheduling architecture.	property of hybrid precoding	Greedy algorithm	[136]

munication will improve the spectrum scarcity and capacity of wireless communication networks. In the upcoming future THz bands are predicted for the development of various applications in both nanoscale communication and classical networking field [138]. The frequency region above and below the THz band is the infrared and microwaves. The THz frequency band (between 0.1 THz and 10 THz) will provide ultra-high-speed data transmission within the areas up to 10 m (PAN). The necessity to deal with this rapid progression of wireless services and related applications has required a large research activity that explores the optimal options for designing of user-oriented context-aware next generation (5G) wireless communication network. The 5G network is not only expected to improve the performance of information and communication devices but also capable to solve many pervasive issues of this field. The several applications related to 5G green communications in THz band are represented in Fig. 28.

4) Green IoT: The goal of the Internet of Things (IoT) is to link millions of users with each other using sensors like RF identification, and is the most evolving technology for making the foundation of the green CRNs. Large energy is required for exchanging information among millions of users. Fig. 26 represents the growth rate of IoT in the Asia Pacific in terms of consumers as well as industry. The goal of Green IoT is to enhance an EE of IoT devices and making the environment healthy. In [139] author projects that the usage of RFID will grow exponentially in the upcoming years. So, if this forecast is even nearby the reality, then a large amount of energy consumption will arise due to the more battery power will be demanded by all active RFIDs. So, to solve this problem, this work focuses on making the IoT technology green by applying numerous approaches. The IoT framework management involves identification, sensing,

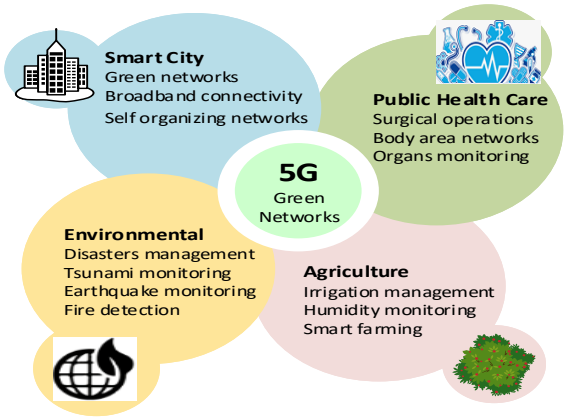


Fig. 28: Future applications related to 5G green networks

analyzing, communication, computation facilities and semantics. Identification is the initial process which guarantees that needed information sends to the accurate destination. Sensing works on collecting information and data, then sends to data centers. Considering different parameters and conditions, this collected data is then scrutinized for different services and operation purposes. The different parameters sensed by sensors are real time traffic, humidity, pressure, rate of pollution, temperature, etc. Combining different networks and objects for performing any task will comes under communication. Technology likes Bluetooth, Wi-Fi, etc are used to complete communication. The microprocessors, microcontrollers, embedded systems, FPGA, CPLD, ASIC, and various software applications perform the computation. Green IoT concept visualizes in Fig. 27. The figure shows that for the design of a smart city every element of a smart city like smart homes, smart agriculture, smart building, etc are connected by the IoT. People will use the internet for different purposes like traffic control, traffic monitoring, medical purpose, etc [140].

The main motive of green IoT is to make the environment green from its designing level to its implementation level. There are two ways to achieve this objective. The first is that the IoT will aid in eradicating the greenhouse effect produced by present applications and second is to optimize itself to reduce the greenhouse effect. Different scientific approaches for designing of green IoT are discussed in [141]. A framework is proposed for constructing the green IoT in [142]. Moreover, green cloud computing [143], green sensor networks [144], green RFIDs and green data-centers plays an important role in designing green IoT.

The IoT is developing technology for future green networks that have capability to change the way we observe IT industry. The green IoT approach is going to have an enormous influence on how we handle with convinced glitches in our everyday life and it is surely going to make our survives easier and better. There are five principles of achieving green IoT. Considering a small network size will save a large amount of energy and improved EE. Select only desired information by proper sensing and eliminate extra data also helpful for energy saving. The integration of different networks and implementation with proper policy making will lead towards green IoT and reduce carbon footprints. There should be a proper tradeoff between parameter is also very necessary. Fig. 29 describe the classification of green IoT techniques and Table XVI represents a detailed comparative survey of these techniques.

J. Green Cognitive Radio Networks Approach

According to the report of CISCO on data uses, in 2017−22, the telecommunication traffic originated by mobile subscribers is predictable to have nearly tenfold [155]. For new mobile subscribers and upcoming data-hungry devices, the availability of unused frequency bands is very low. Because, a report published by the FCC says that, nearly complete accessible bandwidth for wireless network communications has been almost allocated [156], [157]. Also, a study utters that nearly 75% of the assigned bandwidth is not properly used due to the static spectrum allocation scheme [158]. This generates an abundant gap between the existing requirement of additional bandwidth and presents an allocation scheme. The CRNs are the most promising, emerged and key enabling technology of future networks. This technology provides the freedom to utilize underutilized frequency bands by allowing Secondary Users (SUs) to access licensed frequency bands which are not used by Primary Users (PUs) opportunistically [159]. When the primary user comes back to the channel, the SU has to switch to another unoccupied channel belonging from the different network for consistent communication. The performance of SU has affected by the appearance of high prioritized PUs. Fig. 30 represents an idea behind green cognitive radio networks (GCRNs). The Fig.30a shows existing static spectrum allocation policy which fails in heavy traffic load situation and frequency bands are highly utilized. In Fig. 30a, the number of mobile subscribers are more in-network 'A' and cause heavy traffic load, whereas the other networks B, C, D and E have less traffic load with underutilized frequency bands. Fig. 30b, represents the difference between CR and non-CR technology. Fig. 30c represents an optimal utilization of available frequency bands (called white space) by using dynamic spectrum allocation policy. The mobile node can switch from one network to another network after making the decision using the cognitive property. A dynamic spectrum allocation policy requires key functions like dynamic spectrum environment, spectrum sensing, sharing and management respectively, in GCRN as shown in Fig. 30d.

It is very challenging to design a green CRN because it involves the optimization of dynamic spectrum allocation as well as optimal exploitation of green energy. The various functionalities are performed in the energy-efficient CR.

- Spectrum Sensing and Analysis- For designing the energy-efficient spectrum sensing approach the main task is to analyze the performance indicators to find out the presence of the primary user. Various types of spectrum sensing with energy efficiency related issues are represented in Fig. 31.

Journal Pre-proof

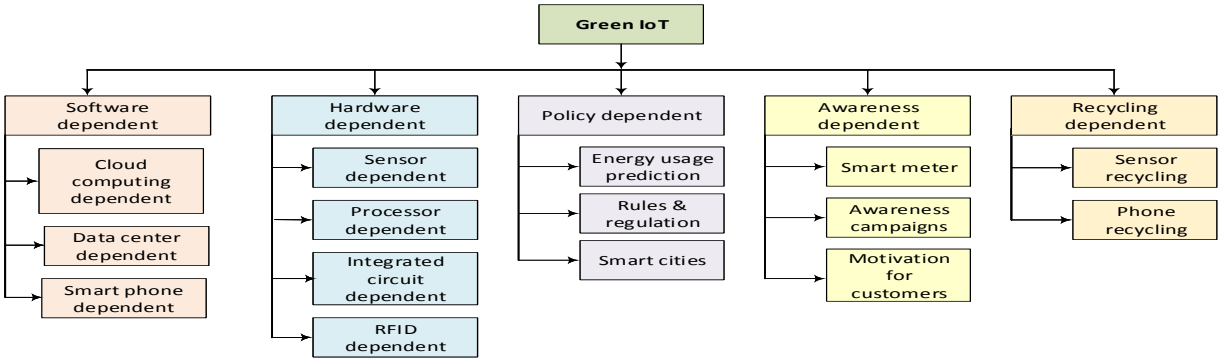


Fig. 29: Classification of Green IoT techniques

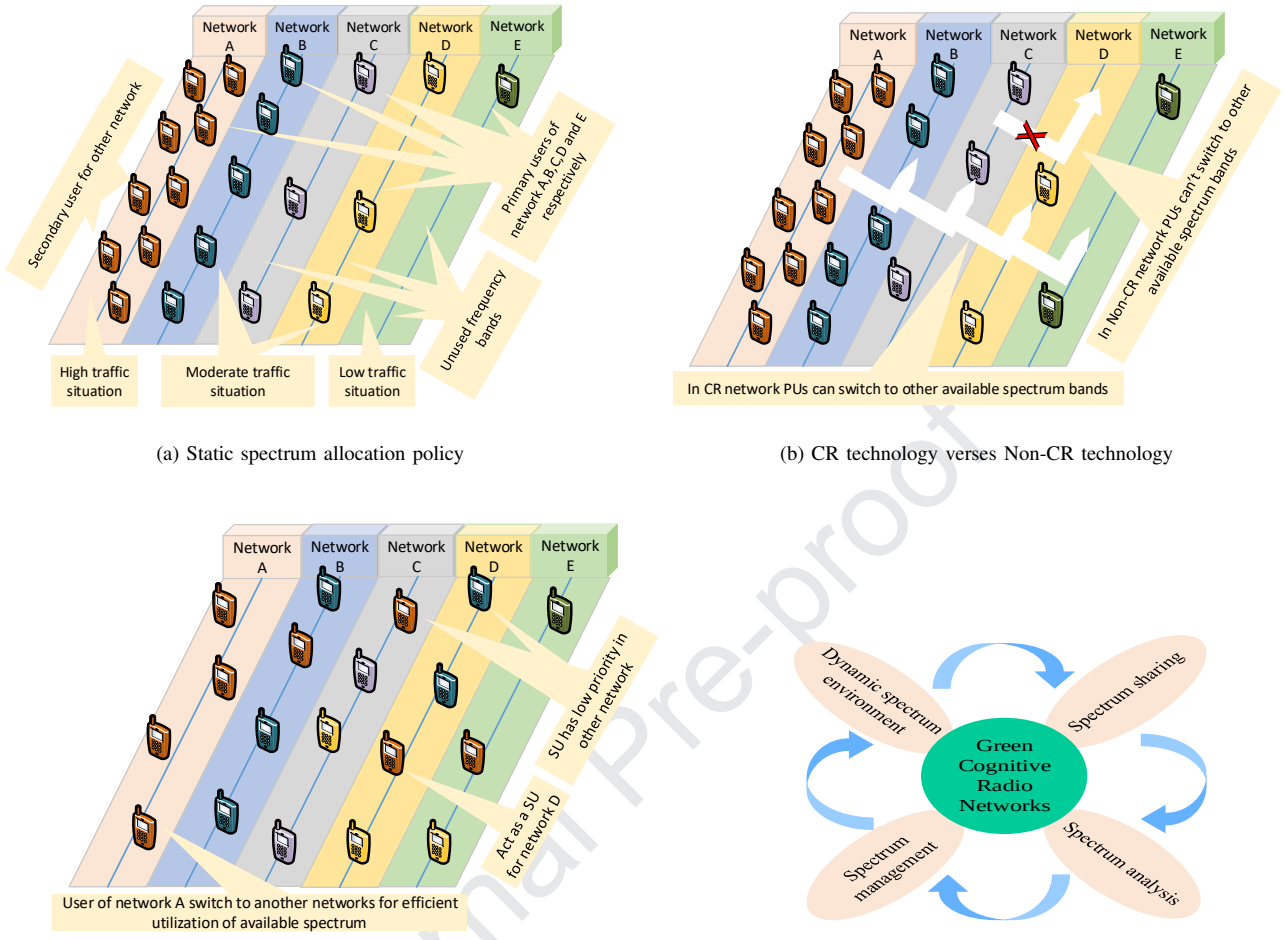
TABLE XVI: Comparative survey of green IoT techniques

Approach	Consider Technology	Energy saving process	Tradeoff	Real time implementation	Metric	Reference
Hardware dependent	Relay	Load balancing among different nodes	Communication delay	Possible	End to end delay	[145]
Software dependent	Sensors	Selective sensing	Privacy, energy overheads	Highly possible	Energy efficiency	[146]
Policy dependent	Information management	Information collected from various section of building to develop energy efficient guidelines.	Data security	Highly possible	Data efficiency	[147]
Software dependent	Cloud computing	To decrease data traffic dynamic packets are downloaded from access points.	QoS may be compromised	Possible	Packet transmission rate	[148]
Software dependent	Smart phones	Calculation of energy consumed by smart phone applications	Security and data privacy	Possible	Energy efficiency	[149]
Software dependent	Data centre	Workload sharing among physically separated data centers	Highly complex	Not possible	Data center efficiency	[150]
Hardware dependent	Integrated circuits	Inbuilt sensors on-chip are used to decrease network traffic load	High cost	Possible	Radio efficiency	[151]
Awareness dependent	Smart metering	Track various types of devices and evaluate the energy consumption rate.	Security and data privacy	Possible	Power efficiency, transmitted power	[152]
Policy dependent	Smart buildings	Guidelines and approaches to reduce energy consumption.	User dependency	Not possible	Energy per bit per area	[153]
Hardware dependent	RFID	Passive sensors are used for evaluation	Communication delays	Possible	Spectral efficiency	[154]

- Green Relaying in Cognitive Radio- In CRNs, authentic end-to-end data transmission among secondary users need a large amount of power, which leads to harmful interference to primary users. Relay technology is capable to boost the overall performance by saving energy. Because of the shorter transmission distance, the path loss is very low in this technology.
- Cooperative Cognitive Radio Networks- Cooperative communication is considered as the most promising solution to tackle the issue of multipath fading. This provides energy-efficient communication and also reduces the detection time due to the usage of spatial diversity. Cooperative communication protocols are working in two modes named as Decode-and-Forward and Amplify-and-Forward. The cooperative communications in CRNs, can be classified into two classes such as cooperation among secondary users and cooperation between the primary user and secondary users respectively.

Now the question is how do CRNs reduce power consumption and provide green communication? So, the answer is for wireless communication system bandwidth efficiency is always a pivotal concern and there exists a rich state of the art on this topic, focusing on improving bandwidth efficiency. The solution lies under Shannon's capacity formula which describes the trade-off between power and bandwidth. The capacity various logarithmically with power, but linearly with bandwidth. It reflects that to decrease power, larger bandwidth will be required or in other words, handle the spectrum dynamically and optimally. This is the principle of CR. Table XVII having a detailed survey of cognitive Radio technique in terms of energy saving.

A complete comprehensive survey of all techniques dis-



(a) Static spectrum allocation policy

(b) CR technology verses Non-CR technology

(c) Dynamic spectrum allocation policy

(d) Key function for Green CR networks

Fig. 30: Concept of Cognitive Radio

cussed in this paper is given in Table XVIII and continue in Table XIX.

IV. Challenges and Existing Practical Implementations of Green Communication

The framework to enhance the EE of telecommunication the system has a holistic view which considers possibly all facts and figures of an ecosystem for green communication. The several challenges related to green communication are listed as:-

- For estimating any energy efficient solution, the required energy for the manufacturing of a communication system (called embodied energy) has a very significant role.
- The surge of EE at a single stage of the communication system might be the reason for the decline at another stage of the system. For instance, the usage of a cognitive radio system may decrease the transmission energy by using proper bandwidth but drain additional power in the process of sensing the spectrum hole of the second network.
- There should be a respectable trade-off between EE and QoS. Like the aforementioned usage of BS sleep modes technique, more energy can be saved during off mode of some cells. But, it is compulsory to stipulate the cell waking time in advance as well otherwise, the QoS may degrade.
- There should be also a proper trade-off between flexibility and EE which depicts in Fig. 32.

The energy-efficient solution depends on certain guidelines, concepts, parameters, energy consumption models and considerations. The energy consumption of the telecommunication system is load-dependent and implementation-dependent. These are the challenges for modeling the energy- efficient system. The EE metric for a green telecommunication system is generally expressed in the form of performance per unit of energy. The performance usually measured in terms of efficiency and throughput for the energy-efficient communication system. Other key challenges for designing the green energy-efficient system are cross-layer adaptation, system reconfiguration, load balancing, and multi-domain scheduling. It is necessary to gauge EE from a new perspective for

TABLE XVII: Comparative survey of CR technique

Objective	Description	Energy saved	Scheme/Algorithm	Reference
Cultivating the EE of cellular networks	The author investigates the cooperation between BSs and television stations in sending UE relying on an opportunistic usage of the TV spectrum. The unoccupied TV spectral frequency band is utilized in an opportunistic manner.	41%	Explores the joint employment of cooperation techniques and cognition	[160]
For CRNs examine the EE with the relation of area spectral efficiency	A constant base station power consumption model is considered for both perfect and imperfect detection in the case of cognitive as well as primary users.	35%	Stochastic geometry approach	[161]
The author has described the method to attained maximum EE in a spectrum sharing atmosphere.	With consideration of EE, the trade-off between throughput and transmit power in a centralized CRNs is explored.	28%	Spectrum-sharing model	[162]
The effect of interference power, interference gain on energy consumption and optimal sensing time of the channel is described.	For multiband CRNs power allocation and optimal sensing time algorithm are proposed.	30%	Multiband sensing-based scheme	[163]
Research work is dedicated to energy-efficient optimum power allocation tactics for green CR in fading channels.	For both delay-insensitive and delay sensitive CR, EE improvement schemes are investigated with the consideration that the channels are suffered from fading and primary user coexists with a secondary users.	25%	Convex optimization methods and Dinkelbach's method	[164]

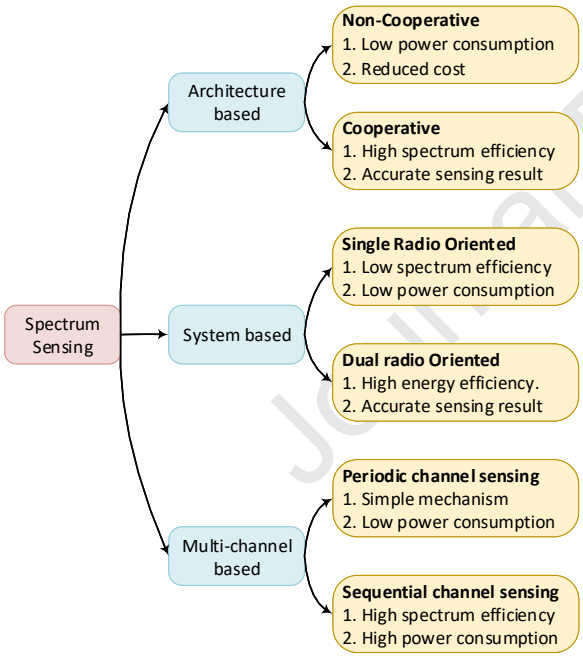


Fig. 31: Various spectrum sensing techniques

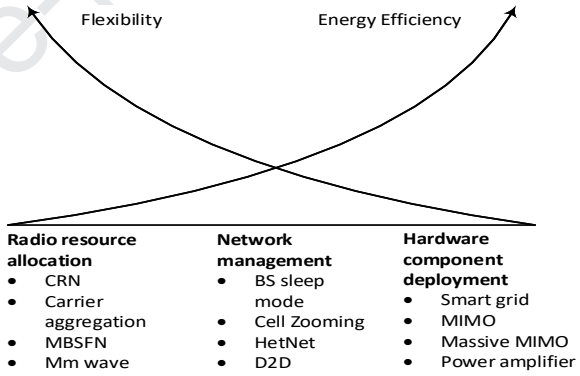


Fig. 32: Tradeoff between flexibility and EE

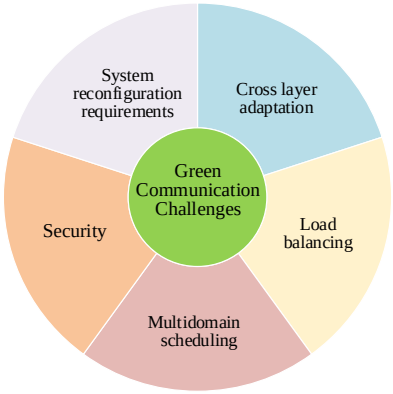


Fig. 33: Green communication related challenges

designing and developing novel energy-efficient architecture. The pathway represented in Fig. 6 to implement in the real world for making communication green is also having several challenges that need to be paid consideration too in this research area.

Several challenges, related to green communication for next generation networks are represented in Fig. 33.

There are some international projects with the aim of exploring possible ways to reduce CO_2 emission, transmitted power of telecommunication system, health effects, and EMR exposure are discussed in Table XX

Journal Pre-proof

TABLE XVIII: Assessment of the Energy-Efficient strategies

Approaches for Green communication	Techniques	Metrics for Green performance evaluation	Saved Energy/Reduced power	Parameter Consideration	Applied Domain	Merits	Demerits
Hardware Component Improvement	Doherty architecture and GaN based amplifier	EE, PA Efficiency	Up to 51%	Peak-to-average power ratio	At infrastructure level	Energy can be saved in a large amount	Less flexible
	Switch mode power amplifier	Energy Consumption Rate	70%	Linearity	At infrastructure level	A large amount of energy can be saved	For hardware replacement large cost required, some upper limits for hardware implementation
Energy Efficient Architecture	Hetnet	CapEx and OpEx		Interference, handoff	Always macro cells remain in active mode, while small cells will come in active mode when traffic increases.		Required high radio interference management
	1. Microcell-Macrocell	CapEx and OpEx	45%	Interference	At infrastructure level	Low implementation cost	Handoff management
	2. Picocell-Macrocell	CapEx and OpEx	62%	Handoff	At infrastructure level	Prolonged handset lifetime	Handoff management
	3. Femtocell-Macrocell	CapEx and OpEx	80%	Interference	At infrastructure level	Higer SNIR	Interference management
	4. Femtocells (urban area)-Marocell (Suburban area)-Portable femtocell (rural area)	CapEx and OpEx	Reduced by 83% to 87%	Handoff	At infrastructure level	Improvement in EE and low transmission power	Handoff management
	5. Femtocell (Highly populated urban area)-Picocell (Less populated urban area)-Microcell (suburban areas)-portable femtocells (rural areas)	CapEx and OpEx	79-80.2%	Interference	At infrastructure level	Low path loss	Interference management
	6. Macrocells (Fully coverage area)-femtocells, picocells, and microcells (border regions)	CapEx and OpEx	10%	Handoff	At infrastructure level	Improve coverage	Handoff management
	Cell Zooming	CapEx and OpEx		Problem of dead zone, coverage holes	BS coverage area continually changes with variation on the real time traffic		Maintain tradeoff between energy saving and performance
	1. Zoom out Active cell in Commercial area	CapEx and OpEx	30%		At infrastructure level	Inexpensive implementation and testing	Handoff management
	2. Zoom out Active cell in Residential area	CapEx and OpEx	38-40%		At infrastructure level	Improve throughput	Interference management
	3. Zoom out Active cell in uniformly distributed area	CapEx and OpEx	25-40%		At infrastructure level	Proper traffic load balancing	Handoff management
	4. Zoom out Active cell in Office area	CapEx and OpEx	45%		At infrastructure level	Improvement in EE	Interference management
Traffic based energy efficient	Base station ON/OFF	Sleep mode energy saving, Energy consumption index, performance indicator, Energy consumption gain		Battery life of UE, coverage	Mobile users transfer connection from off BS to on BS	Energy savings depend on the time period over which each of the BSs is switched off	Maintain trade-off between energy saving and QoS, less accurate in modelling
	1.Distance based evaluation		29%				
	2.Real time Traffic based evaluation.		30-35%				
Time based energy efficient	1. Unicast	Electricity cost, state of battery	38.5-45%	Proper synchronization	LTE BS	High energy reliability, deactivate unused subframe	Degrade battery life of UE

Journal Pre-proof

TABLE XIX: Assessment of the Energy-Efficient strategies

	2.MBSFN		56%				
	3.DTX		80%				
	4.Optimize subframes in each frame		90%				
Frequency based energy efficient	Carrie Aggregation	Latency, gain and axial ratio	Power reduces up to 65%	Compatibility, Complexity	4G LTE	Power reduce by double factor band	Because of early hand-off, radio link failure increases, Difficult for two networks to work together
Momentous Approach	Renewable energy	Cost efficiency	10%, energy savings depend on BSs installation and location, speed of wind	BSs location, other natural energy resources	At infrastructure level	Less expensive, long-term solution for off grid	High replacement cost, for on grid BS limited gain
	Smart Grid	Cost of electricity, on-grid energy consumption	Up to 30%	Resource management, QoS	At infrastructure level	Energy savings through reducing consumption, Improved environment	Required proper infrastructure
Space based energy efficient	MIMO	Spectral efficiency, Energy efficiency	55%	Effect on the UE	At infrastructure level	Improved capacity and spectral efficiency	High circuit power, reduces battery lifetime
	Massive MIMO	Transmission Power, Energy consumption	Depends on number of switched off antenna	Spatial multiplexing, space time block coding	At infrastructure level	Large throughput	Complex channel modeling, more loop interference and training symbols
EM Radiation based energy efficient	SAR Shielding	Power density, Whole body averages SAR	A large amount of energy can be saved	Signals coming from the sensor array	At UE level	Save body organs and tissues from harmful waves	Enhancement depends on a special material and design
	Beamforming	SAR exposure	A large amount of energy saved	Signals coming from the sensor array	At antenna level	Improve system coverage and capacity, improved SINR in the desired direction	High hardware complexity. Advanced processing DSP chip is required. High Battery drain due to large power requirement
Green Cognitive Radio Network		Throughput, network energy cost	Upto 50% by proper spectrum management	Spectrum hole, unutilized bandwidth	Performance depends on real time load distribution information of whole network	User oriented, low cost, various application	Performance depends on spectrum unavailability, Delay is involved due to sensing operation
Futuristic Approach	D2D method	Average sum rate, BS power consumption	15-40% Depends on the number of intermediate devices	Switch off the devices for a deterministic traffic	Device on off depends upon the distance between the transmitter and receiver	Low transmitter and circuit power	Packet encoding and decoding is required, Interference management, D2D mobility management, Synchronization among devices
	Mm wave	Data rate, outage probability, coverage probability	Up to 20%	Privacy, security	At antenna level	Cost-effective	High circuit power and signaling overhead required atmospheric loss, pathloss and shadowing
	IoT	Energy efficiency, end to end delay, weighted throughput	Up to 30%	Compatibility, complexity	At infrastructure level	Reduce human effort, improve security	Complex network, device recognition and difficult to locating in IoT

V. Discussion and Forward Way Research Direction

Various energy-efficient approaches examined in previous sections are valuable to design energy-efficient green communication system and network. Integrating all these approaches into a combined approach can lead the revolt in the field of green communication. The next-generation networks are anticipated to realize the necessities of the mobile subscribers as well as deliver a solution for energy-efficient green communication.

Green communication networks have been broadly subjugated in bestowing a plentiful range of real-time multimedia applications. The Fig.34 demonstrates the extensive summary of upcoming future networking technologies likes Software Defined Networks, Supermassive MIMO, Terahertz Communication, Ultra-Dense Networks, IoT Communication, Quantum Communication, Beamforming and NOMA. However, the future research direction and issues related to green communication networks are also summarized in Fig.34. On the basis of studied literature the future research directions can be summarized as:

- In supermassive MIMO, usage of multiple antennas requires very high power consumption and a great amount of overhead. This problem can be overcome using the spatial diversity concept.
- The EE of terahertz communication in multiple cell scenario requires to be explored to eradicate interference-effect.
- The consumption of high power in communication system circuits is a critical issue in the usage of renewable energy for designing a green communication system.
- The cost factor involved in the practical realization of IoT technology is another significant research subject for the future.
- The power control strategy in hardware and efficient algorithms in the software of the IoT system can be used to achieve optimum SNR and reduce interference.

To realize the green communication network environment, an intelligent advanced energy-efficient technique requires to be developed. Different energy-efficient techniques have been immensely deliberated in this survey. However, there are still new research areas for more effort to be achieved in this field. This part of the section delivers future research directions that entail the consideration of the research community. These new directions are discussed as:

1) Cooperative Heterogeneous Network (HetNet) Architecture: The upcoming generation of networks is likely to provision incorporation of various networks having diverse services and protocols. So, vertical handover, management of interference, enhancement in network capacity with the attention of energy efficient communication system is an important research topic. The tradeoff between spectral and energy efficiency and a tradeoff between network coverage and QoS in HetNet network also requires more investigation.

2) Green Networks: Energy-efficient approaches are the main prerequisite for the designing of GCRN. Thus, a new research area to minimize the EMR effect by the green networks is desired to be investigated. Two key research issues associated with GCRN are (1) decrease in consumption of power and energy overhead instigated by communication among mobile nodes and (2) optimization of dynamic spectrum allocation techniques to improve EE without distressing QoE of mobile subscribers.

3) Experimental Testbeds for Green Networks: For analyzing an impact of suggested energy-efficient approach for green networks, it is vital to test these approaches on realistic testbeds. In this way, the hardware realization issues can be identified and would be able to check the viability of these proposed approaches. The absence of mature green networks modules in prevailing simulators offer new research direction to enterprise an experimental testbed to provision green networks.

4) Security: Auspicious of protected communication in green networks is a significant design objective for consistent advance next-generation wireless networks. It is a challenge to design an energy efficient approach for green communication to handle the EMR and GHG effect jointly with the utilization of RER for saving energy. In the available literature, energy-efficient approaches are focusing on the security of either network or mobile nodes. Secured green communication for both networks and mobile nodes, along with energy-efficient encrypted technique is required in future wireless communication systems.

5) Green networks Model for LTE/Wi-Fi coexistence: LTE network supports the modern wireless networks like carrier aggregation, massive MIMO, HetNet, D2D network with multiple cells of small size. The adaptive energy-efficient approach designed for green networks would be a promising technology in such advanced wireless networks. For example, HetNet comprises multiple cells of different sizes (termed as femto, pico, and microcells). So, the management of interference causes more complexity in networks. The coexistence of two different networks with combining their capabilities and energy- saving proposals would help in the formation of green networks, which is still an energetic research direction and will receive further attention.

Fig. 35 represents the advanced technologies i.e. next step towards futuristic wireless networking. These technologies are blockchain, machine learning, artificial intelligence, big data, advanced beamforming.

VI. Conclusion

This paper scrutinized the trends of the telecommunication system in the last decade which visualize a move towards following energy- efficient green communication for the purpose of designing next-generation networks. For each energy efficient technique, an outline of the existing state of the art research, merits, demerits, open issues, challenges, and possible forward way research direction has been investigated.

Lessons Learned by the survey

- *Lesson 1*: Select an appropriate EE metric for performance measurement at facility, component, access point and network level. Green mobile communication is an important chunk of corporate social obligation which

Journal Pre-proof

TABLE XX: Ongoing projects

International projects/ Consortium on green communication	Web Reference	Approach	Participated country/ infrastructure vendors	Research outcome
FANTASTIC	http://fantastic5g.eu/	Develop a framework based on AI components	European countries	Present a multi-service air interface for enhancing reliability and energy efficiency
GREENNETS	http://www.greennets.eu/	For designing energy efficient architecture advanced tools and software platforms introduce in the market	Germany, Poland, Lithuania	Decrease CO_2 footprint, power consumption and exposure to electromagnetic smog.
TRIBE	http://tribe-h2020.eu	To improve social behavior of users towards EE	European Commission	Reduce energy consumption and emissions between 15% and 30%
CogNet	https://5g-ppp.eu/cognet/	Applying machine learning method for network management in 5G	European Commission and European ICT industry	Design an intelligent system of visions and action, supporting green energy-efficient communication
EUFORIE (European Futures for Energy Efficiency)	http://euforie-h2020.eu	Works on multi-scale analysis for enhancing EE	Finland	Achieve secure, clean, green and energy efficient environment
5 Green	https://wireless.kth.se/5green/	Physical layer signaling, traffic adaptive solutions and higher layer protocols used for minimizing energy consumption	Finland, Italy	Achieve eco-friendly energy efficient 5G networks
GreenTouch	http://www.greentouch.org	Design energy efficient architecture and work on network specifications	-	Target to achieve 98% energy reduction till 2020
5GPPP	https://5g-ppp.eu/	Design energy efficient architecture with converging capabilities for both fixed and mobile accesses	Europe	Saving up to 90% of energy per service provided
LEXNET (Low EMF Exposure Future Networks)	http://www.lexnet.fr/	Focus on cross layer techniques to reduce the EM radiation exposure	European countries	Target to achieve reduced EM radiation exposure up to 50% without compromising QoE and QoS
HORIZON 2020	https://ec.europa.eu/programmes /horizon2020/en/what-horizon-2020	Apply security mechanisms on physical layer	-	Focus on privacy and security issues of 5G architecture
mm MAGIC	https://5g-mmmagic.eu/project/	Consider 6-100 GHz range, novel frame structure, cooperative and adaptive beamforming for designing and addressing the challenges of mobile radio access technology	Key European operators (Telefonica, Orange) and Nokia Samsung, Lucent Ericsson, Alcatel-Huawei, Intel)	Cost and energy-efficient solutions for 5G millimeter wave deployments
The International EMF project	https://5g-mmmagic.eu/https://www.who.int/peh-emf/project/en/	Study the static and time varying magnetic and electric fields between 0-300 GHz frequency range and adverse effects of these on health.	WHO member states	To minimize health hazards because of EM radiation exposure
Green Wireless Communication	https://sri-uq.kaust.edu.sa/Pages /greenwireless.aspx	Progress and performance evaluation of novel wireless channel estimation methods, transceiver design optimization for BS sleeping strategy.	Kingdom of Saudi Arabia	To minimize energy consumption of NGNs.
Wireless@MIT Center	http://wireless.csail.mit .edu/Energy	Work on low-power handsets, energy-improving sensors and wireless medical devices.	MIT Computer Science and Artificial Intelligence Laboratory, Cambridge	Analyze software and hardware components for energy-efficient mobile systems.

endeavors to minimize greenhouse gases and carbon footprint in green ICT services for customers. The government should propose protocols, rules, and regulations to certify the service provider as a green service provider. So, mobile operators have to combat global warming by decreasing their CO_2 emissions.

- *Lesson 2*: The joint application of the BS sleep method with hybrid beamforming provides better cost reduction and flexibility. The increased demand for mobile data and the internet needs a paradigm shift towards broader bandwidth accessibility in GCRN and mm-Wave frequency band. Massive MIMO contains a larger number of antennas which minimize transmitted power. The main attention lies on the utilization of relay for the future green networks as well as small size cells and their role in the formation of energy-efficient communication systems.

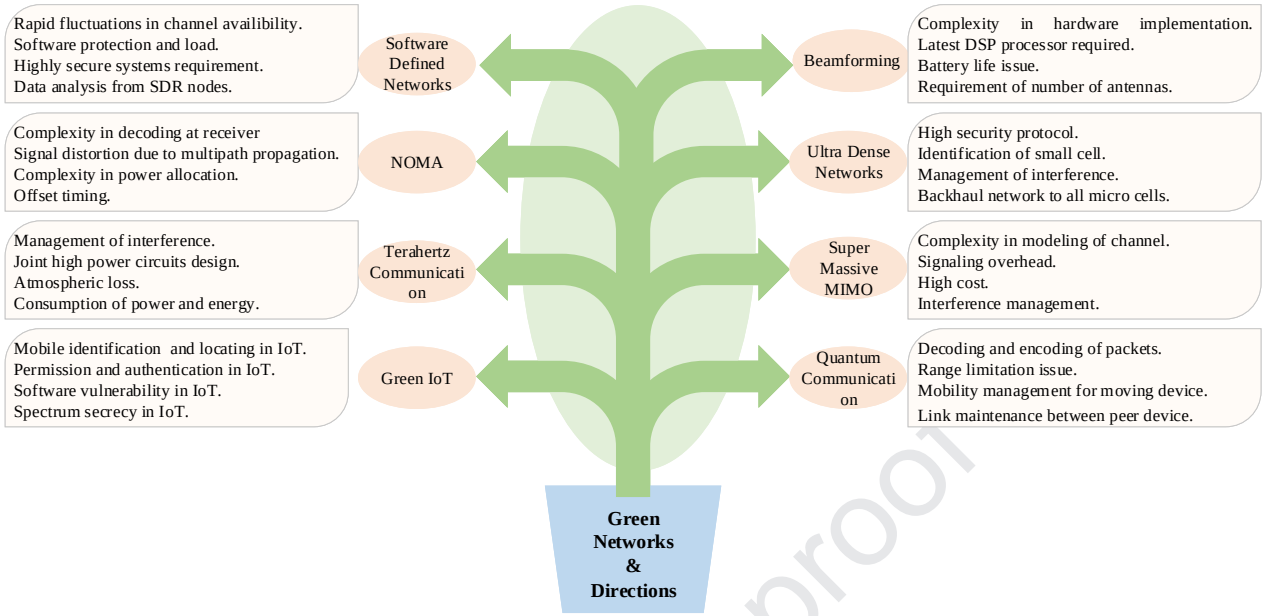


Fig. 34: Green communication future networks and related research directions

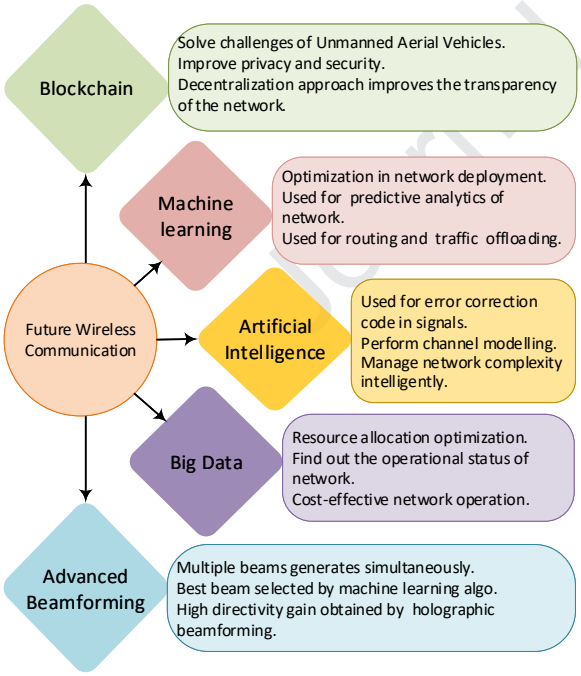


Fig. 35: Advance communication technologies i.e. Next step towards future wireless networks

The green cognitive radio networks have a great potential to handle inevitable demands of extra spectrum in next generation wireless networks. But have some limitation in applying methods in existing communication systems. The most prominent limitations of the current methods are such as practical hardware realization, interference and handoff management, proper traffic load balancing among different candidate networks, non-zero latency in real-time services and applications, appropriate densification limit of a network (how much density of the network is adequate?), lack of self-optimizing and self-configuring network devices, complex circuitry, requirements of system reconfiguration, cost factor and privacy and security management.

- *Lesson 3*: ICT attempts to achieve the target to reduce 13% of the global GHG emissions by 2035 with the help of– i) Development of miscellaneous IoT applications to support efficient energy management ii) Advancement of multimedia applications and mobile broadband and iii) Designing and development of system software by introducing functionality of the network, to analyze the behavior of the network's traffic and improving spectral efficiency gains. To pursue energy-efficient green communication, the impact of the mobile life cycle on the carbon footprint is discussed.

Acknowledgments – The authors would like to thank the Women Scientists Scheme-A under Department of Science and Technology Government of India for its financial support of this work under File No: SR/WOS-A/ET-154/2017. The authors also would like to thank the anonymous reviewers for their feedback.

REFERENCES

[1] (01-11-2018) Internet user population will be increased. [Online]. Available: https://www.cisco.com

[2] (01-02-2019) Growth rate of different technologies. [Online]. Available: https://www.gsma.com/r/mobileeconomy/
[3] (11-11-2018) Growth rate of base station. [Online]. Available: https://www.telecomlead.com
[4] (01-06-2008) gas emmission. [Online]. Available: https://www.theclimategroup.org
[5] I. Humar, X. Ge, L. Xiang, M. Jo, M. Chen, and J. Zhang, "Rethinking energy efficiency models of cellular networks with embodied energy," *IEEE network*, vol. 25, no. 2, pp. 40–49, 2011.
[6] S. Lambert, W. Van Heddeghem, W. Vereecken, B. Lannoo, D. Colle, and M. Pickavet, "Worldwide electricity consumption of communication networks," *Optics express*, vol. 20, no. 26, pp. B513–B524, 2012.
[7] (01-02-2019) Population of mobile subscriber. [Online]. Available: https://www.gsma.com/r/mobileeconomy/asiapacific/
[8] (01-02-2019) Population of mobile subscriber in china. [Online]. Available: https://www.gsma.com/r/mobileeconomy/china/
[9] (01-02-2019) Population of mobile subscriber global. [Online]. Available: https://www.gsma.com
[10] (01-06-2019) addition of mobile users globally. [Online]. Available: https://www.ericsson.com
[11] (11-04-2019) 100s of birds drop from the sky at the hague in netherlands. [Online]. Available: http://www.radiationdangers.com
[12] (03-11-2018) Negetive impact on body organs. [Online]. Available: https://gizadeathstar.com
[13] S. S. Sandhu, A. Rawal, P. Kaur, and N. Gupta, "Major components associated with green networking in information communication technology systems," in *2012 International Conference on Computing, Communication and Applications*. IEEE, 2012, pp. 1–6.
[14] Y. Kremenetskaya, S. Markov, and S. Morozova, "Application of hybrid millimetre wave technology for green wireless communications," in *2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT)*. IEEE, 2018, pp. 649–652.
[15] R. Mahapatra, Y. Nijsure, G. Kaddoum, N. U. Hassan, and C. Yuen, "Energy efficiency tradeoff mechanism towards wireless green communication: A survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 686–705, 2016.
[16] D. Feng, C. Jiang, G. Lim, L. J. Cimini, G. Feng, and G. Y. Li, "A survey of energy-efficient wireless communications," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 167–178, 2013.
[17] P. Gandotra and R. K. Jha, "A survey on green communication and security challenges in 5g wireless communication networks," *Journal of Network and Computer Applications*, vol. 96, pp. 39–61, 2017.
[18] T. Chen, H. Kim, and Y. Yang, "Energy efficiency metrics for green wireless communications," in *2010 International Conference on Wireless Communications & Signal Processing (WCSP)*. IEEE, 2010, pp. 1–6.
[19] L. Suarez, L. Nuaymi, and J.-M. Bonnin, "An overview and classification of research approaches in green wireless networks," *Eurasip journal on wireless communications and networking*, vol. 2012, no. 1, p. 142, 2012.
[20] T. A. Khan, A. Yazdan, and R. W. Heath, "Optimization of power transfer efficiency and energy efficiency for wireless-powered systems with massive mimo," *IEEE Transactions on Wireless Communications*, vol. 17, no. 11, pp. 7159–7172, 2018.
[21] M. R. Mili, L. Musavian, K. A. Hamdi, and F. Marvasti, "How to increase energy efficiency in cognitive radio networks," *IEEE Transactions on Communications*, vol. 64, no. 5, pp. 1829–1843, 2016.
[22] Y. Jiang, Y. Zou, J. Ouyang, and J. Zhu, "Secrecy energy efficiency optimization for artificial noise aided physical-layer security in ofdm-based cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 12, pp. 11 858–11 872, 2018.
[23] J. Ouyang, M. Lin, Y. Zou, W.-P. Zhu, and D. Massicotte, "Secrecy energy efficiency maximization in cognitive radio networks," *IEEE Access*, vol. 5, pp. 2641–2650, 2017.
[24] C. Belady, A. Rawson, J. Pfleuger, and T. Cader, "Green grid data center power efficiency metrics: Pue and dcie," Technical report, Green Grid, Tech. Rep., 2008.
[25] R. C. Zoie, R. D. Mihaela, and S. Alexandru, "An analysis of the power usage effectiveness metric in data centers," in *2017 5th International Symposium on Electrical and Electronics Engineering (ISEEE)*. IEEE, 2017, pp. 1–6.
[26] H. Hamdoun, P. Loskot, T. OâĂŹFarrell, and J. He, "Survey and applications of standardized energy metrics to mobile networks," *Annals of telecommunications-annales des télécommunications*, vol. 67, no. 3-4, pp. 113–123, 2012.
[27] X. Wang, A. V. Vasilakos, M. Chen, Y. Liu, and T. T. Kwon, "A survey of green mobile networks: Opportunities and challenges," *Mobile Networks and Applications*, vol. 17, no. 1, pp. 4–20, 2012.
[28] A. A. Ibrahim, K. P. Kpochi, and E. J. Smith, "Energy consumption assessment of mobile cellular networks."
[29] P. Gandotra, R. K. Jha, and S. Jain, "Green communication in next generation cellular networks: A survey," *IEEE Access*, vol. 5, pp. 11 727–11 758, 2017.
[30] Y. Wu, Y. Chen, J. Tang, D. K. So, Z. Xu, I. Chih-Lin, P. Ferrand, J.-M. Gorce, C.-H. Tang, P.-R. Li *et al.*, "Green transmission technologies for balancing the energy efficiency and spectrum efficiency trade-off," *IEEE Communications Magazine*, vol. 52, no. 11, pp. 112–120, 2014.
[31] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green cellular networks: A survey, some research issues and challenges," *arXiv preprint arXiv:1108.5493*, 2011.
[32] Y. Zhang, Y. Xu, Y. Sun, Q. Wu, and K. Yao, "Energy efficiency of small cell networks: metrics, methods and market," *IEEE Access*, vol. 5, pp. 5965–5971, 2017.
[33] M. Parker and S. Walker, "Roadmapping ict: An absolute energy efficiency metric," *IEEE/OSA Journal of Optical communications and Networking*, vol. 3, no. 8, pp. A49–A58, 2011.
[34] J. Wu, Y. Zhang, M. Zukerman, and E. K.-N. Yung, "Energy-efficient base-stations sleep-mode techniques in green cellular networks: A survey," *IEEE communications surveys & tutorials*, vol. 17, no. 2, pp. 803–826, 2015.
[35] V. Foteinos, D. Kelaidonis, G. Poulios, P. Vlacheas, V. Stavroulaki, and P. Demestichas, "Cognitive management for the internet of things: A framework for enabling autonomous applications," *IEEE vehicular technology magazine*, vol. 8, no. 4, pp. 90–99, 2013.
[36] F. H. Raab, "Efficiency of doherty rf power-amplifier systems," *IEEE transactions on Broadcasting*, no. 3, pp. 77–83, 1987.
[37] H. Claussen, L. T. Ho, and F. Pivit, "Effects of joint macrocell and residential picocell deployment on the network energy efficiency," in *2008 IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*. IEEE, 2008, pp. 1–6.
[38] F. Ghannouchi, M. Ebrahimi, and M. Helaoui, "Inverse class f power amplifier for wimax applications with 74% efficiency at 2.45 ghz," in *2009 IEEE International Conference on Communications Workshops*. IEEE, 2009, pp. 1–5.
[39] B. Berglund, J. Johansson, and T. Lejon, "High efficiency power amplifiers," *Ericsson Review*, vol. 83, no. 3, pp. 92–96, 2006.
[40] M. H. Alsharif, R. Nordin, and M. Ismail, "A review on intelligent base stations cooperation management techniques for greener lte cellular networks," *Journal of Communications*, vol. 9, no. 12, pp. 937–945, 2014.
[41] A. D. Gandhi and M. E. Newbury, "Evaluation of the energy efficiency metrics for wireless networks," *Bell Labs Technical Journal*, vol. 16, no. 1, pp. 207–215, 2011.
[42] D. Rapone, D. Sabella, and M. Fodrini, "Energy efficiency solutions for the mobile network evolution towards 5g: an operator perspective," in *2015 Sustainable Internet and ICT for Sustainability (SustainIT)*. IEEE, 2015, pp. 1–9.
[43] E. Chavarria-Reyes, I. F. Akyildiz, and E. Fadel, "Energy consumption analysis and minimization in multi-layer heterogeneous wireless systems," *IEEE Transactions on Mobile Computing*, vol. 14, no. 12, pp. 2474–2487, 2015.
[44] Y. Lv, H. Zhang, and S. Xueming, "Analysis of base stations deployment on power saving for heterogeneous network," in *2017 IEEE 17th International Conference on Communication Technology (ICCT)*. IEEE, 2017, pp. 1439–1444.
[45] I. Ashraf, L. T. Ho, and H. Claussen, "Improving energy efficiency of femtocell base stations via user activity detection," in *2010 IEEE Wireless Communication and Networking Conference*. IEEE, 2010, pp. 1–5.
[46] F. Cao and Z. Fan, "The tradeoff between energy efficiency and system performance of femtocell deployment," in *2010 7th International Symposium on Wireless Communication Systems*. IEEE, 2010, pp. 315–319.
[47] F. Richter, A. J. Fehske, and G. P. Fettweis, "Energy efficiency aspects of base station deployment strategies for cellular networks," in *2009 IEEE 70th Vehicular Technology Conference Fall*. IEEE, 2009, pp. 1–5.
[48] F. Richter and G. Fettweis, "Cellular mobile network densification utilizing micro base stations," in *2010 IEEE International Conference on Communications*. IEEE, 2010, pp. 1–6.
[49] F. Richter, A. J. Fehske, P. Marsch, and G. P. Fettweis, "Traffic demand and energy efficiency in heterogeneous cellular mobile radio networks,"

Journal Pre-proof

in *2010 IEEE 71st Vehicular Technology Conference*. IEEE, 2010, pp. 1–6.
[50] X. Xu, C. Yuan, W. Chen, X. Tao, and Y. Sun, "Adaptive cell zooming and sleeping for green heterogeneous ultradense networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 2, pp. 1612–1621, 2018.
[51] I. G. Karnila, H. Mahmudah, O. Puspitorini, A. Wijayanti, and N. A. Siswandari, "Analysis of cell zooming technique using adaptive traffic load balancing method on macrocell base station," in *2018 International Electronics Symposium on Engineering Technology and Applications (IES-ETA)*. IEEE, 2018, pp. 157–162.
[52] R. Balasubramaniam, "Cell zooming techniques for power efficient base station operation," Ph.D. dissertation, San Diego State University, 2012.
[53] R. I. Yuanita, O. Puspitorini, and H. Mahmudah, "Analysis technique of cell zooming using lowerto-higher scheme on centralized algorithm towards green cellular network in surabaya," in *2016 International Electronics Symposium (IES)*. IEEE, 2016, pp. 132–137.
[54] Z. Niu, "Tango: Traffic-aware network planning and green operation," *IEEE Wireless Communications*, vol. 18, no. 5, pp. 25–29, 2011.
[55] J. Lorincz, A. Capone, and D. Begusic, "Impact of service rates and base station switching granularity on energy consumption of cellular networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, no. 1, p. 342, 2012.
[56] X. Guo, Z. Niu, S. Zhou, and P. Kumar, "Delay-constrained energy-optimal base station sleeping control," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1073–1085, 2016.
[57] J. Gong, J. S. Thompson, S. Zhou, and Z. Niu, "Base station sleeping and resource allocation in renewable energy powered cellular networks," *IEEE Transactions on Communications*, vol. 62, no. 11, pp. 3801–3813, 2014.
[58] C. Liu, B. Natarajan, and H. Xia, "Small cell base station sleep strategies for energy efficiency," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 3, pp. 1652–1661, 2016.
[59] M. F. Hossain, K. S. Munasinghe, and A. Jamalipour, "A protocooperation-based sleep-wake architecture for next generation green cellular access networks," in *2010 4th International Conference on Signal Processing and Communication Systems*. IEEE, 2010, pp. 1–8.
[60] A. Bousia, A. Antonopoulos, L. Alonso, and C. Verikoukis, ""green" distance-aware base station sleeping algorithm in lte-advanced," in *2012 IEEE International Conference on Communications (ICC)*. IEEE, 2012, pp. 1347–1351.
[61] M. A. Marsan, L. Chiaraviglio, D. Ciullo, and M. Meo, "Switch-off transients in cellular access networks with sleep modes," in *2011 IEEE international conference on communications workshops (ICC)*. IEEE, 2011, pp. 1–6.
[62] M. H. Alsharif, R. Nordin, and M. Ismail, "Classification, recent advances and research challenges in energy efficient cellular networks," *Wireless Personal Communications*, vol. 77, no. 2, pp. 1249–1269, 2014.
[63] T. Chen, Y. Yang, H. Zhang, H. Kim, and K. Horneman, "Network energy saving technologies for green wireless access networks," *IEEE Wireless Communications*, vol. 18, no. 5, pp. 30–38, 2011.
[64] (16-11-2018) Extended cell dtx. [Online]. Available: http://www.3gpp.org
[65] M. H. Alsharif, R. Nordin, and M. Ismail, "Survey of green radio communications networks: techniques and recent advances," *Journal of computer networks and communications*, vol. 2013, 2013.
[66] A. Ghandri, Y. Boujelben, and M. B. Jemaa, "Dynamic mbsfn subframe allocation algorithm for bursty video traffic in lte-advanced network," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2018, pp. 1–6.
[67] K. Kanwal, G. A. Safdar, M. Ur-Rehman, and X. Yang, "Energy management in lte networks," *IEEE Access*, vol. 5, pp. 4264–4284, 2017.
[68] D. Migliorini, G. Stea, M. Caretti, and D. Sabella, "Power-aware allocation of mbsfn subframes using discontinuous cell transmission in lte systems," in *2013 IEEE 78th Vehicular Technology Conference (VTC Fall)*. IEEE, 2013, pp. 1–5.
[69] (16-11-2018) Energy saving techniques for lte. [Online]. Available: http://www.3gpp.org
[70] I. F. Akyildiz, D. M. Gutierrez-Estevez, R. Balakrishnan, and E. Chavarria-Reyes, "Lte-advanced and the evolution to beyond 4g (b4g) systems," *Physical Communication*, vol. 10, pp. 31–60, 2014.
[71] (16-11-2018) Radio regulations, international telecommunication union std. [Online]. Available: http://www.itu.int
[72] S. Rostami and K. Arshad, "A novel spectrum-aggregating technique in cognitive radio networks," in *International conference on Advance Computing and Communication Systems, Decemmber*, 2013.
[73] B. Barakat and K. Arshad, "Energy efficient carrier aggregation for lte-advanced," in *2015 IEEE 8th GCC Conference & Exhibition*. IEEE, 2015, pp. 1–5.
[74] Z. Shen, A. Papasakellariou, J. Montojo, D. Gerstenberger, and F. Xu, "Overview of 3gpp lte-advanced carrier aggregation for 4g wireless communications," *IEEE Communications Magazine*, vol. 50, no. 2, pp. 122–130, 2012.
[75] H. A. H. Hassan, L. Nuaymi, and A. Pelov, "Renewable energy in cellular networks: A survey," in *2013 IEEE online conference on green communications (OnlineGreenComm)*. IEEE, 2013, pp. 1–7.
[76] F. Ahmed, M. Naeem, W. Ejaz, M. Iqbal, A. Anpalagan, and H. Kim, "Renewable energy assisted traffic aware cellular base station energy cooperation," *Energies*, vol. 11, no. 1, p. 99, 2018.
[77] M. Alsharif and J. Kim, "Hybrid off-grid spv/wtg power system for remote cellular base stations towards green and sustainable cellular networks in south korea," *Energies*, vol. 10, no. 1, p. 9, 2017.
[78] Y. Zhang, M. Meo, R. Gerboni, and M. A. Marsan, "Minimum cost solar power systems for lte macro base stations," *Computer Networks*, vol. 112, pp. 12–23, 2017.
[79] M. Alsharif, J. Kim, and J. Kim, "Green and sustainable cellular base stations: An overview and future research directions," *Energies*, vol. 10, no. 5, p. 587, 2017.
[80] D. Niyato, X. Lu, and P. Wang, "Adaptive power management for wireless base stations in a smart grid environment," *IEEE Wireless Communications*, vol. 19, no. 6, pp. 44–51, 2012.
[81] H. A. H. Hassan, A. Pelov, and L. Nuaymi, "Integrating cellular networks, smart grid, and renewable energy: Analysis, architecture, and challenges," *IEEE access*, vol. 3, pp. 2755–2770, 2015.
[82] X. Wang, Y. Zhang, T. Chen, and G. B. Giannakis, "Dynamic energy management for smart-grid-powered coordinated multipoint systems," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1348–1359, 2016.
[83] T.-Y. Wu and T. Chang, "Interference reduction by millimeter wave technology for 5g-based green communications," *IEEE Access*, vol. 4, pp. 10 228–10 234, 2016.
[84] W. Liu, X. Li, and M. Chen, "Energy efficiency of mimo transmissions in wireless sensor networks with diversity and multiplexing gains," in *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 4. IEEE, 2005, pp. iv–897.
[85] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5g cellular: It will work!" *IEEE access*, vol. 1, pp. 335–349, 2013.
[86] Y. Gai, L. Zhang, and X. Shan, "Energy efficiency of cooperative mimo with data aggregation in wireless sensor networks," in *2007 IEEE Wireless Communications and Networking Conference*. IEEE, 2007, pp. 791–796.
[87] M. Di Renzo, "Energy efficiency metrics and performance tradeoffs of green wireless networks," *Green Communications: Principles, Concepts and Practice*, pp. 43–54, 2015.
[88] B. C. Banister, M. Brehler, P. Gaal, M. Kitazoe, and K. Bhattad, "Energy saving mode with maintained number of advertised transmit antennas," Aug. 15 2017, uS Patent 9,736,707.
[89] K. Kanwal and G. A. Safdar, "Reduced early handover for energy saving in lte networks," *IEEE Communications Letters*, vol. 20, no. 1, pp. 153–156, 2016.
[90] T. Chen, H. Zhang, Z. Zhao, and X. Chen, "Towards green wireless access networks," in *2010 5th International ICST Conference on Communications and Networking in China*. IEEE, 2010, pp. 1–6.
[91] W. Ejaz, S. K. Sharma, S. Saadat, M. Naeem, A. Anpalagan, and N. Chughtai, "A comprehensive survey on resource allocation for cran in 5g and beyond networks," *Journal of Network and Computer Applications*, p. 102638, 2020.
[92] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser mimo systems," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1436–1449, 2013.
[93] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user mimo systems: Is massive mimo the answer?" *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3059–3075, 2015.
[94] H. Q. Ngo, L.-N. Tran, T. Q. Duong, M. Matthaiou, and E. G. Larsson, "On the total energy efficiency of cell-free massive mimo," *IEEE*

Journal Pre-proof

Transactions on Green Communications and Networking, vol. 2, no. 1, pp. 25–39, 2018.

[95] N. Lu, Y. Jiang, F. Zheng, and X. You, "Energy efficient power control for the two-tier networks with small cells and massive mimo," in *2016 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*. IEEE, 2016, pp. 127–132.

[96] J. Chen, X. Chen, T. Liu, and L. Lei, "Toward green and secure communications over massive mimo relay networks: Joint source and relay power allocation," *IEEE Access*, vol. 5, pp. 869–880, 2017.

[97] Y. Xin, D. Wang, J. Li, H. Zhu, J. Wang, and X. You, "Area spectral efficiency and area energy efficiency of massive mimo cellular systems," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 3243–3254, 2016.

[98] W. Xu, S. Li, S. Wang, Z. Feng, J. Lin, and A. V. Vasilakos, "Joint parameter selection for massive mimo: An energy-efficient perspective," *IEEE Access*, vol. 4, pp. 3719–3731, 2016.

[99] M. Alonzo, S. Buzzi, and A. Zappone, "Energy-efficient downlink power control in mmwave cell-free and user-centric massive mimo," in *2018 IEEE 5G World Forum (5GWF)*. IEEE, 2018, pp. 493–496.

[100] O. Saatlou, M. O. Ahmad, and M. Swamy, "Optimal power allocation for massive mu-mimo downlink tdd systems," in *2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*. IEEE, 2016, pp. 1–4.

[101] L. Chen, F. R. Yu, H. Ji, B. Rong, X. Li, and V. C. Leung, "Green full-duplex self-backhaul and energy harvesting small cell networks with massive mimo," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3709–3724, 2016.

[102] (30-01-2019) Effect of electromagnetic radiation on human-health. [Online]. Available: https://www.airestech.com

[103] (01-11-2018) Health effects from radio frequency electromagnetic fields. [Online]. Available: http://www.hpa.org.uk

[104] K. Agarwal and Y.-X. Guo, "Interaction of electromagnetic waves with humans in wearable and biomedical implant antennas," in *2015 Asia-Pacific Symposium on Electromagnetic Compatibility (APEMC)*. IEEE, 2015, pp. 154–157.

[105] M. Younis, T. Rommel, F. Bordoni, G. Krieger, and A. Moreira, "On the pulse extension loss in digital beamforming sar," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 7, pp. 1436–1440, 2015.

[106] B. M. Hochwald and D. J. Love, "Minimizing exposure to electromagnetic radiation in portable devices," in *2012 Information Theory and Applications Workshop*. IEEE, 2012, pp. 255–261.

[107] W. Ni, P.-H. Chiang, and S. Dey, "Energy efficient hybrid beamforming in massive mu-mimo systems via eigenmode selection," in *2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*. IEEE, 2017, pp. 400–406.

[108] O. Tervo, L.-N. Tran, H. Pennanen, S. Chatzinotas, M. Juntti, and B. Ottersten, "Energy-efficient coordinated multi-cell multi-group multicast beamforming with antenna selection," in *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2017, pp. 1209–1214.

[109] L. Zhou, Z. Xu, W. Jiang, and W. Luo, "Joint multicast beamforming and user scheduling in large-scale antenna systems," *IET Communications*, vol. 12, no. 11, pp. 1307–1314, 2018.

[110] S. Tariq, D. Psychoudakis, O. Eliezer, and F. Khan, "A new approach to antenna beamforming for millimeter-wave fifth generation (5g) systems," in *2018 Texas Symposium on Wireless and Microwave Circuits and Systems (WMCS)*. IEEE, 2018, pp. 1–5.

[111] M. Scarfi *et al.*, "International and national expert group evaluations: biological/health effects of radiofrequency fields," *International journal of environmental research and public health*, vol. 11, no. 9, pp. 9376–9408, 2014.

[112] R. Gómez-Villanueva, H. Jardón-Aguilar, and R. L. y Miranda, "State of the art methods for low sar antenna implementation," in *Proceedings of the Fourth European Conference on Antennas and Propagation*. IEEE, 2010, pp. 1–4.

[113] L. Ragha and M. Bhatia, "Evaluation of sar reduction for mobile phones using rf shields," *International Journal of Computer Applications*, vol. 1, no. 13, pp. 80–85, 2010.

[114] D. Ying, D. J. Love, and B. M. Hochwald, "Beamformer optimization with a constraint on user electromagnetic radiation exposure," in *2013 47th Annual Conference on Information Sciences and Systems (CISS)*. IEEE, 2013, pp. 1–6.

[115] Y. Yang, Y. Zhang, L. Dai, J. Li, S. Mumtaz, and J. Rodriguez, "Transmission capacity analysis of relay-assisted device-to-device overlay/underlay communication," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 1, pp. 380–389, 2017.

[116] M. Jung, K. Hwang, and S. Choi, "Joint mode selection and power allocation scheme for power-efficient device-to-device (d2d) communication," in *2012 IEEE 75th vehicular technology conference (VTC Spring)*. IEEE, 2012, pp. 1–5.

[117] X. Tao, X. Xiao, and J. Lu, "A qos-aware power optimization scheme in ofdma systems with integrated device-to-device (d2d) communications," *Engine*, 2012.

[118] M. Belleschi, G. Fodor, and A. Abrardo, "Performance analysis of a distributed resource allocation scheme for d2d communications," in *2011 IEEE GLOBECOM Workshops (GC Wkshps)*. IEEE, 2011, pp. 358–362.

[119] P. Mach, Z. Becvar, and T. Vanek, "In-band device-to-device communication in ofdma cellular networks: A survey and challenges," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 1885–1922, 2015.

[120] G. Fodor, E. Dahlman, G. Mildh, S. Parkvall, N. Reider, G. Miklós, and Z. Turányi, "Design aspects of network assisted device-to-device communications," *IEEE Communications Magazine*, vol. 50, no. 3, pp. 170–177, 2012.

[121] P. Li and S. Guo, "Incentive mechanisms for device-to-device communications," *IEEE Network*, vol. 29, no. 4, pp. 75–79, 2015.

[122] S. Mumtaz, H. Lundqvist, K. M. S. Huq, J. Rodriguez, and A. Radwan, "Smart direct-lte communication: An energy saving perspective," *Ad Hoc Networks*, vol. 13, pp. 296–311, 2014.

[123] Z. Zhou, M. Dong, K. Ota, J. Wu, and T. Sato, "Energy efficiency and spectral efficiency tradeoff in device-to-device (d2d) communications," *IEEE Wireless Communications Letters*, vol. 3, no. 5, pp. 485–488, 2014.

[124] Y. Xu and S. Wang, "Mode selection for energy efficient content delivery in cellular networks," *IEEE Communications Letters*, vol. 20, no. 4, pp. 728–731, 2016.

[125] C. Kai, H. Li, L. Xu, Y. Li, and T. Jiang, "Energy-efficient device-to-device communications for green smart cities," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 4, pp. 1542–1551, 2018.

[126] P. Gandotra and R. K. Jha, "Device-to-device communication in cellular networks: A survey," *Journal of Network and Computer Applications*, vol. 71, pp. 99–117, 2016.

[127] A. Ramezani-Kebrya, M. Dong, B. Liang, G. Boudreau, and S. H. Seyedmehdi, "Joint power optimization for device-to-device communication in cellular networks with interference control," *IEEE Transactions on Wireless Communications*, vol. 16, pp. 5131–5146, 2017.

[128] A. Shahid, K. S. Kim, E. De Poorter, and I. Moerman, "Self-organized energy-efficient cross-layer optimization for device to device communication in heterogeneous cellular networks," *IEEE Access*, vol. 5, pp. 1117–1128, 2017.

[129] Y. Jiang, Q. Liu, F. Zheng, X. Gao, and X. You, "Energy-efficient joint resource allocation and power control for d2d communications," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6119–6127, 2016.

[130] A. Habbal, S. I. Goudar, and S. Hassan, "A context-aware radio access technology selection mechanism in 5g mobile network for smart city applications," *Journal of Network and Computer Applications*, vol. 135, pp. 97–107, 2019.

[131] A. M. Hamed and R. K. Rao, "Evaluation of capacity and power efficiency in millimeter-wave bands," in *2016 International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS)*. IEEE, 2016, pp. 1–6.

[132] T. S. Rappaport, Y. Xing, G. R. MacCartney, A. F. Molisch, E. Mellios, and J. Zhang, "Overview of millimeter wave communications for fifth-generation (5g) wireless networks with a focus on propagation models," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 12, pp. 6213–6230, 2017.

[133] S. A. R. Naqvi, S. A. Hassan, H. Pervaiz, Q. Ni, and L. Musavian, "Self-adaptive power control mechanism in d2d enabled hybrid cellular network with mmwave small cells: An optimization approach," in *2016 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2016, pp. 1–6.

[134] Z. Wei, X. Zhu, S. Sun, Y. Huang, A. Al-Tahmeesschi, and Y. Jiang, "Energy-efficiency of millimeter-wave full-duplex relaying systems: Challenges and solutions," *IEEE Access*, vol. 4, pp. 4848–4860, 2016.

[135] M. Hussain and N. Michelusi, "Energy efficient beam-alignment in millimeter wave networks," in *2017 51st Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2017, pp. 1219–1223.

[136] M. Feng, S. Mao, and T. Jiang, "Dynamic base station sleep control and rf chain activation for energy-efficient millimeter-wave cellular

systems," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 10, pp. 9911–9921, 2018.
[137] Z. Chen, X. Ma, B. Zhang, Y. Zhang, Z. Niu, N. Kuang, W. Chen, L. Li, and S. Li, "A survey on terahertz communications," *China Communications*, vol. 16, no. 2, pp. 1–35, 2019.
[138] K. Tekbıyık, A. R. Ekti, G. K. Kurt, and A. Görçin, "Terahertz band communication systems: Challenges, novelties and standardization efforts," *Physical Communication*, vol. 35, p. 100700, 2019.
[139] R. Arshad, S. Zahoor, M. A. Shah, A. Wahid, and H. Yu, "Green iot: An investigation on energy saving practices for 2020 and beyond," *IEEE Access*, vol. 5, pp. 15 667–15 681, 2017.
[140] A. Srivastava, M. S. Gupta, and G. Kaur, "Green smart cities," *Green and Smart Technologies for Smart Cities*, pp. 1–18, 2019.
[141] L. Da Xu, W. He, and S. Li, "Internet of things in industries: A survey," *IEEE Transactions on industrial informatics*, vol. 10, no. 4, pp. 2233–2243, 2014.
[142] M. Maksimovic, "Greening the future: Green internet of things (g-iot) as a key technological enabler of sustainable development," in *Internet of Things and Big Data Analytics Toward Next-Generation Intelligence*. Springer, 2018, pp. 283–313.
[143] A. Jain, M. Mishra, S. K. Peddoju, and N. Jain, "Energy efficient computing-green cloud computing," in *2013 International Conference on Energy Efficient Technologies for Sustainability*. IEEE, 2013, pp. 978–982.
[144] S. Li, L. Da Xu, and X. Wang, "Compressed sensing signal and data acquisition in wireless sensor networks and internet of things," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 4, pp. 2177–2186, 2013.
[145] Y. Cho, M. Kim, and S. Woo, "Energy efficient iot based on wireless sensor networks," in *2018 20th International Conference on Advanced Communication Technology (ICACT)*. IEEE, 2018, pp. 294–299.
[146] R. Sukjaimuk, Q. Nguyen, and T. Sato, "A smart congestion control mechanism for the green iot sensor-enabled information-centric networking," *Sensors*, vol. 18, no. 9, p. 2889, 2018.
[147] M. A. Zamora-Izquierdo, J. Santa, and A. F. Gómez-Skarmeta, "An integral and networked home automation solution for indoor ambient intelligence," *IEEE Pervasive Computing*, vol. 9, no. 4, pp. 66–77, 2010.
[148] Y. H. Lee and I. Marsic, "A sensing platform for a uhf rfid system," *International Journal of Scientific & Technology Research*, vol. 4, no. 8, pp. 114–121, 2015.
[149] M. Razzaque and S. Dobson, "Energy-efficient sensing in wireless sensor networks using compressed sensing," *Sensors*, vol. 14, no. 2, pp. 2822–2859, 2014.
[150] P. Sathyamoorthy and E. Ngai, "Energy efficiency as a orchestration service for the internet of things," in *Proc. 11th Swedish Nat. Comput. Netw. Workshop*, 2015, pp. 41–44.
[151] M. Ghamkhari and H. Mohsenian-Rad, "Optimal integration of renewable energy resources in data centers with behind-the-meter renewable generator," in *2012 IEEE International Conference on Communications (ICC)*. IEEE, 2012, pp. 3340–3344.
[152] D. Bol, J. De Vos, F. Botman, G. Streel, b. Bernard, D. Flandre, and J.-D. Legat, "Green socs for a sustainable internet-of-things," in *2013 IEEE Faible Tension Faible Consommation*. IEEE, 2013, pp. 1–4.
[153] J. Li, Y. Liu, Z. Zhang, J. Ren, and N. Zhao, "Towards green iot networking: Performance optimization of network coding based communication and reliable storage," *IEEE Access*, vol. 5, pp. 8780–8791, 2017.
[154] M. Moreno, B. Úbeda, A. Skarmeta, and M. Zamora, "How can we tackle energy efficiency in iot basedsmart buildings?" *Sensors*, vol. 14, no. 6, pp. 9582–9614, 2014.
[155] V. Cisco, "Cisco visual networking index: Forecast and trends, 2017–2022," *White Paper*, 2018.
[156] E. Fcc, "Docket no 03-222 notice of proposed rule making and order," 2003.
[157] F. Hu, B. Chen, and K. Zhu, "Full spectrum sharing in cognitive radio networks toward 5g: A survey," *IEEE Access*, vol. 6, pp. 15 754–15 776, 2018.
[158] H. Eltom, S. Kandeepan, R. J. Evans, Y. C. Liang, and B. Ristic, "Statistical spectrum occupancy prediction for dynamic spectrum access: a classification," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, p. 29, 2018.
[159] M. Gupta and K. Kumar, "Progression on spectrum sensing for cognitive radio networks: A survey, classification, challenges and future research issues," *Journal of Network and Computer Applications*, 2019.
[160] J. Zhu and Y. Zou, "Cognitive network cooperation for green cellular networks," *IEEE Access*, vol. 4, pp. 849–857, 2016.
[161] S. R. Sabuj and M. Hamamura, "Energy efficiency analysis of cognitive radio network using stochastic geometry," in *2015 IEEE Conference on Standards for Communications and Networking (CSCN)*. IEEE, 2015, pp. 245–251.
[162] P. S. Tripathi and R. Prasad, "Energy efficiency in cognitive radio network," in *Wireless VITAE 2013*. IEEE, 2013, pp. 1–5.
[163] Z. Shi, T. Tan, K. C. Teh, and K. H. Li, "Energy efficient cognitive radio network based on multiband sensing and spectrum sharing," *IET Communications*, vol. 8, no. 9, pp. 1499–1507, 2014.
[164] F. Zhou, N. C. Beaulieu, Z. Li, J. Si, and P. Qi, "Energy-efficient optimal power allocation for fading cognitive radio channels: Ergodic capacity, outage capacity, and minimum-rate capacity," *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2741–2755, 2016.

Journal Pre-proof

Journal Pre-proof

AUTHOR DECLARATION

Akanksha Srivastava, Mani Shekhar Gupta & Gurjit Kaur
Department of E&CE,
Delhi Technological University,
Shahbad Daulatpur, Main Bawana Road,
Delhi, India, PIN- 110042.
April 22, 2020

Subject: Regarding Conflict of Interest for manuscript entitled "Energy Efficient Transmission Trends Towards Future Green Cognitive Radio Networks (5G): Progress, Taxonomy and Open Challenges".

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been financial support for this work that could not have influenced its outcome.
We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

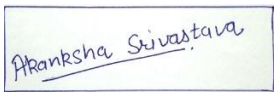
We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). He/she is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs.

We confirm that we have provided a current, correct email address which is accessible by the Corresponding Author and which has been configured to accept email from (akankshasrivastava30@gmail.com).

Journal Pre-proof

Regards



Akanksha Srivastava
(Corresponding author)



Mani Shekhar Gupta
(Co-author)



Dr. Gurjit Kaur
(Co-author)

Journal Pre-proof

Energy Efficient Transmission Towards Future Green Cognitive Radio Networks (5G): Progress, Taxonomy and Open Challenges

Akanksha Srivastava[1], Mani Shekhar Gupta[2], Gurjit Kaur[3]
[1, 2, 3]Department of Electronics and Communication Engineering
[1, 3]Delhi Technological University, Delhi, India
[2]National Institute of Technology, Hamirpur, Himachal Pradesh, India
Email id: [1]akankshasrivastava30@gmail.com, [2]mi_sr87@yahoo.com, [3]gurjitkaur@dtu.ac.in

Authors' detail:



Akanksha Srivastava has received her B. Tech. degree in Electronics and Communication Engineering from Uttar Pradesh Technical University (now Dr. A.P.J. Abdul Kalam Technical University), Lucknow, India in 2010, and her M. Tech. degree in Electronics and Communication Engineering from Dr. A.P.J. Abdul Kalam Technical University, Lucknow, India in 2014. She is pursuing her Ph.D. degree in Electronics and Communication Engineering from Delhi Technological University, Delhi, India. She is also working as a Principal Investigator in a project under Women Scientists Scheme-A (WOS-A) funded by Department of Science and Technology, Government of India (vide File No. SR/WOS-A/ET-154/2017) at the department of electronics and communication engineering, Delhi Technological University, Delhi. From 2014 to 2018, she was an Assistant Professor in the Department of Electronics and Communication Engineering at Pranveer Singh Institute of Technology, Kanpur, India. She has various publications in conferences, journals and book chapters, etc. Her current research includes Green Communication, Energy Efficient Wireless Networks, Cognitive Radio Networks, Wireless Communication, and Heterogeneous Networking.



Mani Shekhar Gupta has received his B. Tech. degree in Electronics and Communication Engineering from Uttar Pradesh Technical University (now Dr. A.P.J. Abdul Kalam Technical University), Lucknow, India in 2009 and his M. Tech. degree in Electronics and Communication Engineering from National Institute of Technology, Hamirpur, India in 2011. He is currently pursuing his Ph.D. degree in Electronics and Communication Engineering from National Institute of Technology, Hamirpur, India. He worked as an Assistant Professor in Department of Electronics and Communication Engineering at Pranveer Singh Institute of Technology, Kanpur, India, from 2011- 2017. He has qualified national level Graduate Aptitude Test in Engineering (GATE) and was a recipient of the GATE Financial Scholarship from 2009 to 2011. Currently, he is a recipient of research fellowship from Ministry of Human Resource Development (MHRD), Government of India. He has various publications in conferences,

journals and book chapters etc. His current research includes Cognitive Radio Networks, Vehicular networks, Spectrum and mobility management, Wireless communication, Next Generation Networks, Resource allocation and Heterogeneous networking.



Gurjit Kaur is an Associate Professor in the Department of Electronics & Communication Engineering at the Delhi Technological University (DTU), Delhi, India and having 16 years of research and teaching experience in well-reputed institutes. She was awarded a 'Gold Medal' by former President of India Dr. A P J Abdul Kalam for being the overall topper of the Punjab Technical University, Jalandhar in B.Tech program. She has completed her M. Tech degree in 2003 and Ph.D. in 2010 from PEC University of Technology, Chandigarh. She has also authored many books at the international and national level. She contributed various book chapters in well-renowned publishers' books i.e., IGI, Springer, CRC, etc. For her outstanding research work she received Prof. Indira Parikh 50 Women in Education Leader Award in World Education Congress, Mumbai, India in 2017. Due to her pioneer research, her name is also listed in Who's Who in Science and Engineering Directory, USA and her biography was also published by International Biographic Centre, Great Britain as 2000 Outstanding Scientists. She is also a reviewer of various reputed journals. Her research interests mainly include Optical CDMA, Wireless Communication system, high-speed interconnect and IoT.

Journal Pre-proof

Proceedings of the Fourth International Conference on Trends in Electronics and Informatics (ICOEI 2020)
IEEE Xplore Part Number: CFP20J32-ART; ISBN: 978-1-7281-5518-0

Ensemble learning-based approach for crack detection using CNN

Vibhu Kailkhura*, S. Aravindh†, Shashank Shekhar Jha‡, N. Jayanthi§
Department of Electronics & Communication Engineering
Delhi Technological University
*vibhu bt2k16@dtu.ac.in, †saravindh bt2k16@dtu.ac.in, ‡jhashashank012@gmail.com, §njayanthi@dce.ac.in

***Abstract*— Crack detection is of pivotal importance in civil engineering and other related applications. Traditional methods of human inspection are tedious and severely limited. Automated crack detection by conventional image processing techniques is challenging due to their inability to discriminate crack features from background noise. Inhomogeneous lighting, shadows, and surface finish hinder the performance of digital image processing methods. The use of convolutional neural networks has helped achieve remarkably better results in the field of computer vision. Ensemble learning is an approach to aggregate the results of a number of individual models for classification or regression. Ensemble learning for crack detection has been implemented using deep convolutional neural networks (DCNN) in this paper. The models are evaluated on a number of performance metrics, namely–(i) accuracy, (ii) precision, (iii) recall (iv) Matthews correlation coefficient (MCC), (v) AUROC, and (vi) F1 score. Experimental results show the robustness of the ensembling method and offer promising scope in crack detection. They outperform the current best performance on open source concrete crack dataset. The ensemble models achieved much better performance than their individual counterparts with the best ensemble achieving a validation accuracy of 99.67%.**

Keywords—crack detection, convolutional neural network (CNN), transfer learning, ensemble learning

I. INTRODUCTION

Concrete structures suffer from decay over time by various means. Cracks develop on the surface, which need to be properly detected. A need for Structural Health Monitoring(SHM) arises as the properties of concrete depends on many factors and are often hard to predict in practice. SHM is a process aimed at providing real-time and accurate information about structural conditions and performance on a proactive basis. SHM is very economical and provides good returns. The cost of installing and implementing an SHM system is only 0.5% to 3% of total construction and repair costs, respectively [1]. Having an SHM system in place can save a lot of money and human lives. Manual detection of cracks cannot set a uniform standard and is subject to an individual's expertise. Automated crack detection overcomes this problem by precise inspection without human error. The application of traditional algorithms of image processing for crack detection led to satisfactory results.

Yamaguchi et al., 2010 [2] used a percolation based image processing method for fast crack detection. A terminate and skip procedure was used where only the crack pixels were checked instead of the background. Circularity of percolated region was then found to set the threshold for classifying pixels. Salman et al., 2013 [3] proposed to use a Gabor filter of a given orientation, which was convolved with the input pre-processed image. After the completion of convolution, the real component of the response out of the kernel was threshold to generate a binary output image. Finally, the binary images resulting from the differently oriented filters were combined by logical OR operation to produce an output image that contained detected crack segments. Yiyang et al., 2014 [4] proposed a glass crack detection algorithm using image processing techniques. The crack information was obtained by making use of image pre-processing, segmentation, and feature extraction of the glass crack image. The pre-processing task involved image smoothing filtering and image sharpening. A twice threshold segmentation was used in [5], and road marking was removed by the Otsu threshold algorithm. Segmentation was done by the improved adaptive iterative threshold segmentation algorithm. Shan et al., 2016 [6] proposed a Canny-Zernike combination algorithm for image processing. SIFT (Scale Invariant Feature Transform) was used for feature extraction. The nearest neighbor algorithm was applied to match the feature points extracted by the SIFT algorithm. Inhomogeneous illumination of the surface and discriminative crack features being confused with background noise are a challenging task for image processing methods.

The image processing algorithms have now been largely outperformed by artificial neural network-based methods. Of the various neural networks in use, the most significant for computer vision is the convolutional neural network (CNN). It offers the advantage of lesser computation in terms of a lesser number of weights to train for a single kernel that can be applied repeatedly to all the portions of the image. It is also able to learn features automatically. The first use of CNNs for crack detection was implemented by Zhang et al., 2016 [7]. They applied CNN on images downsized from a road image of 3264 x 2248 and achieved substantial accuracy. Li et al. [8] used a modified version of GoogLeNet for concrete crack detection. A deep learning-based CNN architecture was proposed to detect cracks in concrete surfaces in [9]. A comparative study was conducted to show the performance difference between CNN and traditional image processing techniques for the purpose of crack detection [10]. The vision-based method using deep CNN performed quite better than the traditional Canny and Sobel edge detection technique. Li et al., 2019 [11] used AlexNet as the CNN architecture for crack detection, and the model was trained for optimum learning parameters.

Ensemble learning is the approach to use multiple models for accurate prediction. The results of the individual models are combined by various methods to achieve better accuracy. The technique of ensembling has been applied in various domains. Shahin at al. [12] discuss the improved accuracy of their approach using ensemble learning for the problem of skin lesion classification. Ensemble models yield promising results when the individual models have significant diversity among them [13].

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 05:00:11 UTC from IEEE Xplore. Restrictions apply.

In this paper, ensemble learning is applied with the motive of crack detection. The paper is structured in smaller sections for better understanding. Section II discusses the concept of Convolutional Neural Networks. Section III is based on ensemble learning. Section IV elaborates the methodology adopted in this paper for crack detection. Section V includes the specifications of the dataset used. Section VI lists the results. Analysis of results is done in section VII. Section VIII presents the conclusion.

II. CONVOLUTIONAL NEURAL NETWORKS

A Convolutional Neural Network (CNN) is a deep learning algorithm that takes an image as an input, learns the various features of that image, and then classifies them into their respective classes. The CNN architecture is similar to the pattern of the neurons in a human brain. It enables machines to perceive the world as humans do by applying relevant filters to capture the spatial and temporal dependencies of an image successfully. The convolution layer and pooling layer are used in feature learning. The Fully Connected layers and the softmax layer are used for the classification task. All the mentioned components are explained in detail as follows:

A. Convolution Layer

It is the core building block of a Convolutional Network and does most of the computational tasks. The input image is essentially an array of pixel values. The kernel or a filter is also an array of numbers (called weights or parameters). The depth of this filter is the same as that of the input. The filter starts convolving around the input by multiplying the pixel values element-wise. These multiplications are then summed up to a single number. This process is then repeated for every unique location. After sliding the filter over all locations, an activation or feature map is obtained. The objective of this operation is to extract high-level features such as edges from the input image.

B. Pooling Layer

The pooling layer reduces the images into a form that is easier to process, without losing critical features. It is responsible for reducing the spatial size of the feature map. Through dimensionality reduction, the computation cost to process the data is reduced. This layer also controls overfitting to some extent. There are two types of pooling available - max pooling and average pooling. Studies have shown that max pooling performs a lot better than average pooling. The convolution layers and pooling layers together form structural units which can be increased to suit the complexity of the image but at the cost of increased computational power.

C. Fully Connected Layer

The Fully Connected layer represents the typical layer of a neural network wherein the activation of a neuron in next layer is determined by its connections with all neurons in the present layer. The FC layers are attached to the end of the network. They take the output of the preceding layers as input and the last FC layer gives an N-dimensional vector output, where N is the number of classes. Each number in this vector represents the confidence of model for a particular class.

D. Softmax Layer

The softmax function limits the output between 0 to 1 and allows it to be directly interpreted as a probability. Since they are interpreted as a probability, they must sum to unity. A softmax layer is typically the final layer in a CNN.

III. ENSEMBLE LEARNING

Stochastic optimization is employed by deep convolutional neural networks (DCNNs) to learn. However, due to their sensitivity to small fluctuations in the training dataset, their performance gets limited as the variance in predictions becomes high. Random noise from training data also gets modeled into the learning process leading to overfitting of data. The objective is to reduce model variance and improve robustness and generalization. Training multiple, diverse models and pooling in their predictions is a practical approach to reduce variance. This is the basis of the ensemble learning process, which yields results that are superior to any individual model [14]. Ensemble learning is known to improve the performance of a model or reduce the probability of an unfortunate selection of a poor one [15]. However, merely using ensemble learning does not guarantee better results. Only smartly chosen ensemble models would provide a substantial improvement in the results.

Most of the ensemble learning algorithms work on different data distribution schemes to individual models and the different combination methods of constituent classifier results. In the proposed work, the individual models are trained by using the whole of the training data.

IV. METHODOLOGY

For the purpose of crack detection, the concept of ensemble learning is applied as it provides improved results. A flow diagram illustrating the steps involved is shown in Fig.1.

Various pre-trained CNNs are used as the base models which vary in architecture and performance. A fastai library [16] is used for deep learning, which is an open-source library built on top of Pytorch [17] and provides powerful learning techniques. The pre-trained CNN models with ImageNet [18] are used. These models are fine-tuned for the specific task of crack detection. The fastai library provides good default values for the hyperparameters involved and the option to train different layers of the network at different learning rates. Hence, the initial layers which recognize basic image features are trained at lowest learning rate and the final layers at the highest learning rate. All the models are trained on the training data and their performance is analyzed on the common validation set.

For building ensemble models, a combination that is expected to show improved results should be selected. The selection criteria has been used to choose the base model combination is based on the following two characteristics-

- The diversity in architecture of the individual base models
- Analyzing the confusion matrix of each model on the crack dataset. The critical parameters of false negatives and false positives were taken into account while deciding the combinations of ensemble models which could lead to improvement.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 05:00:11 UTC from IEEE Xplore. Restrictions apply.

The selected model combinations are then used to build ensemble models. The model combinations are indexed in table III. The ensembling methods used in this paper include the most popular techniques of averaging, voting and weighted voting [15]. Using these methods on all the model combinations, various ensemble models are built. The ensemble models are then validated on the same validation dataset used for individual base models and the performance is analyzed.

Let there be *M* number of models for classifying the sample into *C* classes. Then the various ensembling methods may be formulated as below.

1) Averaging

This method is based on taking the simple algebraic average of the predictions of the base models and then treating the average as the final prediction of the ensemble model. If the prediction of the i^{th} model for the j^{th} class to an image input *X* be given as $y_{i,j}(X)$ then the ensemble prediction for the j^{th} class will be expressed as

$$Y_{e,j}(X) = \frac{1}{M}\sum_{i=1}^{M} y_{i,j}(X) \quad (1)$$

where the subscript *e* denotes the prediction by the new ensemble model.

The final classification label of the ensemble model for image input *X* is generated as

$$\hat{Y}_e(X) = class(\max(Y_{e,1}(X), Y_{e,2}(X), \dots, Y_{e,C}(X))) \quad (2)$$

2) Majority Voting

In this method, the individual model predictions are first used to generate classification labels. For each output label $\hat{y}_{i,j}(X)$ generated by a base model corresponding to full support to j^{th} class, the vote $V_j(X)$ of the j^{th} class increases by 1. The class which gains the highest votes is taken to be the final predicted class by the ensemble model.

The final classification label of the majority voting ensemble model can be represented as

$$\hat{Y}_e(X) = class(\max(V_1(X), V_2(X), \dots, V_C(X))) \quad (3)$$

3) Weighted Voting

In this method, the votes of a model to a certain label are weighted in proportion to the performance of the model. Equivalently, the vote of a model with higher accuracy is given more preference.

The individual model predictions are first used to generate classification labels. For each output label $\hat{y}_{i,j}(X)$ generated by a base model corresponding to full support to j^{th} class, the vote $V_j(X)$ of the j^{th} class increases by w_i.

The symbol w_i denotes the weight factor assigned to the vote of i^{th} model and is given as

$$w_i = \frac{a_i}{\sum_{i=1}^{M} a_i} \quad (4)$$

The term a_i is the validation accuracy of the i^{th} model.

The final classification label of the weighted voting ensemble model can be represented as

$$\hat{Y}_e(X) = class(\max(V_1(X), V_2(X), \dots, V_C(X))) \quad (5)$$

V. DATASET AND TRAINING

For training and validation, concrete crack dataset [11] is used. This dataset [11] for training and validation was built out of 1250 images of bridge towers and anchor chambers of a suspension bridge in Dalian, Liaoning, China. The images were captured using a smartphone at a distance of 0.5-1.0m to the surface and were cropped into 256x256 pixel dimension images. Finally, a dataset of 60,000 images was prepared after manual classification.

The images include cracks in varied lighting conditions and situations for enhanced robustness of model. The dataset specifications are given in table I. Sample figures from the dataset are shown in Fig.2.

TABLE I

SPECIFICATION OF THE DATASET USED

Type	Training	Validation	Total
Positive	24k	6k	30k
Negative	24k	6k	30k
Total	48k	12k	60k

VI. EXPERIMENTS AND RESULTS

The models have been trained using the fastai library with Tesla P100-PCie GPU for acceleration on Google Colab, which is a cloud computing platform for machine learning.

The best ensemble model in our experiments outperforms the previous best results as shown in table II.

Let TP, TN, FP and FN denote the true positives, true negatives, false positives and false negatives for the model. Then the performance metrics may be expressed in these terms. The metrics used are listed below.

A. Accuracy

It is a measure of the number of correct predictions made over all kinds of predictions made.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (6)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 05:00:11 UTC from IEEE Xplore. Restrictions apply.

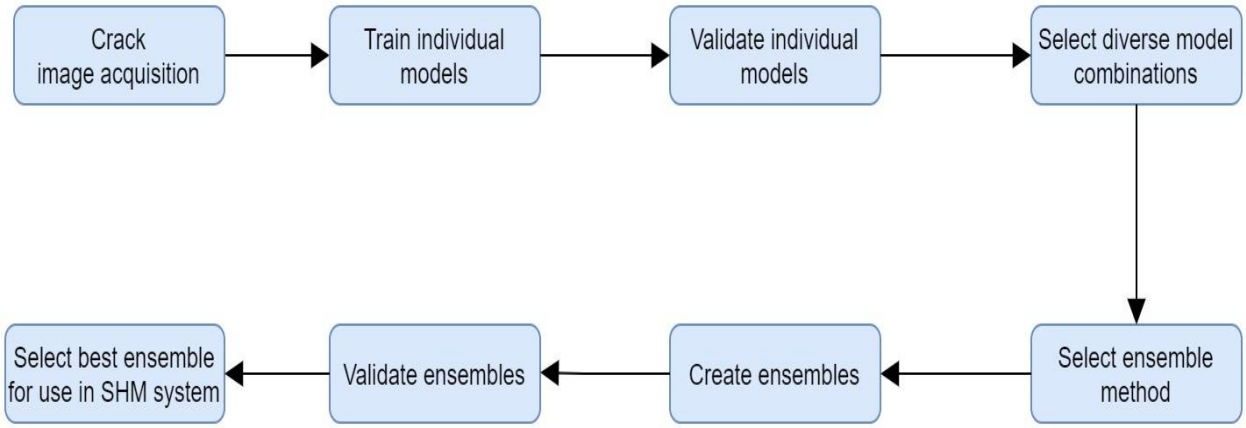


Fig. 1. Flow Diagram for the technique

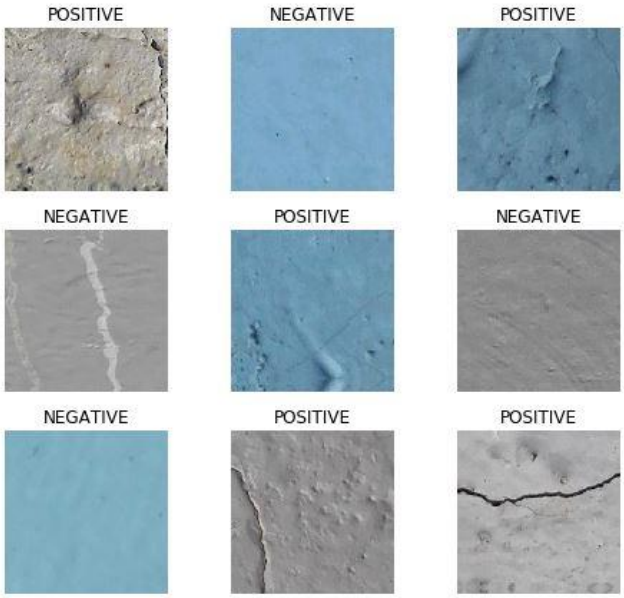


Fig. 2. Sample Images from the dataset

B. Precision

Precision is a measure of what proportion of predicted positives were truly positive.

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

C. Recall

Recall is a measure of what proportion of actual positives were classified correctly.

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

D. Matthews Correlation Coefficient (MCC)

MCC is a correlation between observed and predicted binary classifications.

$$MCC = \frac{(TP.TN) - (FP.FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{9}$$

E. AuROC

AuROC is a measure of how well the probabilities of the positive classes are separated from the negative classes.

F. F1 Score

F1 score is the harmonic mean of recall and precision. It maintains a balance between recall and precision.

$$F1\ Score = 2\,.\frac{Precision.Recall}{Precision + Recall} \tag{10}$$

TABLE II
PERFORMANCE COMPARISON ON DATASET WITH PREVIOUS WORK

Work	**Accuracy**
Li et al. [8]	99.39
Li et al. [11]	99.06
Proposed Ensemble Model	99.67

Results based on the ensembling method are shown for various models in table IV, V, VI and VII.

VII. DISCUSSION

The comparative studies in the form of tables provide a deep insight for analysis. The results are discussed based on the three ensemble methods used. The model combinations are indexed in table III. The models m1-m5 are the individual constituent models. The models m6-m11 are the combinations of two models. Similarly models m12-m14 represent the three-model combinations and m15-m16 are the four-model combinations. Model m17 is the total-ensemble model built using all of the models m1-m5. All ensemble combinations (m6-m17) have been implemented using the three ensembling techniques of averaging, majority voting and weighted voting. Table IV shows the performance of the base models. The results based on the ensembling methods are discussed below.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 05:00:11 UTC from IEEE Xplore. Restrictions apply.

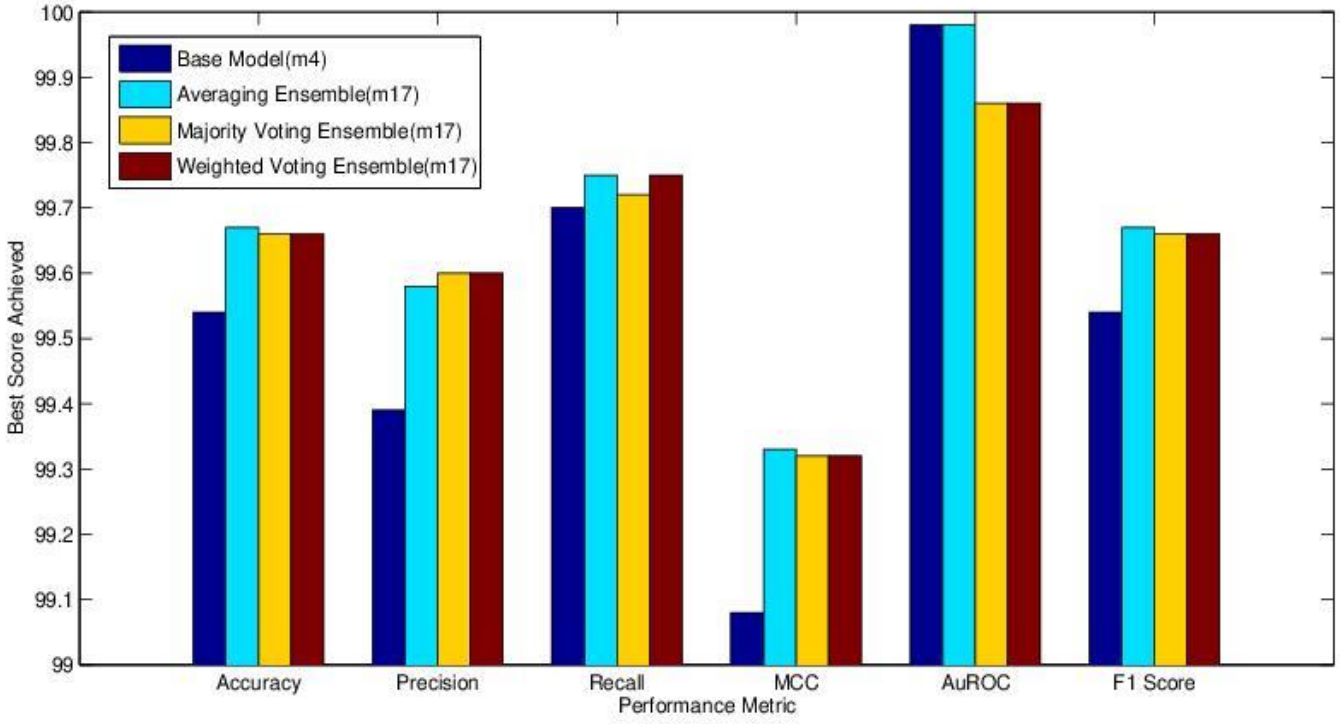


Fig. 3. Performance improvement of ensemble model m17 over the best single model m4

TABLE III

MODEL COMBINATIONS LIST

Model Index	Base Models
m1	ResNet 34
m2	VGG 19
m3	SqueezeNet
m4	DenseNet
m5	AlexNet
m6	DenseNet, SqueezeNet
m7	ResNet, DenseNet
m8	DenseNet, VGG
m9	VGG, ResNet
m10	ResNet, SqueezeNet
m11	SqueezeNet, VGG
m12	ResNet, VGG, DenseNet
m13	DenseNet, VGG, SqueezeNet
m14	SqueezeNet, ResNet, DenseNet
m15	DenseNet, ResNet, VGG, AlexNet
m16	SqueezeNet, AlexNet, VGG, ResNet
m17	Total-ensemble

A. Averaging Based Ensemble Models

The results on the concrete crack dataset can be assessed very well by accuracy since this dataset is balanced in terms of number of images of both classes. Accuracy helps to determine the degree of correct prediction of a model. This includes correct prediction efficiency for both classes-positive and negative. From table IV, it has been noted that, the maximum accuracy for any single model is 99.54 for m4 (DenseNet). However, the top accuracy of 99.67 is achieved using ensembling as shown in table V. Almost all averaging-based ensembles perform better than the best model in the combination.

The top precision of 99.60 is achieved by model m16 which is an improvement over the 99.52 precision value of m1 (ResNet). The improvement in precision value implies that the false alarms in the system are brought down. This helps to reduce the efforts of the visual inspection team which will be given lesser false alarms hence leading to better utilization of resources. Among the single models, the recall value of 99.73 is highest for m2 (VGG19) while ensemble models m8 and m13 have recall of 99.75. Since recall represents the ratio of predicted cracks out of total cracks, a high value of recall is very important. The F1 score presents a combined perspective of precision and recall, and hence higher F1 score is generated when model provides lesser false alarms and is able to detect most of the cracked images correctly. The highest value for F1 score among the single models is that by m4 (DenseNet) with a value of 99.54. This value is significantly improved to 99.67 in the total-ensemble model (m17). The AuROC metric helps to understand the quality of the predictions of the model and how well it can differentiate the two classes. The AuROC value of 99.98 of m4 is maintained in most of the ensemble models. This shows that the models are confident of the predictions they make. The MCC metric may be used when detection of both classes is to be given equal importance. It represents the correlation between the predicted and true classes and is very useful when the dataset is unbalanced. The model m4 provides the best value of 99.08 among the individual models, while the total-ensemble achieves the MCC score of 99.33.

B. Majority Voting Based Ensemble Models

Table VI shows the performance of ensemble models based on majority voting. In this method, the class label which obtains maximum votes from all the models is termed as the true label for the given data sample.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 05:00:11 UTC from IEEE Xplore. Restrictions apply.

TABLE IV

PERFORMANCE OF INDIVIDUAL MODELS

Model	Accuracy	Precision	Recall	MCC	AuROC	F1 score
m1	99.48	**99.52**	99.43	98.95	99.97	99.47
m2	99.42	99.12	**99.73**	98.85	99.97	99.43
m3	99.23	99.15	99.30	98.45	99.95	99.23
m4	**99.54**	99.39	99.70	**99.08**	**99.98**	**99.54**
m5	99.32	99.40	99.23	98.63	99.95	99.32

TABLE V

PERFORMANCE OF ENSEMBLE MODELS BASED ON AVERAGING

Model	Accuracy	Precision	Recall	MCC	AuROC	F1 score
m6	99.58	99.50	99.67	99.17	99.97	99.58
m7	99.57	99.52	99.62	99.13	99.98	99.57
m8	99.54	99.34	**99.75**	99.08	99.98	99.54
m9	99.55	99.48	99.62	99.10	99.98	99.55
m10	99.58	99.55	99.60	99.15	99.97	99.58
m11	99.52	99.34	99.70	99.03	99.97	99.52
m12	99.59	99.48	99.70	99.18	99.98	99.59
m13	99.60	99.45	**99.75**	99.20	99.98	99.60
m14	99.63	99.57	99.70	99.27	99.98	99.63
m15	99.63	99.53	99.73	99.27	99.98	99.63
m16	**99.67**	**99.60**	99.73	**99.33**	99.98	**99.67**
m17	**99.67**	99.58	**99.75**	**99.33**	99.98	**99.67**

TABLE VI

PERFORMANCE OF ENSEMBLE MODELS BASED ON MAJORITY VOTING

Model	Accuracy	Precision	Recall	MCC	AuROC	F1 score
m6	99.58	99.50	99.67	99.17	99.71	99.58
m7	99.57	99.52	99.62	99.13	99.76	99.57
m8	99.54	99.34	**99.75**	99.08	99.75	99.54
m9	99.55	99.48	99.62	99.10	99.73	99.55
m10	99.58	99.55	99.60	99.15	99.75	99.58
m11	99.52	99.34	99.70	99.03	99.73	99.52
m12	99.61	99.48	99.73	99.22	99.81	99.61
m13	99.57	99.39	**99.75**	99.13	99.81	99.57
m14	99.64	99.57	99.72	99.28	99.79	99.64
m15	99.63	99.53	99.73	99.27	**99.86**	99.63
m16	99.65	99.57	99.73	99.30	99.84	99.65
m17	**99.66**	**99.60**	99.72	**99.32**	**99.86**	**99.66**

TABLE VII

PERFORMANCE OF ENSEMBLE MODELS BASED ON WEIGHTED VOTING

Model	Accuracy	Precision	Recall	MCC	AuROC	F1 score
m6	99.54	99.39	99.70	99.08	99.71	99.54
m7	99.54	99.39	99.70	99.08	99.76	99.54
m8	99.54	99.39	99.70	99.08	99.75	99.54
m9	99.48	99.52	99.43	98.95	99.73	99.47
m10	99.48	99.52	99.43	98.95	99.75	99.47
m11	99.42	99.12	99.73	98.85	99.73	99.43
m12	99.61	99.48	99.73	99.22	99.81	99.61
m13	99.57	99.39	**99.75**	99.13	99.81	99.57
m14	99.64	99.57	99.72	99.28	99.79	99.64
m15	99.61	99.48	99.73	99.22	**99.86**	99.61
m16	99.61	99.53	99.68	99.22	99.84	99.61
m17	**99.66**	**99.60**	99.72	**99.32**	**99.86**	**99.66**

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 05:00:11 UTC from IEEE Xplore. Restrictions apply.

TABLE VIII

TABULAR COMPARISON FOR PERFORMANCE IMPROVEMENT IN METRICS

Model Type	Accuracy	Precision	Recall	MCC	AuROC	F1 score
Single Model	99.54(m4)	99.52(m1)	99.73(m2)	99.08(m4)	99.98(m4)	99.54(m4)
Averaging Ensemble	99.67(m16,m17)	99.60(m16)	99.75(m8,m13,m17)	99.33(m16,m17)	99.98(most models)	99.67(m16,m17)
Voting Ensemble	99.66(m17)	99.60(m17)	99.75(m8,m13)	99.32(m17)	99.86(m15,m17)	99.66(m17)
Weighted Voting Ensemble	99.66(m17)	99.60(m17)	99.75(m13)	99.32(m17)	99.86(m15,m17)	99.66(m7)

It can be observed that for even number of models in an ensemble, there is a chance of a tie for the number of votes that each label gets. Hence, the confidence of prediction to the given label is considered to resolve such a tie. No such corrective measure is required for voting ensembles with an odd number of models. As can be seen, almost every ensemble model outperforms the individual best model in the combination.

As can be seen from table VI, the maximum performance is again achieved by the total-ensemble model (m17) on all the parameters except the recall metric. The individual maximum recall of 99.73 for model m2 is outperformed by voting based m8 and m13 models with 99.75 recall score. The recall value for m17 is 99.72 which is almost the same as that of m2. The AuROC value of the best voting ensemble at 99.86 (for m15 and m17) is lower than the individual best of 99.98(for m4). However, the value is still very high and can be termed as a confident distinction ability between the two classes.

The voting ensembles improve the accuracy of 99.54 (for m4) to 99.66 (for m17), the precision of 99.52 (for m1) to 99.60 (for m17), the recall value of 99.73 (for m2) to 99.75 (for m8 and m13), f1 score of 99.54 (for m4) to 99.66 (for m17) and the MCC value of 99.08 (for m4) to 99.32 (for m17).

C. Weighted Voting Based Ensemble Models

The weighted voting method provides higher vote value to model with better accuracy. The results are tabulated in table VII. As expected, the weighted voting results of ensembles with 2 constituent models (m6-m11) are the same as that of the better performing model among the two in the ensemble. For three or more model combinations(m12-m17), the weighted voting ensemble outperforms the best model in the combination.

The weighted voting ensembles improve the accuracy of 99.54 (for m4) to 99.66 (for m17), the precision of 99.52 (for m1) to 99.60 (for m17), the recall value of 99.73 (for m2) to 99.75 (for m13), f1 score of 99.54 (for m4) to 99.66 (for m17) and the MCC value of 99.08 (for m4) to 99.32 (for m17). It is observed that the best AuROC value of 99.86 (for m15 and m17) is lower than the individual best of 99.98 (for m4). However, the value is still very high and demonstrates a good confidence in the predictions.

Hence, it can be concluded that the ensemble models markedly improve the performance than the individual models and the total-ensemble model (m17) shows best results for most of the performance metrics.

D. Comparative analysis of ensemble methods for crack detection

Since the concrete crack dataset is a balanced one, the accuracy provides a good direct estimate of performance. The performance improvement in the metrics by various ensemble methods is compared in table VIII. Among the constituent models, DenseNet (m4) is the best performing, while the total-ensemble model (m17) is the best among the ensembles. The accuracy improvement is from 99.54 to 99.67. The recall improvement is from 99.73 to 99.75 in the m13 model based on all the three methods of ensemble. The precision value is improved by all the three ensemble methods from 99.52(for m1) to 99.60 (for m17). The f1 score is again improved best by averaging method from 99.54 (for m4) to 99.67 (for m16, m17). The MCC metric which gives equal focus to both class labels without any bias, is improved from 99.08 (for m4) to 99.33 (for m16, m17) in the averaging ensemble. The AuROC value is already high and maintained in averaging ensembles which denotes the quality of the classifying decision.

It is observed that the best performing single base model on the given dataset is DenseNet (m4). The total-ensemble model(m17) achieves improvement in almost all the metrics. The performance improvement of the total-ensemble model (m17) over the best single model m4 is shown in Fig.3.

It is important to note that roughly all the parameters are boosted the most by the averaging method of ensemble. This is intuitively expected as the algebraic aggregation of predictions is done and then converted to class labels. It is observed that m16 and m17 based on averaging perform the same on most of the metrics except precision and recall. However, m17 outperforms m16 in other ensembling methods. This shows the strength of the averaging method which helps to achieve improved results. Owing to the simplicity and the algebraic nature of combination, for most of the cases, a simple averaging method may be used to build an optimal ensemble model with a number of diversified models.

VIII. CONCLUSION

In this paper, ensemble learning has been proposed for the problem of concrete crack detection. The individual CNN models are trained and then various ensemble models based on ensemble methods are built. The ensemble models have shown to boost the performance of the individual models. The best ensemble model gives benchmark results with an accuracy of 99.67% on 12k images used for validation. The results offer promising scope of application of more ensemble-based algorithms for crack detection.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 05:00:11 UTC from IEEE Xplore. Restrictions apply.

IX. REFERENCES

[1] P. Roshan, A. Kumar, D. Tewatia, and S. Pal, "Review paper on structural health monitoring: Its benefit and scope in india," Journal of Civil Engineering and Environmental Technology, vol. 2, no. 2, pp. 109–112, 2015.

[2] T. Yamaguchi and S. Hashimoto, "Fast crack detection method for large-size concrete surface images using percolation-based image processing," Machine Vision and Applications, vol. 21, no. 5, pp. 797–809, 2010.

[3] M. Salman, S. Mathavan, K. Kamal, and M. Rahman, "Pave-ment crack detection using the gabor filter," in 16th interna-tional IEEE conference on intelligent transportation systems (ITSC 2013). IEEE, 2013, pp. 2039–2044.

[4] Z. Yiyang, "The design of glass crack detection system based on image preprocessing technology," in 2014 IEEE 7th Joint International Information Technology and Artificial Intelligence Conference. IEEE, 2014, pp. 39–42.

[5] L. Peng, W. Chao, L. Shuangmiao, and F. Baocai, "Research on crack detection method of airport runway based on twice-threshold segmentation," in 2015 Fifth International Conference on Instrumentation and Measurement, Computer, Communica-tion and Control (IMCCC). IEEE, 2015, pp. 1716–1720.

[6] B. Shan, S. Zheng, and J. Ou, "A stereovision-based crack width detection approach for concrete surface assessment," KSCE Journal of Civil Engineering, vol. 20, no. 2, pp. 803–812, 2016.

[7] L. Zhang, F. Yang, Y. D. Zhang, and Y. J. Zhu, "Road crack detection using deep convolutional neural network," in 2016 IEEE international conference on image processing (ICIP). IEEE, 2016, pp. 3708–3712.

[8] S. Li and X. Zhao, "Convolutional neural networks-based crack detection for real concrete surface," in Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2018, vol. 10598. International Society for Optics and Photonics, 2018, p. 105983V.

[9] G. Yao, F.-J. Wei, J.-Y. Qian, and Z.-G. Wu, "Crack detection of concrete surface based on ƒnewline convolutional neural net-works," in 2018 International Conference on Machine Learning and Cybernetics (ICMLC), vol. 1. IEEE, 2018, pp. 246–250.

[10] Y.-J. Cha, W. Choi, and O. Buy¨uk¨ozt¨urk,¨ "Deep learning-based crack damage detection using convolutional neural networks," Computer-Aided Civil and Infrastructure Engineering, vol. 32, no. 5, pp. 361–378, 2017.

[11] S. Li and X. Zhao, "Image-based concrete crack detection using convolutional neural network and exhaustive search technique," Advances in Civil Engineering, vol. 2019, 2019.

[12] A. H. Shahin, A. Kamal, and M. A. Elattar, "Deep ensemble learning for skin lesion classification from dermoscopic im-ages," in 2018 9th Cairo International Biomedical Engineering Conference (CIBEC). IEEE, 2018, pp. 150–153.

[13] D. Opitz and R. Maclin, "Popular ensemble methods: An em-pirical study," Journal of artificial intelligence research, vol. 11, pp. 169–198, 1999.

[14] T. G. Dietterich, "Ensemble methods in machine learning," in International workshop on multiple classifier systems. Springer, 2000, pp. 1–15.

[15] R. Polikar, "Ensemble learning. scholarpedia 4 (1): 2776," 2008.

[16] J. Howard and S. Gugger, "fastai: A layered api for deep learning," 2020.

[17] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf,¨ E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chil-amkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," 2019.

[18] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009, pp. 248–255.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 05:00:11 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Environment Monitoring System for Agricultural Application using IoT and Predicting Crop Yield using Various Data Mining Techniques

Gautam Gupta
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
gg0918914@gmail.com

Rashi Setia
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
rashika.setia8c@gmail.com

Archana Meena
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
pooranlal1973@gmail.com

Bhavnesh Jaint
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
bhavneshjaint@dtu.ac.in

Corresponding Author: Gautam Gupta, gg0918914@gmail.com

***Abstract*— Nowadays, Internet is facing an exponential increase in the number of electronic devices interfaced through it. Previously, there is a limit of connecting only mobile and PCs (personal computer) with the internet but now millions of devices can be connected using the concept of IoT (internet of things). Hence, IoT is able to communicate data between machine to machine and the data that is previously available to private server now easily available to the internet so that it can be accessed efficiently. This paper show the increase in the usage of IoT in monitoring like applications. Smart farming is the technique which intend to provide all the necessary resources for the specified amount of time. These resources required are light intensity, ambient temperature required, relative humidity, soil moisture content, pH reading of the soil. The central idea is to sense all these parameters one by one and take the final decision accordingly. A sensor node should be developed for sensing all the required resources parameters and subsequently send the data to the cloud for further processing. After getting all the required resources values our final aim is to predict the crop production using various data mining techniques. As the human lives are primarily depended on the food resources, the agricultural process need to be efficient and this efficiency can be enhanced if there is an accurate amount of resources. By getting the accurate amount of resources, Data mining techniques such as Random forest, KNN, SVM are used to analyze the crop production in advance such that farmer always have an upper hand on it and by comparing with the previous trend he will be able to detect which kind of parameter is accurate and which is not. A User defined site is designed to monitor all these values of agricultural process.**

***Keywords*- Internet of Things, Wi-Fi, ESP8266 Wi-Fi module, sensor node, Data Mining, Prediction, random Forest, Support Vector Machine (SVM), K-Nearest Neighbor (KNN).**

I. INTRODUCTION

Agriculture is the foremost only field industry where the recent improvement in technology has not been welcomed in huge number. One of the main reason regarding this issue is poor lethargic condition of the farmers in developing countries say India. The scarcity of agricultural products is also increasing day by day due to two main reasons, first one is overpopulation and the second one is urbanization. Considering the case of overpopulation, it increases the demand of agricultural resources and with the case of increasing population it emphasis the better growth of farming products but due to globalization the most of agricultural land is taken by big industries and they convert these land into non-agricultural land. Basically the rural area for farming is decreasing day by day and with that decrease the agricultural resources are also decreasing. Therefore, the need of an hour is to take a look at the situation carefully and emphasis is given to improve the crop production by consuming the resources conservatively without wasting the existing resources. Smart farming is one such methodology to do it. In this type of farming the crop is fed up with the sufficient quantity of resources required for it to grow and by taking the calculated amount of time. In normal Traditional farming method, there is a time based irrigation system i.e. farmer know exactly that after certain limited time (few days) he needs to irrigate his crop. But due to the lack of knowledge farmer is not capable of irrigating his/her field at the exact duration of time which significantly lead to a wastage of water. Hence to overcome such issue smart farming method is employed. To overcome the issue of traditional farming, a soil moisture detection sensor is implanted at the observation site to get the soil moisture content and with this the crop is irrigated accordingly. Similarly, soil moisture sensor, temperature sensor, Humidity sensor and pH sensor are used around the plant to get the desired value, which helps in reducing the exploitation of extra resources used initially in traditional farming. Soil moisture is important in determining the physical

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:57:29 UTC from IEEE Xplore. Restrictions apply.

structure of the plant whereas all the other sensors play a major role in photosynthesis. This paper basically suggest an idea to generate the sensor node which is able to send the data to the central node i.e. to the thing speak cloud. This paper consist of methodology used in smart farming and after visualizing the exact percentage of resources required for crop to appropriately predict the crop yield using various data mining techniques. Data Mining [1] is a procedure to retrieve useful information from very enormous datasets. This technique is used to predict the crop production in advance so that the farmers could take an immediate measures accordingly as the crop demands. It also helps government to get the value of exact MSP (minimum selling price) for farmers. It aims to predict the crop prediction by using various regression and classification methods like multiple linear regression, SVM, Random forest, KNN etc. With these techniques of data mining, the effect of various parameters (sensor node parameters in general) on the prediction of crop production has been studied.

II. LITERATURE SURVEY

In paper [2] author tries to explains the requirement of WSN in agricultural field so as to increase the crop productivity. This paper [2] also give emphasis on the need of smart farming techniques in today's scenario which ultimately analyze the architecture of proposed system by monitoring the environment parameters. The paper [3] suggest the usage of IoT in the field of agriculture. It basically shows how the intermediate chains of agriculture are connected and how IoT can be useful in doing it. It also tell us about all the technologies that are designed by using the concept of IoT. In paper [4] author explained the system using the feature of GPS to monitor the agricultural field. The prime aim of GPS is to look at the data sensed at sensor node and transfer this data using RS232 that are obtained from sensor node to the central node. The proposed system in paper [5] shows the data of light intensity factor around the crop and saving it in a excel sheet. The data stored in database is further used to compare and analyze it later on. In this system only the light intensity parameter is targeted as their only parameter to control the crop production. The author in paper [6] tried to sort out the problem of unequal rain distribution in crop production by controlling the parameters like soil moisture sensor and pH sensor. Arduino is used as a controlling unit which is further also used to communicate the values to the cloud part. The paper [7] shows the problem of food scarcity in Africa. It significantly put forward a schematic way which would predict total crop production needed to meet the African native citizens' needs. The model of Artificial Neural Network is built and subsequently the prediction is done. It helps in predicting the annual amount of crops needed by 2021. The paper [8] defined the comparison algorithms which one can use to make a prediction of kharif crop in suburban zones of India. The algorithms is being executed on an open source platform named WEKA. Afterwards the various type of error is calculated of all the algorithms and being compared individually. In paper [9] author proposed smart agriculture based on IoT and cloud computing methods therefore the agriculture information cloud is constructed with different dynamic resources so as to achieve load balancing. Large amount of data obtained through RFID, and wireless communication are supported through the agriculture information cloud schematic model. In paper [10] the limitation of IoT is discussed. Actually there is communication delay between the cloud server and the end user application. So, paper proposes the model of AI-enabled edge computing to work for IoT applications. It improves the speed and range for the IoT applications. In paper [12] author tries to convey the various data mining techniques one can use while dealing with concept of IoT techniques and provide an overview of what is data mining and different techniques one must use while dealing with it having maximum efficiency.

III. PROPOSED WORK

Our proposed work is divided mainly into two broad categories. First is getting the sensor data from senor node used at the data transmission end and uploading each senor value to thingspeak so as to monitor it for further evaluation. Secondly, after getting sensor data values utilizing it for analyzing the crop yield production will have an upper hand on the future prediction of crop production.

A. Working with sensor node and thingspeak

The main central idea of the work is explained in Fig.1 given below which contains various components namely Relative humidity & ambient air temperature sensor, Soil Moisture content sensor, ph sensor, CO2 sensor, Microcontroller Unit (MCU) along with Wi-Fi module, Wi-Fi router, and thingspeak cloud. Finally the user defined website to monitor the details. Moisture content of the soil is measured with the use of electrode fabricated in it. The main responsible factor through which the soil moisture content changes is the resistance between the two terminals of the electrode. Whenever the soil moisture changes the resistance between the terminal changes. The change in the value of resistance and is the amount of soil moisture content percentage in the soil. A chip is also fabricated inside it which is used to convert the change in resistance to some analog voltage. It is fed with 3.3v input to give an output voltage in the range 0-3.3v. Output voltage signal is fed to MCU unit respectively. Soil moisture sensor have working voltage less than 5V DC and current limit is < 20mA. DHT11 is multivariable sensor which shows the reading of the relative humidity content in percentage and temperature of the atmosphere. It is capable of measuring both the humidity and temperatures single handedly. It senses the relative humidity using capacitors and temperature using the temperature dependent resistors through which change in the output voltage of the resistance gives the relative change in temperature from the previous value. The output waveform of this sensor is digital in nature hence noise must be removed

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:57:29 UTC from IEEE Xplore. Restrictions apply.

subsequently to protect the actual output wave from getting distorted. The digital nature of output pulse force the connection of the sensor to one of the input pins of MCU which is digital in nature. As the resulting output wave is of digital in nature, hence the data is updated after every 2 seconds timeframe. Humidity range of this sensor is 20-80% with 5% accuracy and temperature range of DHT11 is 0-50 degree Celsius with 2 degree accuracy. The sampling rate of DHT11 is 1 Hz and current limit of sensor is 2.5mA. MQ-2 sensor is used to sense the amount of CO2 in surrounding atmosphere. It is metal oxide semiconductor type gas sensor also known as chemiresistors as the sensing is based upon the change in the value of resistance as the gas comes directly in contact with the sensor material. Using the voltage divider rule, the concentration of gas can be found later on in voltage terms. It can detect gas concentration at 200pm-1000pm. It is fed with 5V dc input.

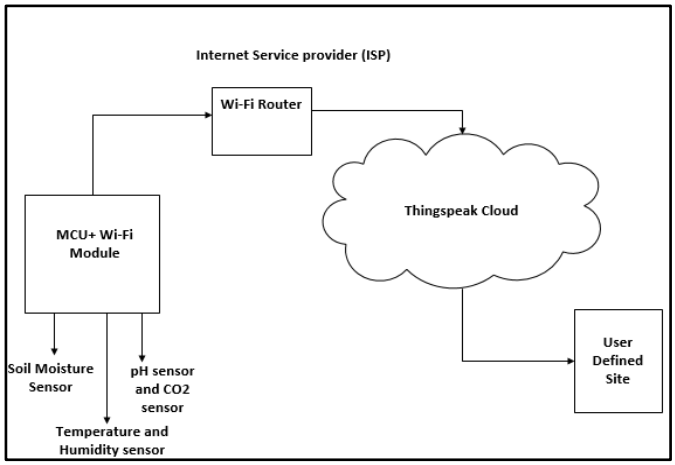


Fig 1. Block Diagram of proposed System

NodeMCU is microcontroller unit here. It is a board which functions on ESP8266 Wi-Fi module. Apart from General pins input-output it also come with additional feature i.e. Pulse Width Modulation capability and analog-digital converter (ADC) all these components are embedded in a single piece of board. NodeMCU basically collects the data from all the sensors that are mentioned above at its GPIOs pins. Data from humidity and temperature sensor is given at GPIO 0 while analog data from the soil moisture sensor is given at the analog input pin of the NodeMCU which having 10 bit of inbuilt ADC. As 10 bit gives a maximum count of 1023 units so it basically mapped the 0-3.3v value to 0-1023 count. It is based on Tensilica 32-bit RSIC microcontroller. The clock speed of the NodeMCU is 80MHz. The SRAM capacity of the device is 80KB. To code the MCU unit Arduino IDE software is used. To connect the NodeMCU board with Wi-Fi, hotspot of the cell phone could also be used which worked as router. Id and password are entered corresponding to the hotspot so that it can connected to the active internet connection.

Thingspeak is used as a cloud source in the above mentioned system. It is basically a platform service which allows you to visualize and analyze live data in the cloud. Certain number of channels are created in this platform where it analyzes the raw data obtained from sensors. Initially, 5 fields are created to analyze the data. The field names are soil moisture content, surrounding temperature, relative humidity content, CO2 and ph. correspondingly, API key is required to code the board that is used in our system. But with the help of thingspeak they are generate automatically. All the data are send in the form analog values so that the graphical representation of the sensor data can be plotted effectively.

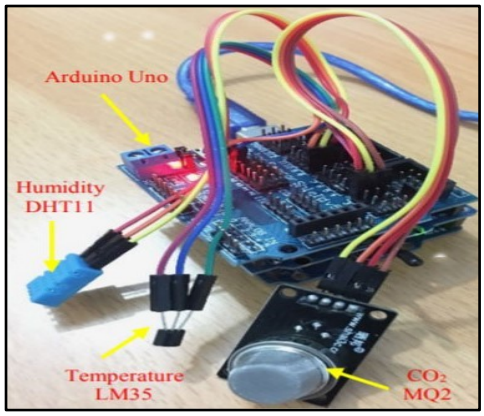


Fig 2. Interfacing of sensors

B. Implementation of Data mining

After sensing the values of all the sensors used above, the measurements of each parameters are stored in the form of datasets. After storing the value, various data mining techniques are used to predict the exact crop production in future. To predict the Crop production three major steps have to be done. These are datasets details, data preprocessing and building prediction model.

1. Datasets used

All the datasets were collected from the previously sensed value from the sensors used at observation site. Temperature, humidity, soil moisture, CO2 percentage and pH value are the common datasets used in the analysis. In addition to these values monthly rainfall dataset is taken from the publicly available records issued by Indian government. Area needed for crop production is also available at the government records. In addition to all these values yield of the crop is also taken under consideration so as to compare it with predicted value with per year yield (in tonnes/hectare).

2. Datasets preprocessing

Datasets preprocessing was done to collect all the datasets of the sensors and additional parameters in Microsoft excel. For every crop it consists of various columns namely: temperature, pH, Soil moisture, relative humidity, CO2 percentage, monthly

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:57:29 UTC from IEEE Xplore. Restrictions apply.

rainfall. By dividing the value of production by area under production, a value is obtained as somehow close to yield. Therefore, by removing the production column and adding yield column or vice-versa and finally saving the file in .csv extension.

3. *Data mining techniques*

Regression is the process performing a statistical modelling through which one can estimate the dependence of dependent variable over one or more than one independent variable. Predictant is the name given to dependent variable and predictors to independent variable. Considering production as dependent (target) variable and all others parameters whether it is area, temperature, soil moisture etc. as independent variables. Now, regression is somewhat similar to classification techniques. Algorithms used in this research paper are Random forest classification, KNN (K-Nearest neighbor) classification and SVM (Support Vector machine) classification.

4. *Constructing the prediction model using sci-kit learn libraries*

This paragraph discuss the building of prediction model of various data mining techniques namely multiple linear regression, K-nearest neighbor (KNN), Support vector machine (SVM) and random forest regression as discussed above using sci-kit learn and py-earth libraries. Sci-kit learn is an open source platform tool for data mining and data analysis via interfacing with python [10]. It provides classification, clustering and regression algorithms. Fig 3. Shows how to construct a prediction model for various data mining techniques as discussed above using sci-kit learn.

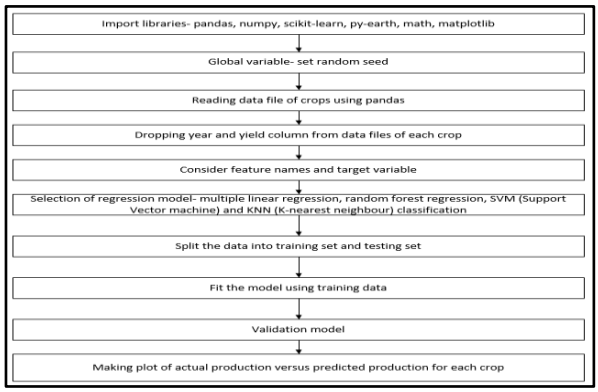


Fig 3. Procedure to construct the prediction model

In prediction model first of all the libraries are imported and then the global variable is created by random seeding method. As our dataset contain year and yield column but it also has production column which is Yield/Area. Hence, both these column are dropped. Now the code of above three classification is added and the data is divided inti training and testing data. Now training data is considered for evaluation and then respective algorithm is added to the data values and plot of actual vs predicted plot is created at last.

IV. RESULTS AND OBSERVATIONS

In the above mentioned work the output of the above methodology depicted in part A of the proposed work is depicted as the variation of values vs observation time and is monitored in thingspeak cloud server. The matlab representation of data sensed at the agricultural field is shown in the form of fields in channel. In fig 4, Temperature is plotted at thingspeak server, similarly in subsequent figures from 5-8 all remaining sensors plot have been obtained.

The plot of ambient temperature of surrounding and time interval is plotted and the values of temperature is recorded in Fahrenheit.

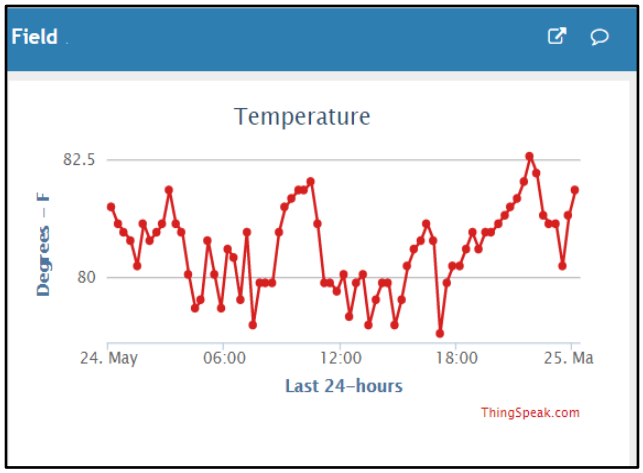


Fig 4. Surrounding air temperature

The plot of relative humidity and time interval is plotted and it is observed that maximum value recorded is 40 percent in humidity.

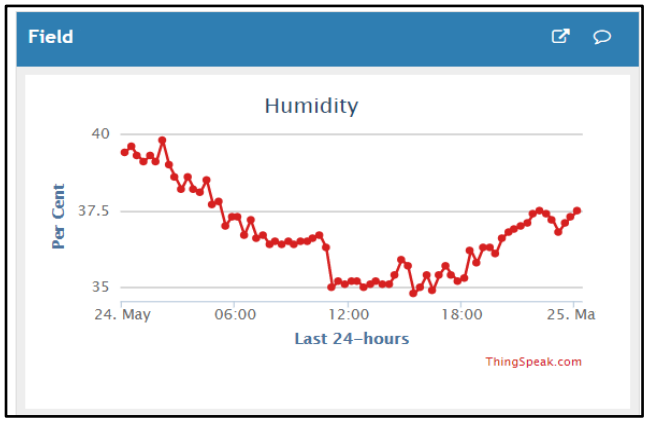


Fig 5. Surrounding air relative humidity

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:57:29 UTC from IEEE Xplore. Restrictions apply.

The plot of soil moisture content vs time interval is plotted and it is noted that soil moisture content is also in percentage.

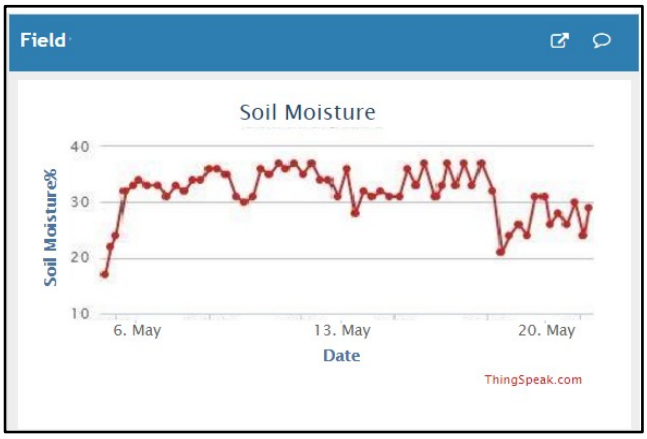


Fig 6. Soil moisture reading

The plot of CO2 vs. time interval is plotted and it is observed that maximum concentration of gas reaches at midway between 3-4 pm.

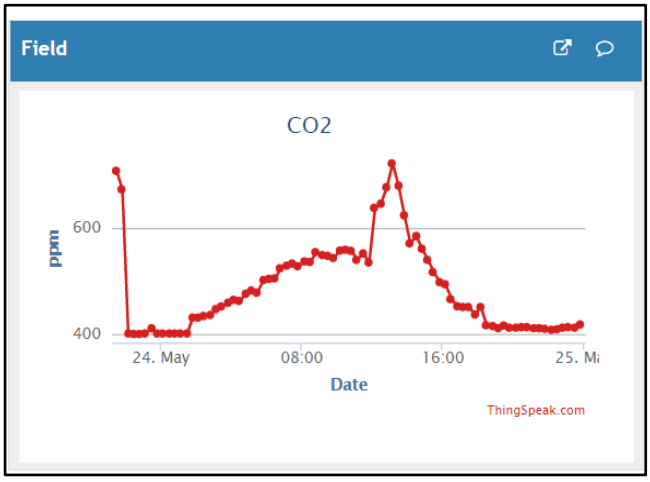


Fig 7. CO2 content in surrounding air

The plot of pH values vs. time interval is plotted and it is concluded that soil is alkaline in nature coz majority of time the value crosses 7.

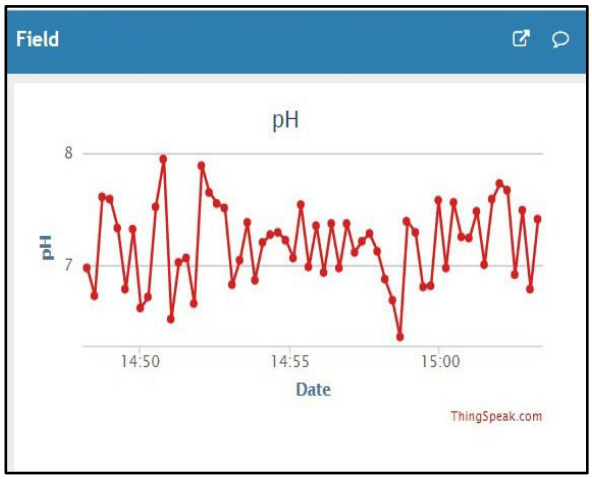


Fig 8. pH of soil

The result of the proposed work done in part-B i.e. visualization of predicted crop production using various data mining techniques mentioned above is shown in the form of plot between actual vs. predicted crop production. Each algorithms are applied on the given datasets and corresponding comparison plot is drawn as shown below.

The plot correspond to SVM is drawn and as SVM gives accurate data for more than 10000 data values peak is observed at this stage.

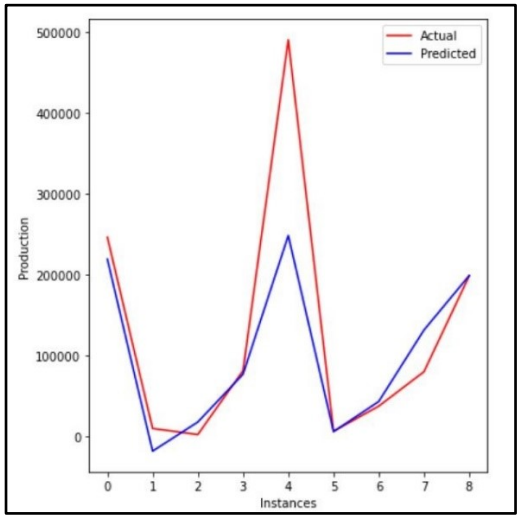


Fig 9. SVM classification- comparison between predicted and actual value of crop production

The plot correspond to KNN is drawn between the actual and predicted crop production. At small value instances the actual and predicted graph is exact replica but it fail to do so at higher values.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:57:29 UTC from IEEE Xplore. Restrictions apply.

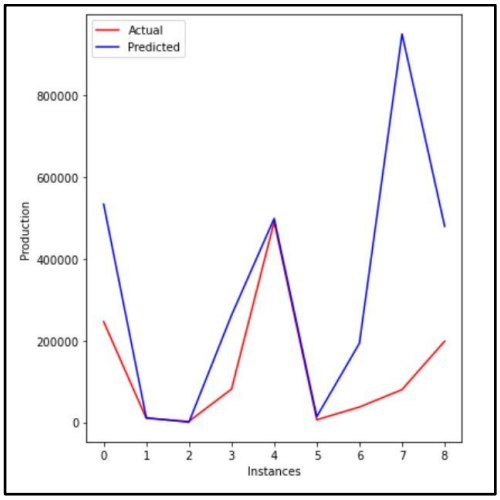


Fig 10. KNN classification- comparison between predicted and actual value of crop production

The plot correspond to Random forest classification which from graph it is concluded that it gives better accurate results on the small data values due to its accuracy in measurement of lower order data values as compared to higher values in actual vs. predicted crop production.

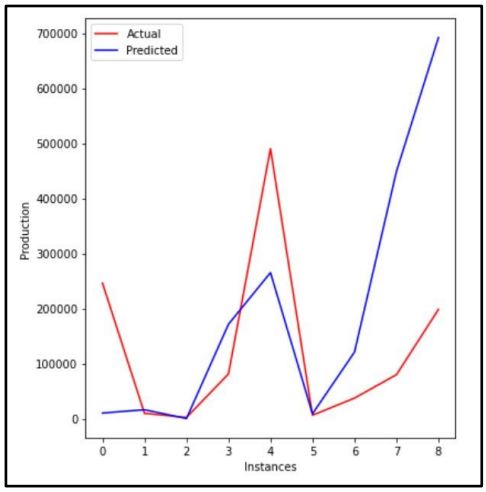


Fig 11. Random Forest- comparison between predicted and actual value of crop production

Based on the results of all three techniques an observation is done showing the accuracy percentage of calculated values with that of theoretical values. The result of observation is thus noted of various data values of dataset and are compared. A total of 11000 data values records are taken for observation.

TABLE I. ACCURACY COMPARISON OF RANDOM FOREST, SVM, KNN

NO OF DATA VALUES USED	Percentage of Accuracy of shown algorithms		
	Random Forest	SVM	KNN
0	0	0	0
2000	70.18	67.13	66.01
3000	70.56	65.4	67.49
4000	70.88	58.34	49.9
5000	68.46	61.58	56.6
6000	55.97	62.92	58.9
7000	63.89	63.01	62.0
8000	58.79	64.79	60.0
9000	60.25	65.25	78.69
10000	59.54	66.23	81.3
11000	56.2	68.8	87.9

In this table it is inferred that on the small number of data the random forest classification is the best to apply for prediction and for large no of data values KNN is recommended.

At the end a custom made webpage is designed having login page and crop prediction page. In crop prediction page after entering all the values of parameters and selecting the type of technique will get the content of Nitrogen, Potassium and phosphorus in soil related to that specific technique.

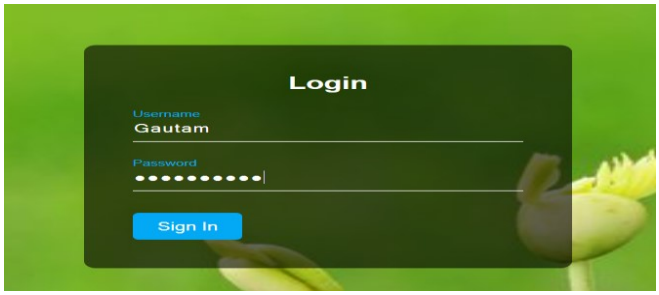


Fig 12. Login page of user defined site

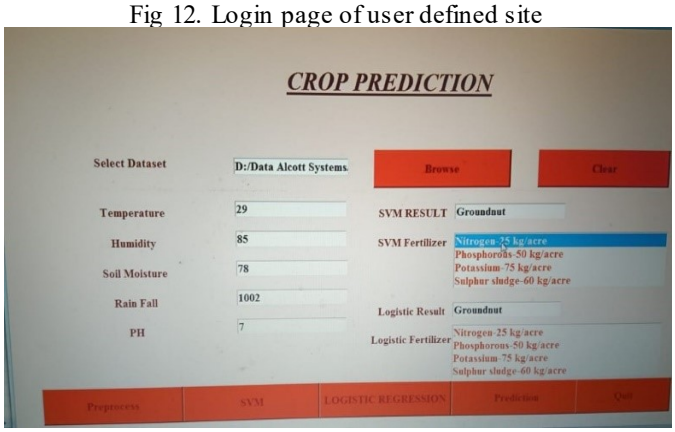


Fig 13. Crop prediction web page

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:57:29 UTC from IEEE Xplore. Restrictions apply.

V. CONCLUSION AND FUTURE SCOPE

This paper showed the transformation of system which is capable of collecting the data from senor node with the help of IoT in the field of agriculture. This system successfully able to sense the data and sending the data locally to thingspeak cloud which is further access by user in his/her custom made website. Data mining techniques are used to predict the crop production in advance so that farmer always got to know the extent of accuracy of their farming techniques. Various methods of regression and classification to get the output and their plot are plotted successfully. In future whether showing the data in custom made website can show the data in mobile app. Further, it detects the disease of the crop by using the concept of image detection. With that suitable type of fertilizer and pesticides is used to eliminate the disease. Artificial Intelligence and neural network could also be used as a regression method that promises to give more precise results.

ACKNOWLEDGMENT

We would like to thank Mrs. Bhavnesh Jaint, for his constant motivation and guidance during the project. Without his timely help and support, the completion of this paper wouldn't have been possible. We would also like to thank the publisher of the research papers from where we have taken references.

REFERENCES

[1] W. Klosgen, J.M. Zytkow, Handbook of data mining and knowledge discovery, Oxford University Press, 2002.

[2] Kiruthika M, Shweta T, Mritunjay O, Kavita S, "Parameter monitoring for the precision agriculture", International Journal Of The Scientific Research and Innovation 2015.

[3] O.P. Uma Maheshwari, M Savitha, "A Study of IoT (Internet of Things) in agriculture", International Journal of Innovative research in computer and communication Engineering.

[4] Dr. N. Suma, S. R. Samson, S. Saranya, G. Shanmugapriya, R. Subhashri " IoT Based Smart Agriculture Monitoring System", International Journal on Recent and Innovation Trends in computing and communication.

[5] Sandip Khot, M. Gaikwad, "Development of cloud based Light Intensity Monitoring system for Green house Using Raspberry Pi" , IEEE International conference on computing communication control and automation (ICCUBEA) Aug 2016

[6] Sheetal V, A Bakshi, Tanvi T, "Green House by using IoT and Cloud computing", IEEE International conference on Recent trends in Electronics, information & communication technology May 2016

[7] Aymen E Khedr, Mona Kadry, Ghada Walid(2015), "Proposed framework for Implementing Data Mining Techniques to enhance Decisions in Agriculture sector Applied case on Food Security Information Center Ministry of Agriculture, Africa", international conference on Communication, management and Information technology (ICCM).

[8] Kuljit Kaur, Kalnwalpreet Singh Atwal "Effect of Temperature and Rainfall on Paddy Yield using data mining ", 2017 7th international conference on Cloud Computing, Data Science and Engineering Confluence.

[9] F. TongKe, "Smart Agriculture Based on Cloud Computing and IoT", in Journal of Convergence Information Technology, Volume 8, Number 2, 2013.

[10] Sivaganesan, D. (2019). DESIGN AND DEVELOPMENT AI-ENABLED EDGE COMPUTING FOR INTELLIGENT-IOT APPLICATIONS. Journal of trends in Computer Science and Smart technology (TCSST), 1(02), 84-94.

[11] Joseph, S. I. T. (2019). SURVEY OF DATA MINING ALGORITHM'S FOR INTELLIGENT COMPUTING SYSTEM. Journal of trends in Computer Science and Smart technology (TCSST), 1(01), 14-24.

[12] Scikit-learn, Machine learning in python retrieved from http: //scikitlearn. Org.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:57:29 UTC from IEEE Xplore. Restrictions apply.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/342877873

Estimation of COVID-19 Prevalence under Closed Population using Capture-Recapture approach

Article *in* International Journal of Science and Research (IJSR) · January 2018

DOI: 10.21275/SR20629043404

CITATIONS
0

READS
18

2 authors, including:



Rakesh Kumar
Delhi Technological University
7 PUBLICATIONS **0** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Performance Measurement View project

Population Statistics View project

All content following this page was uploaded by Rakesh Kumar on 12 July 2020.

The user has requested enhancement of the downloaded file.

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2019): 7.583

Estimation of COVID-19 Prevalence under Closed Population using Capture-Recapture approach

Rakesh Kumar[1], R.S.Mishra[2]

Delhi Technological University, Delhi, India

Abstract: ***The nation-wide lockdown due to COVID-19 created the situation of closed population. This paper presents a general theory for estimating the size of a closed population. The theory is based on a log-linear model, with model parameters representing the natural logarithm of subpopulation size and the natural logarithm of subpopulation capture probabilities. Lastly the paper contains illustration of use of log-linear to closed population.***

Keywords: COVID-19, log-linear, capture-recapture

1. Introduction

Estimating population sizes has long been of interest, from the estimation of the human or ecological population size within regions or countries to the hidden number of corona patients in the state. Total enumeration of the population, for example, via a census, is often infeasible or simply impractical. However, a series of partial enumerations or observations of the population is often possible. This has led to the ideas of capture-recapture methods, which have been extensively used within ecology to estimate the size of wildlife populations, with an associated measure of uncertainty, and are most effectively applied when there are multiple capture occasions. The capture-recapture technique originally developed for animal studies and later has been applied by practitioners and researchers to human populations under the term `multiple-record system' for an arbitrary number of lists and under the term `dual-system especially for two lists. The earliest references to such applications can be traced back to Sekar & Deming (1949) for two samples, Wittes & Sidel (1968) for three samples, Wittes (1974) for four samples and Fienberg (1972) for five samples.

Since state circumstances will vary in terms of disease activity and capacity to respond, caution should be exercised in directly linking severity assessment at a global to national level to actions at the state level. It is likely that information will be limited early in the pandemic while the demand for information simultaneously escalates. If pandemic surveillance is to provide sufficient information and data to assess severity, country/state need to review their existing surveillance procedure and capacity to address the weaknesses to be prepared for pandemic surveillance. Essential components of an effective pandemic corona surveillance system will include: early detection and investigation; comprehensive assessment; and Monitoring. Delhi government launched 5T plan to stop the spread of the novel corona virus. The 5T plan includes Testing, Tracing, Treatment, Teamwork and Tracking based on the South Korean strategy Test, Trace and Isolate. But the step testing requires basis. India's apex medical body Indian Council of Medical Research (ICMR) intensified the random sampling of people who display flu-like symptoms but don't have any history of travel to outbreak zones to determine whether community transmission is taking place. Later ICMR experts aren't convinced about following South Korea's strategy, which has reported more recoveries from Covid-19 than new infections through aggressive testing without lockdowns. Indian experts said they will revise the testing protocol depending on the situation. The ICMR in its study highlighted that Covid-19 containment activities need to be targeted in districts reporting positive cases among Severe Acute Respiratory Infections (SARI) patients and stated that intensifying sentinel surveillance for Covid-19 among SARI patients may be an efficient tool to effectively use resources towards containment and mitigation. Testing: Until now testing was very limited. Only symptomatic people who had history of travel were tested. But now Delhi government plans to aggressively start, testing drives in Delhi areas which are being seen as Corona virus hot spot. Also there is going to be random testing of one lakh people in the next few days. Delhi is the first state to carry out random testing. Second T - Tracing as testing ramps up every single contact of positive cases through cellphone triangulation will be traced. Twenty four hour tabs will be kept on those who are in quarantine to ensure it's being followed. Such focused actions will help in staying at least two steps ahead of virus.
ICMR (Indian Council of Medical Research) conducted a test in few states of country and its research shown that there are many cases of corona virus where it is difficult to find the source i.e. they have no travel history and not get in contact with any such person who has travel history. The Ministry of Home Affairs (MHA) of India stated that the government has adopted a special strategy to fight coronavirus, with specific instructions for the district and state authorities. The strategic approach for COVID-19 containment broadly involves defining the area of operation, applying perimeter control, delineating containment and buffer zones. Besides, there is active search for cases, contact tracing, quarantine, clinical management and awareness generation among public.

Without discussing the methodology, It is estimated by Indian professor Ramanan Laxminarayan, Director and Senior Fellow, The Center for Disease Dynamics, Economics & Policy that by July 2020, 300 to 500 million Indians will get infected and a couple of million Indians can die of corona virus cases and fatalities. A data journalist Rukmini S. Has also presented her observation on projections of corona virus cases concluded that if the rate of COVID-19 spread continues, 30,000 deaths can occur in

Licensed Under Creative Commons Attribution CC BY

India by May 2020 and also mentioned that various variables like lockdowns in India and the enforcement of them are great challenges that statisticians cannot deal with.

Random Sampling

Random sampling is basic to many computing applications in science, statistics, and engineering. Random sampling is used on those occasions when processing the entire dataset is unnecessary and too expensive in terms of response time or resource usage. The savings generated by sampling may arise either from reductions in the cost of retrieving the data from the database or from subsequent "post-processing" of the sample. Sampling is useful for applications which are attempting to estimate some aggregate property of a set of records, such as the total number of records which satisfy some predicate. Thus random sampling is typically used to support statistical analysis of a dataset, either to estimate parameters of interest or for hypothesis testing. Random sampling has three common justifications: reduced database retrieval costs (or time), reduced "post-processing" computational costs, or reduced cost of physical inspections. So, random sampling is a basic sampling technique through which each individual is chosen entirely by chance and each member of the population has an equal probability of being included in the sample. It is the only way to be sure that the sample is not biased and is representative of the population as a whole, and even more complex sampling schemes such as systematic and stratified sampling are based on this basic scheme. ICMR's testing of a small random sample may not be able to capture it. As India is divided its 755 districts into green, orange and red zones. It will entail making sure that green zones are indeed green by deploying testing strategies. Delhi government announced a five-point action plan of "five Ts -- testing, tracing, treatment, team-work and tracking -- and monitoring" to contain the virus spread and said one lakh random rapid anti-body blood tests for COVID-19 tests will be conducted in the city's hotspot areas, the similar plan is followed by other states also.

2. Literature Review

The past literature is reviewed keeping in view for this study is concerned as well as COVID-19. Pandemic is not new for the world, but the lockdown concept is new for country like India experience in year 2020. Estimating population sizes has long been of interest, from the estimation of the human or ecological population size within regions or countries to the hidden number of corona patients in the state. Total enumeration of the population, for example, via a census, is often infeasible or simply impractical. However, a series of partial enumerations or observations of the population is often possible. This has led to the ideas of capture-recapture methods, which have been extensively used within ecology to estimate the size of wildlife populations, with an associated measure of uncertainty, and are most effectively applied when there are multiple capture occasions.

Capture-recapture methods are used in various areas, from epidemiology (Coeli et al., 2000) to zoological studies of populations (Fernandez, 1995) or even in the analysis of the variation of measurement intervals of closed communities of diverse magnitudes (Cormack, 1992). The capture-recapture techniques are used for closed populations, where the effects of birth rate or mortality and migration are not taken into account, in this case it is considered no change in population size. They can also be used for open populations, where population changes caused by birth rate, mortality and migration are taken into account (Cormack, 1989). Sekar and Deming (1949) proposed their research concerned with estimating birth and death rates in an area near Calcutta, India. The research based report of Robles, S. C. et al. (1988) described about usefulness of capture-recapture methods used to estimate completeness at the Ontario Cancer Registry. The method was applied in two fashions; first, using three data sources in a modeling approach and second, using two data sources and standard, simple capture-recapture methods. The modeling approach is more flexible, since several variables that influence cancer registration can be considered and can be used to identify reporting patterns of different data sources. Wittes, J. T (1974) in his paper considered the problem of estimating the total size of a target population from which a study sample has been obtained by merging names from several routinely collected lists. Each list is assumed to be a simple random sample from the target population and all lists are assumed independent. Hook, E. B. (1980) presented his work with goal to compare various methods for estimating the live-birth prevalence of spina bifida (cystica) using Bernoulli census analysis, allowing for under ascertainment in the sources used, and to determine if a stable estimate or at least a useful confidence boundary on the population live-birth prevalence rate could be derived from prenatal diagnosis of spina bifida undertaken as part of a cost-benefit analysis. Brenner, H. (1994) highlighted the drawback of routine systems of disease monitoring, such as morbidity registries, are often prone to specific threats of validity, such as in imperfect diagnoses or imperfect record linkage. Through quantitative assessment of the performance of the two-source capture-recapture method for disease monitoring in the presence of imperfect record linkage. International Working Group for Disease Monitoring and Forecasting (1995) evaluates the potential epidemiologic uses of capture-recapture, which include the primary area of determining disease frequency. Capture-recapture may be a means to effectively "count" new cases (incidence) or count existing cases (prevalence). In addition to generating an estimate of population size, another application of capture-recapture is to assess the costs of ascertainment relative to the degree of accuracy. Fisher N. et al. (1994) worked on routinely collected data from local agencies to provide a reliable estimate of the size and psychiatric morbidity of the homeless population of a given central London by using capture-recapture analysis. A multiple sample, log-linear capture-recapture method was applied to a defined geographical area during 6 months. The method calculates the total homeless population from the sum of the population actually observed and an estimate of the unobserved population. Sheila M. Bird and Ruth King (2018) with changed name of Capture-Recapture Estimation with Multiple Systems Estimation (MSE) advocated that the MSE approach has been preferred when estimating "capture-shy" or hard to-reach populations, including those caught up in the criminal justice system; or homeless; or trafficked; or civilian casualties of war. The paper focuses on whether the uncertainty which qualified their estimates was sufficiently narrow to orient public policy or not, if not, what options

Licensed Under Creative Commons Attribution CC BY

were available and/or taken to reduce the uncertainty or to seek external validation. *Jones H.E.et al.(2016) by using* Capture–recapture analysis recommended for estimating the prevalence of problem drug use or people who inject drugs .The work aim to demonstrate how naive application of Capture–recapture can lead to highly misleading results, and to suggest how the problems might be overcome. The method further validated by a case study of estimating the prevalence of people who inject drugs in Bristol, UK, using Capture–recapture. *Leclerc, P. (2014) already published similar type of work as Jones H.E.et al.(2016) carried but study area was* Island of Montréal, Canada, Using a Six-Source Capture-Recapture Model.

Even after worthiness of this method Stephen N Jarvis et al. (2000) stated with prior proof in his research paper with illustration that there is also serious drawbacks in the use of the mark/recapture technique when applied to injuries. Through example of two sample mark/recapture estimation in an epidemiological setting it is shown to violate virtually all the requirements for valid use of the technique. Further mentioned that very little can be deduced accurately about the scale or characteristics of an unobserved group by the use of mark/recapture applied to two overlapping health event registers. Begon, M. (1983) as a critic to the methods of Jolly he mentioned that there has also been a consistent interest in the method on the part of statisticians, a sporadic interest in field-based critiques of the method, and a considerable tendency, especially since the mid-seventies, to refer to Jolly's paper but make no use of its contents.

3. Capture-Recapture Method and Log-Linear Model

Pierre Simon Laplace (1749-1827) used similar to capture-recapture approach to estimate the size of human population of France. Further Lincoln-Peterson develop a technique with the name capture-recapture

Closed Population: A closed population is one in which the total number of individuals is not changing through births, deaths, immigration or emigration. But in practice, most real populations are not closed, it's a chance the in complete lockdown, the population is definitely closed. An open population is one that is (or could be) changing during the course of study, because of any combination of births, deaths, immigration or emigration. The 'closed captures' data types all assume the population of interest is closed during the sampling period (White *et al*. 1982). Meaning, the models assume that no births or deaths occur and no immigration or emigration occurs. Typically, we refer to a closed population as one that is free of unknown changes in abundance, as we can usually account for known changes.

The assumption is in context of human being is differ from animals in that way that- human being are self conscious and able to report about their illness and checked (tagged), human being always available whenever called for recounting. But this COVID-19 effected the mind and thought of people in such a way that people use to hide themselves from the society and that become a challenging task to the health department to do mapping to identify them from the huge crowd of population, even if it is closed due to lockdown. The 'closed captures' data types all assume the population of interest is closed during the sampling period (White *et al*. 1982). Meaning, the models assume that no births or deaths occur and no immigration or emigration occurs. Typically, we refer to a closed population as one that is free of unknown changes in abundance, as we can usually account for known changes.

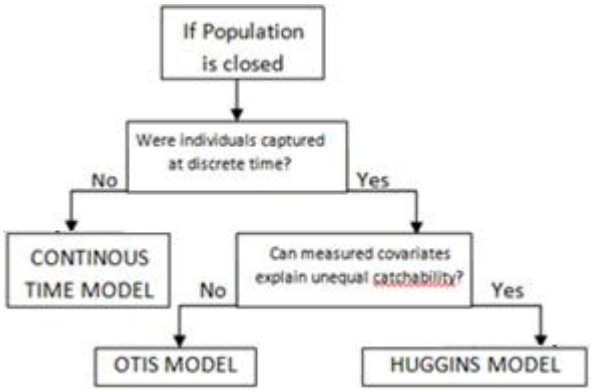


Figure 1: Flow Chart of the methods of CAPTURE-RECAPTURE

Consider a closed population with size N and the people are indexed by 1,2,. . . , N in any arbitrary fashion. In a usual t-sample capture-recapture experiment, people are captured and marked in each trapping sample (or trapping occasion), and we only notice in each sample whether each person was captured or not without recording the sequential order or the time when it was captured. Thus the capture history can be expressed as a N x t matrix (X_{ij}), where X_{ij} = I [the i^{th} person was caught in the j^{th} sample], i = 1, 2,. . . , N, j = 1, 2.. . , t. Extensive models have been discussed for the above discrete-time type data.

If the capture time and the sequential order for each individual can be recorded, then a continuous-time model is needed. Anderson and Wilson (1989) have done some simulation to compare the existing estimators for continuous-time models. Becker (1984), Becker and Heyde (1990) proposed an estimation procedure using a martingale approach.

In the complete lockdown situation the whole population is considered as closed population and the new birth in this period is neglected because the record is unavailable for this period. A closed population is one in which the total number of individuals is not changing through births, deaths, immigrations and emigrations and all these were not in assumption in corona period. In practice, most real populations are not closed. Sometimes, however, the changes over the time period of interest are small enough that the assumption of closure is a reasonable approximation, and the effects of violating that assumption are minimal. For that reason, the analysis of capture-recapture data from closed populations continues to be a topic of interest to researchers.
However closure is subdivided into two components: (i) geographic and (ii) demographic. The geographic closure of population by a boundary, which limits the population, and demographic is about birth, death, immigration and emigration.

Licensed Under Creative Commons Attribution CC BY

The models of Otis et al.(1978) denoted M_0, M_t, M_b, M_h, M_{bh}, M_{tb}, M_{th}, and M_{tbh}.

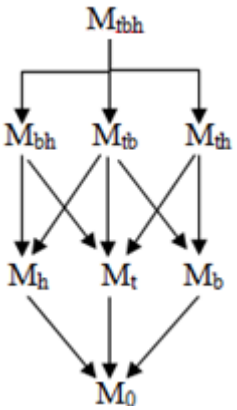
Model M_0 assumes a single common probability of capture. The subscripts t, b, and h correspond to temporal variation in common probability of capture, behavioral response to common probability of capture, and individual response to capture, respectively. Multiple subscripts indicate a combination of capture responses.

The above mentioned eight models summarised by symbol, source of variation in capture probabilities , and the associated estimators, if any , program CAPTURE uses theses names for estimators.

Table 1: Discrete-Time Models and Associated Appropriate Estimator with Log Linear form

Model	Source of variation in capture probabilities	Appropriate estimator	Log Linear form
M_0	None	Null (All estimators given below are valid)	$log\ (P_{ij}) = \alpha$
M_t	Time	Darroch (1958)	$log\ (P_{ij}) = \mu_j$
M_b	Behaviour	Zippin (1956)	$log\ (P_{ij}) = \alpha + \gamma\ Y_{ij}\ (\alpha_i = \alpha)$
M_h	Heterogeneity	Jackknife (1978)	$log\ (P_{ij}) = \alpha_i$
M_{bh}	behaviour heterogeneity	Generalized Removal (Otis et al. 1978)	$log\ (P_{ij}) = \alpha_i + \gamma\ Y_{ij}$
M_{tb}	time behaviour	None	$log\ (P_{ij}) = \mu_j + \gamma\ Y_{ij}$
M_{th}	time heterogeneity	None	$log\ (P_{ij}) = \alpha_i + \mu_j$
M_{tbh}	time behaviour heterogeneity	None	$log\ (P_{ij}) = \alpha_i + \mu_j + \gamma\ Y_{ij}$

Relation of the model



Each arrow points from one model to an immediate special case of the model. Test of assumptions are based partly on these relations among the model

Figure 2: Relation of the Model

There are also capture-recapture models available for the open population like Jolly-Seber model, Modern open population model, Multi State Model, Target Recovery Model and Joint modeling of live and dead recovery, with different assumptions. These models are widely used by researchers to count the animal populations as well as human also.

Log-Linear Models

The log-linear model approach was originally proposed by Fienberg (1972) and has been widely used in many applications (Cormack 1989; Agresti 1994; Evans, Bonett, and McDonald 1994; Coull and Agresti 1999; IWGDMF 1995). In this approach, the data are regarded as a form of an incomplete 2^t contingency table (t is the number of lists) for which the cell corresponding to those individuals uncaptured by any of the samples is missing. Then various log-linear models are fitted to the observed cells and the chosen model is projected onto the unobserved cell by assuming that there is no t-sample interaction. We use the three-sample case as an illustration. There are seven observed cells: z_{111}, z_{110}, z_{101}, z_{100}, z_{011}, z_{010}, and z_{001}, where z_{100} is the number of individuals captured in sample 1 only and z_{011} is the number of individuals caught in samples 2 and 3 but not in sample 1. A similar interpretation pertains to other capture histories. The missing cell z_{000} denotes the number of missed individuals. The log-linear approach models the logarithm of the expected value of each observable category. The most general log-linear model for the three-sample case is

$$log\ E(z_{ijk}) = u + u_1 I(i = 1) + u_2 I(j = 1) + \underline{u}_3 I(k = 1) + u_{12} I(i = j = 1) + u_{13} I(i = k = 1) + u_{23} I(j = k = 1) + u_{123} I(i = j = k = 1); \quad \text{----------(1)}$$

where I(A) is an indicator function for event A. The above is a reparameterization of the eight expected values. We have seven observed cells, whereas there are eight parameters in the right-hand side of equation (1). Therefore, a commonly adopted assumption is that there is no three-sample interaction term, i.e., $u_{123} = 0$, which allows an extrapolation formula for the missing cases using the fitted values of the observable cells. The independent model with main effect terms is equivalent to model M_t in Table 1. The log-linear models corresponding to the multiplicative models M_b and M_{tb} are discussed in Cormack (1989), who also showed how to t various models using the GLIM package (GLIM is a statistical package(marketed by NAG) which analyses generalised linear models by iterated reweighted least squares). Many authors stated that random sampling is required in capture studies. This assumption stems from ball and urn experiments, in which marked and unmarked balls are shaken completely, and a random sample is taken at the end of each sampling occasion. Traditional sampling methods include procedures for drawing random samples. Uses of the procedures require knowledge of the sampling probabilities (for finite populations). Deliberate control over the elements to be sampled is required. It is unrealistic to think that if any COVID patient is captured (marked) in one corner of the trapping grid may be captured subsequently in the opposite corner of the boundary. The concept of random sampling does not apply to situations assumed in Models M_b, M_h, M_{bh}, M_{tb}, M_{th}, and M_{tbh}. The goal of these models is to provide an analysis of the sampled data in the face of behavioural response and heterogeneity, both of which are contrary to the traditional role of random sampling.

4. Data Source For Illustration

According to ICMR the data of the following people should be considered for COVID-19 testing.

1) Symptomatic (fever, sore throat, running nose, dyspnea, etc) individuals who have travelled to affected countries

Licensed Under Creative Commons Attribution CC BY

(China, Hong Kong, Japan, South Korea, Singapore, Iran, Italy etc.)
2) Close contacts of confirmed positive cases of COVID-19 infection.
3) All individuals evacuated and quarantined from Wuhan, China and Diamond Princess ship, Japan.

It includes information on COVID-19 service activities by private and public hospitals and community health services from 25th March 2020 to 31st May 2020 (Lock-down period in India). This study may be use to estimate the number of people who were infected by the corona virus (COVI-19) in an outbreak in whole India and for more accuracy study area may be carried out at state wise. For example if the study area is Delhi selected with approximately 16787941 population in 11 different districts. Our data are restricted to census records of government. In total, 271 cases were reported from the following three sources. (1) A-list: records based on Symptomatic (fever, sore throat, running nose, dyspnea, etc) individuals who have travelled to affected countries (China, Hong Kong, Japan, South Korea, Singapore, Iran, Italy etc.) . There were 135 ascertained cases. (2) B-List: records provided by Close contacts of confirmed positive cases of COVID-19 infection. The records included 122 cases reported by doctors in local hospitals. (3) C-list: records based on All individuals evacuated and quarantined from Wuhan, China and Diamond Princess ship, Japan.. A total of 126 cases were in the C-list.

Table 2: Data on corona virus infection

A	B	C	Data
1	1	1	Z_{111}= 28
1	1	0	Z_{110}= 21
1	0	1	Z_{101}= 17
1	0	0	Z_{100}= 69
0	1	1	Z_{011}= 18
0	1	0	Z_{010}= 55
0	0	1	Z_{001}= 63
0	0	0	-----

The data are given in Table 2 Several log-linear models were fitted to these data. Except for the saturated model, the log-linear models, which do not take into account heterogeneity (i.e. models (AC, B), (BC, A), (AB, C), (AB, BC), (AB, AC) and (BC, AC)), do not fit the data well. The quasi-symmetric and three partial quasi-symmetric models (see Darroch *et al*., 1993), which incorporate heterogeneity effects, fit the data well. The four adequate models produce approximately the same estimates, 1300, with an approximate estimated s.e. of 520. This relatively large estimated i.e. shows the data are actually insufficient to model the heterogeneity effect.

Estimation Procedure

$$\hat{C} = 1 - \frac{1}{t} \sum_{k=1}^{t} \frac{S_k}{n_k}$$

Equation for estimator of sample coverage ----------(2)

The singletons cannot contain any overlap information. Therefore, the sample coverage can be Interpreted as a measure of overlap. When neither dependence exists, it follows from the definition of C that $C = D/N$, where D denotes the average of overlapped cases, i.e.

$$D = \frac{1}{t} \sum_{k=1}^{t} \{ \sum_{k=1}^{N} I [\sum_{j \neq k} Xij > 0] = M - \frac{1}{t} \sum_{k=1}^{t} S_k$$

----------(3)

Here, M denotes the number of different cases identified in at least one of the lists. Hence, in the independent case, a valid estimator is

$$N_0 = D/\hat{C} \qquad \text{----------(4)}$$

the sample coverage is estimated to be $\hat{C} = 51.27\%$ and the average of the overlapped cases is D= 208.67. If we incorrectly assume that the three samples are independent, then from (4) an estimate would give $N_0 = D/\hat{C} = 407$, which is slightly larger than the estimate of 388 based on an independent log-linear model. We feel these data, with estimated sample coverage of 51%, do not contain enough information to correct for undercount. Thus, this example presents a very valuable data set with the advantage of a known true parameter.

5. Discussion

We discuss first the extent to which this study provides evidence for or against the use of mark/recapture techniques in the epidemiological setting. We then consider the relationship of these findings to other similar studies. To what extent then does the present study meet the "requirements" for the mark/ recapture method? As far as possible, the study has attempted to define a "closed" population—Otis condition "a". Note here that "closed" does not infer that the COVID-19 patient population of Delhi is unchanged—rather that the reference population for individual cases is the same for both lists. In this study the reference population is COVID-19 patients from Delhi addresses with following people should be considered for COVID-19 testing.
1) Symptomatic (fever, sore throat, running nose, dyspnea, etc) individuals who have travelled to affected countries (China, Hong Kong, Japan, South Korea, Singapore, Iran, Italy etc.)
2) Close contacts of confirmed positive cases of COVID-19 infection.
3) All individuals evacuated and quarantined from Wuhan, China and Diamond Princess ship, Japan.

Indian Council of Medical Research (ICMR) affirms, while there is no readily available test kit to detect coronavirus infection, doctors and researchers go by a series of tests to confirm the infection i.e. Swab test, Nasal aspirate and tracheal aspirate test. Indian Council of Medical Research (ICMR) suggests that anyone with a recent travel history to COVID-19 affected countries or someone who has been in close contact with infected patients is also advised screening. Despite the extended time over which our lists are compiled, an individual's availability for both lists is almost concurrent. The crucial assumption, that an individual in the population from which one list is a sample, is also in the population from which the other list is a sample, is thus

Licensed Under Creative Commons Attribution CC BY

satisfied. In any area of epidemiology or, in fact, any human monitoring system involving population estimates of the number of people who have a certain characteristic, capture-recapture should be considered a complementary or an alternate approach to traditional counting technology. One application of capture-recapture, in addition to generating an estimate of population size, is to assess the costs of ascertainment relative to the degree of accuracy obtained. A primary goal of monitoring is the ability to obtain accurate and timely incidence rates at a reasonable cost. The more expensive it is to accurately estimate the counts of disease, the less likely it is to assess that disease and the smaller the geographic coverage. When setting out to determine incidence rates, therefore, one should first determine the amount of precision desired, as greater degrees of precision are likely to be more expensive. The appropriate degree of precision required for estimating incidence rates will vary by the use(s) to which the data are put. There are no right or wrong decisions as to how many samples should be examined, or the degree of ascertainment needed, as this is a function of both the acceptable degree of precision and the amount of resources available to carry it out.

We strongly believe that with any disease monitoring system, whether communicable or non-communicable, there should be the means to formally evaluate the degree of undercount. With capture-recapture, we have this technology. All disease monitoring systems should be able to estimate the proportion of the diseases diagnosed in the community, region, or nation that were identified by the monitoring system. Without this information, it is difficult, if not impossible, to determine the degree to which the rates of disease are compromised by missing cases. Capture-recapture is not a panacea. However, it offers public health officials the opportunity to get closer to the truth if the methods are used appropriately. There is much to be learned in the development of capture-recapture models for use in epidemiology. The assumptions need to be assessed and the models tested. Epidemiologists and biostatisticians working with human disease need to continue the dialogue with those who have developed the approaches using capture-recapture. Much remains to be learned, and there is great potential. As indicated by the IWGDMF (1995), the log-linear model approach has many advantages as follows: (a) all models are under a unified framework; (b) model selection can be easily implemented and carried out in a flexible fashion, based on data and guided by prior information; (c) tests are available for comparing models; (d) dependence can be incorporated by adding proper interactions; and (e) all inference is within the mainstream of statistical data analysis. An untestable assumption involved is that the highest-order interaction does not exist. The IWGDMF commented that the existence of heterogeneity in three-list data might result in the lack of a reliable estimate. Another concern is that two equally fitted models might produce quite different estimates. As the number of lists increases, the number of adequate models increases rapidly and thus model selection causes further problems.

References

[1] Anderson, D. R., and Wilson, K. R., 1989: Capture-recapture population estimation using continuous time models. Progress Report.
[2] Becker, N. G., 1984: Estimating population size from capture-recapture experiments in continuous time. Austral. J. Statist. 26, 1-7.
[3] Becker, N. G., and Heyde, C. C., 1990: Estimating population size from multiple recapture experiments. Stochastic Process Appl. 36, 77-83.
[4] Bird SM, King R., "Multiple Systems Estimation (or Capture-Recapture Estimation) to Inform Public Policy", *Annual review of statistics and its application. 2018 Mar; 5: 95-118*, doi: 10.1146/annurev- statistics-031017-100641
[5] Begon, M. (1983). Abuses of Mathematical Techniques in Ecology: Applications of Jolly's Capture- Recapture Method. Oikos, 40(1), 155. doi:10.2307/3544213
[6] Coeli, C.M., Veras, R.P. and Coutinho, E.S.F., 2000. Metodologia de captura-recaptura: uma opção para a vigilância das doenças não transmissíveis na população idosa. Cadernos de Saúde Pública, vol. 16, no. 4, pp. 1071-1082. PMid:11175530. http://dx.doi.org/10.1590/S0102-311X2000000400025.
[7] Cormack, R.M., 1992. Interval estimation for mark-recapture studies of closed populations. Biometrics, vol. 48, no. 2, pp. 567-576. PMid:1637979. http://dx.doi.org/10.2307/2532310.
[8] Cormack, R.M., 1989. Log-linear models for capture-recapture. Biometrics, vol. 45, no. 2, pp. 395-413. http://dx.doi. org/10.2307/2531485.
[9] Fernandez, F.A.S., 1995. Métodos para estimativas de parâmetros populacionais por captura, marcação e recaptura. In: P.R. PERES-NETO, J.L. Valentin and F.A.S. Fernandez, eds. Programa de pós-graduação em ecologia. Rio de Janeiro: Instituto de Biologia, UFRJ, pp. 1-26. http://dx.doi.org/10.4257/ oeco.1995.0201.01.
[10] Fienberg, S. E. (1972) The multiple recapture census for closed populations and incomplete 2^k contingency tables, Biometrika, 59, pp. 591-603.
[11] Hook, E. B., Albright, S. G., & Cross, P. K. (1980). Use of bernoulli census and log-linear methods for estimating the prevalence of spina bifida in livebirths and the completeness of vital record reports in New York state. American Journal of Epidemiology, *112(6), 750– 758.* doi:10.1093/oxfordjournals.aje.a113047
[12] H. Brenner (1994). Application of Capture-Recapture Methods for Disease Monitoring: Potential Effects of Imperfect Record Linkage. Methods Inf Med 1994; 33(05): 502-506. DOI: 10.1055/s-0038-1635053.
[13] International Working Group for Disease Monitoring and Forecasting (1995). Capture-Recapture and Multiple-Record Systems Estimation II: Applications in Human Diseases. American Journal of Epidemiology, 142(10), 1059–1068. doi:10.1093/oxfordjournals.aje.a117559
[14] Jones H.E., Welton NJ, Ades AE, Pierce M, Davies W, Coleman B, Millar T, Hickman M. Problem drug use prevalence estimation revisited: heterogeneity in capture–recapture and the role of external evidence.

Licensed Under Creative Commons Attribution CC BY

Addiction (Abingdon, England). 2015 Dec 28; 111(3): 438-447. doi: 10.1111/add.13222
[15] Jarvis SN, Lowe PJ, Avery A, *et al.*Children are not goldfish—mark/recapture techniques and their application to injury data Injury Prevention 2000;**6:**46-50.
[16] Leclerc P, Vandal AC, Fall A, Bruneau J, Roy É, Brissette S, Archibald C, Arruda N, Morissette C., "Estimating the Size of the Population of Persons Who Inject Drugs in the Island of Montréal, Canada, Using a Six-Source Capture-Recapture Model", Drug Alcohol Depend. 2014 Sep 1; 142:174-80. Epub 2014 Jun 24. doi: 10.1016/j.drugalcdep.2014.06.022.
[17] Nigel Fisher, Stuart W Turner, Robert Pugh, Colin Taylor, Estimating numbers of homeless and homeless mentally ill people in north east Westminster by using capture-recapture analysis , BMJ. 1994 Jan 1; 308(6920): 27–30.doi: 10.1136/bmj.308.6920.27
[18] Robles, S. C., Marrett, L. D., Aileen Clarke, E., & Risch, H. A. (1988). An application of capture-recapture methods to the estimation of completeness of cancer registration. Journal of Clinical Epidemiology, 41(5), 495–501. doi:10.1016/0895-4356(88)90052-2
[19] Sekar, C. & Deming W. E. (1949) On a method of estimating birth and death rates and the extent of registration, Journal of the American Statistical Association, 44, pp. 101-115.
[20] Wittes, J. T. & Sidel, V. W. (1968) A generalization of the simple capture-recapture model with applications to epidemiological research, Journal of Chronic Diseases, 21, pp. 287-301.
[21] Wittes, J. T. (1974) Applications of a multinomial capture-recapture method to epidemiological data, Journal of the American Statistical Association, 69, pp. 93-97.
[22] White, G.C., Anderson, D. R., Burnham, K. P., and Otis, D. L. (1982) Capture-recapture and removal methods for sampling closed populations. Los Alamos National Laboratory Publication LA-8787-NERP. Los Alamos, NM.
[23] White, G.C., Anderson, D. R., Burnham, K. P., and Otis, D. L. (1982) Capture-recapture and removal methods for sampling closed populations. Los Alamos National Laboratory Publication LA-8787- NERP. Los Alamos, NM.
[24] Wittes, J. T., Colton, T., & Sidel, V. W. (1974). Capture-recapture methods for assessing the completeness of case ascertainment when using multiple information sources. *Journal of Chronic Diseases, 27(1-2), 25–36.* doi:10.1016/0021-9681(74)90005-8

Author Profile



Rakesh Kumar (rakeshtech1@gmail.com) Graduated in Industrial Engineering and Master's in Mechanical Engineering with specialization in Computer Integrated Manufacturing from India. He also obtained Post Graduate Diploma in IPR. He is currently an Assistant Professor and research scholar at Delhi Technological University in Department of Mechanical. His research interest includes Total Quality Management, Reliability Engineering, Operations and Supply chain management, Fuzzy Logic, AI and Decision Science and Intellectual Property Rights. http://orcid.org/0000-0001-5060-1259



R.S. Mishra (Professor) (rsmishra@dce.ac.in) received the Ph.D. degree in Mechanical engineering from IIT Delhi, Indiaand currently at Delhi Technological University serving since 1997. He having 150 research papers in peer reviewed in International journals and 75 research papers in proceedings of International and National conferences. Prior to joining DTU, he served 09 years at different engineering colleges of India, viz, REC Hamirpur, PAU Ludhiana (Punjab)and HAU, Hisar (Haryana.). His research area includes Green Technology, Refrigeration & Air Conditioning, Metal Cutting Technology, Total Quality Management, Solar Energy Technology, Bio Fuels, power Plant engineering.

Licensed Under Creative Commons Attribution CC BY

Materials Today: Proceedings xxx (xxxx) xxx

Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

Experimental studies of vapour compression refrigeration system with eco-Friendly primary refrigerant and brine mixed with nano particles as secondary refrigerant

Uma Shankar Prasad [a,*], R.S. Mishra [b], R.K. Das [a]

[a] IIT(ISM), Dhanbad, India
[b] Delhi Technological University (DTU), Delhi, India

ARTICLE INFO

Article history:
Received 3 March 2020
Received in revised form 3 June 2020
Accepted 4 June 2020
Available online xxxx

Keywords:
Alternative refrigerants
Eco friendly refrigerants
Nano-refrigerant
Energy exergy analysis
Irreversibility analysis

ABSTRACT

This paper is to study the thermal analysis of vapor compression refrigeration system using R134a in primary circuit and Al_2O3-Water based nano-fluids in secondary circuit. The study provides the information of the secondary fluids input conditions, Performance parameters of vapour compression refrigeration system using eco-friendly refrigerants and without nano materials, Effect of overall evaporator heat transfer coefficient using different nano particles mixed with R718 in the secondary evaporator circuit and R134a in primary circuit, the compressor power consumption and the system overall energy performance. Such an analysis with the help of engineering equation solver software (EES) tool can be conveniently useful to compare the thermal performance of different nano-particles (i.e., Al_2O3, TiO_2) based nano fluid as a secondary fluid in a vapor compression refrigeration system.
© 2020 Elsevier Ltd. All rights reserved.
Selection and Peer-review under responsibility of the scientific committee of the 2nd International Conference on Nanoscience and Nanotechnology.

1. Introduction

Nanofluids are a new class of fluids engineered by dispersing nanometer-sized materials (nanoparticles, nanofibers, nanotubes, nanowires, nanorods, nanosheet, or droplets) in base fluids. In other words, nanofluids are nanoscale colloidal suspensions containing condensed nanomaterials. Nanofluids have been found to possess enhanced thermophysical properties such as thermal conductivity, thermal diffusivity, viscosity, and convective heat transfer coefficients compared to those of base fluids like oil or water. It has demonstrated great potential applications in many fields. For a two-phase system, there are some important issues we have to face. One of the most important issues is the stability of nanofluids, and it remains a big challenge to achieve the desired stability of nanofluids (Fig. 1).

1.1. Brief literature review

Lots of research work has been done for replacing "old" refrigerants with "new" refrigerants Ahamed et al. [1] emphasized on use of hydrocarbons and mixture compressor shows much higher exergy destruction as compared to rest of components in the vapour compression refrigeration system and this exergy destruction can be minimized by using of nanofluid and nanolubricants in compressor. Anand [2] did detailed exergy analysis of 2 ton of refrigeration capacity window air conditioning test rig with R22 as working fluid and reached to the conclusions , that irreversibility in system components will be highest when the system is 100% charged and lowest when 25% charged and irreversibility in compressor is highest among system. Arora et al. [3] developed numerical model of actual vapour compression refrigeration system with liquid vapour heat exchanger and did energy and exergy analysis on the same in the specific temperature range of evaporator and condenser and concluded that R502 is the best refrigerant compared to R404A and R507A and compressor is the worst component and liquid vapour heat exchanger is best component of the system in case of exergy transfer. Bolaji et al. [4] had done experimentally comparative analysis of R32, R152a and R134a refrigerants in vapour compression refrigerator and concluded that R32 shows lowest performance whereas R134a and R152a showing nearly same performance but best performance was obtained of system using R152a. Chopra et.al [5] carried out analysis of energetic and exergetic based comparison Multiple Evaporators with

* Corresponding author.
E-mail address: ushankar@piemr.edu.in (U. Shankar Prasad).

https://doi.org/10.1016/j.matpr.2020.06.083
2214-7853/© 2020 Elsevier Ltd. All rights reserved.
Selection and Peer-review under responsibility of the scientific committee of the 2nd International Conference on Nanoscience and Nanotechnology.

Please cite this article as: U. Shankar Prasad, R. S. Mishra and R. K. Das, Experimental studies of vapour compression refrigeration system with eco-Friendly primary refrigerant and brine mixed with nano particles as secondary refrigerant, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2020.06.083



Fig. 1. (a) VCRS setup.

compound compression and flash intercooler with individual and multiple throttle valves and find out the effect of flash intercooler on second law performances. Han et al. [6] Under different working conditions experimental results revealed that there could be replacement of R407C in vapour compression refrigeration system having rotor compressor with mixture of R32/R125/R161 showing higher COP, less pressure ratio and slightly high discharge compressor temperature without any modification in the same system. Mishra et al. [7–9] performed numerical analysis of vapour compression refrigeration system using R134a, R143a, R152a, R404A, R410A, R502 and R507A, and discussed the effect of evaporator temperature, degree of subcooling at condenser outlet, superheating of evaporator outlet, vapour liquid heat exchanger effectiveness and degree of condenser temperature on COP and exergetic efficiency. They reported that evaporator and condenser temperature have significant effect on both COP and exergetic efficiency and also found that R134a has the better performance while R404a has poor performance in all respect (Table 1).

1.2. Problem statement

From the above literature review following points are left by different authors-

- Different arrangements & combinations of refrigerants with nano particles are still left for research.
- Al_2O_3 , TiO_2 & CuO response is better and they may improve the COP of refrigerator by introducing a secondary circuit in the evaporator unit.

But still there is vast area is to explore for performance improvement in VCRS at low cost and low input (Table 2).

Table 1
Performance parameters of vapour compression refrigeration system using eco-friendly refrigerants and without nano materials.

Refrigerants	COP	ED_Ratio	ETA_Second
R134a	2.973	2.447	0.2901
R404a	2.705	3.158	0.2717
R407a	2.754	2.333	0.30

Table 2
Effect of overall evaporator heat transfer coefficient using different nano particles mixed with R718 in the secondary evaporator circuit and R134a in primary circuit.

Nano particles	Overall heat transfer coefficient (U_e)	Difference in water temperature (°C)	Difference in brine temperature (°C)
Copper oxide	1372.14	15.71	11.58
Copper	1367.18	14.55	11.44
Al_2O_3	1318.0	14.62	11.5
Ti_2O_3	1242.43	14.43	11.32
Without nano particles	691.04	12.24	9.29

1.3. Materials and methodology

To improve thermal performance of vapour compression air conditioning systems by improving:

- First law efficiency- According to first law of thermodynamic energetic efficiency /COP is defined as the ratio of net air conditioning effect to the per unit power consumed. First law analysis restricted to calculate only coefficient of performance of the systems and,
- Second law efficiency- The concept of exergy was given by second law of thermodynamics. Second law efficiency is the exergy of the heat abstracted in to the evaporators from the space to be cooled and exergy of fuel is actual compressor work input.

In this paper our main focus on performance of three suspended nano particles i.e., Al_2O_3, TiO_2, & CuO in water based refrigerant in secondary circuit and in primary circuit we are using eco-friendly refrigerants (Table 3).

Table 3
Effect of overall evaporator heat transfer coefficient using different nano particles mixed with R718 in the secondary evaporator circuit and R134a in primary circuit.

Nano particles	Overall heat transfer coefficient (U_e)	% improvement
Copper oxide	1325.7	91.41
Copper	1324.25	91.4099
Al_2O_3	1318.70	90.6077
Ti_2O_3	1242.43	79.6
Without nano particles	691.04	—

Please cite this article as: U. Shankar Prasad, R. S. Mishra and R. K. Das, Experimental studies of vapour compression refrigeration system with eco-Friendly primary refrigerant and brine mixed with nano particles as secondary refrigerant, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2020.06.083

1.4. Results

In this paper, first law and second law analysis of vapour compression refrigeration systems using multiple evaporators and single compressor and single expansion valve with eco-friendly refrigerants in the system and secondary refrigerant (water used in secondary circuit with and without nano-particles mixed with water used as refrigerant) have been presented with the help of EES software results.

1.5. Conclusion

Various nano-particles with R134a refrigerant can be used as an excellent refrigerant to improve the heat transfer characteristics in a refrigeration system. A successful model has been designed and the basic theoretical heat transfer analysis of the refrigeration system has been done.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

I Mr. Uma Shankar Prasad, Research scholor, IIT(ISM), Dhanbad 149 and Assistant Professor at PIEMR, Indore gratefully acknowledge my Research guide Prof. R.K das, Mechanical Department IIT (ISM), Dhanbad and Prof. R.S Mishra, Mechanical Department DTU, New Delhi for providing funds for the research to review this article.

References

[1] U. Ahamed, R. Saidur, H.H. Masjuki, A review on exergy analysis of vapor compression refrigeration system, Int J Renewable and sustainable energy reviews 15 (2011) 1593–1600.
[2] S. Anand, S.K. Tyagi, Exergy analysis and experimental study of a vapour compression refrigeration cycle, Int J Therm Anal Calorim, 2012, 110:961-971.
[3] AkhileshArora, S.C. Kaushik, Theoretical analysis of a vapour compression refrigeration system with R502, R404A and R507A, Int J Refrigeration. 2008 (31) (2008) 998–1005.
[4] B.O. Bolaji, M.A. Akintunde, T.O. Falade, Comparative analysis of performance of three ozone-friends HFC refrigerants in a vapor compression refrigerator, Int J Sustainable Energy & Environment, Vol-2, pp.61-64, 2011.
[5] Chopra et.al, Energetic and Exergetic Based Comparison Multiple Evaporators with Compound Compression and Flash Intercooler with Individual or Multiple Throttle Valves International Journal of Advance Research & Innovations (1), 2013, 73-81.
[6] Han, et.al, Cycle performance study on R32/ R125/ R161 as an alternative refrigerant to R407C. Int J Applied Thermal Engineering, 2007, 27:2559-2565.
[7] Mishra et.al, Methods for improving thermal performances of vapour compression Refrigeration system using eleven ecofriendly refrigerants, ISTE conference on "Technological Universities and Institutions in New Knowledge Age: Future Perspectives and Action plan, ISTE, 149, 9, Delhi Technological University, 2013.
[8] Mishra, Irriversibility analysis of multi-evaporators vapour compression refrigeration systems using new and refrigerants: R134a, R290, R600, R600a, R1234yf,R502, R404a and R152a and R12, R502, Int. J. Adv. Res. Innovations 1 (2013) 180–193.
[9] Mishra, Thermodynamic performance evaluation of multiple evaporators, single compressor, single expansion valve and liquid vapour heat exchanger in vapour compression refrigeration systems using thirteen ecofriendly refrigerants for reducing global warming and ozone depletion, Int. J. Adv. Res. Innovations 1 (2014) 163–171.

Please cite this article as: U. Shankar Prasad, R. S. Mishra and R. K. Das, Experimental studies of vapour compression refrigeration system with eco-Friendly primary refrigerant and brine mixed with nano particles as secondary refrigerant, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2020.06.083

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/335459256

Experimental Validation of Fully Informed Particle Swarm Optimization Tuned Multi-Loop L-PID Controllers for Stabilization of Gantry Crane System

Chapter · January 2020

DOI: 10.1007/978-981-13-8618-3_61

CITATION
1

READS
32

6 authors, including:



Sudarshan K Valluru
Delhi College of Engineering (Now Delhi Technological University) Govt.of NCT Delhi
37 PUBLICATIONS **78** CITATIONS
SEE PROFILE



Daksh Dobhal
Delhi Technological University
4 PUBLICATIONS **1** CITATION
SEE PROFILE



Kumar Kartikeya
Delhi Technological University
4 PUBLICATIONS **1** CITATION
SEE PROFILE



Manpreet Kaur
Bow Valley College
9 PUBLICATIONS **1** CITATION
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Design of Multi-Loop L-PID and NL-PID Controllers: An Experimental Validation View project

meta algorithms application to nonlinear dynamical systems View project

All content following this page was uploaded by Sudarshan K Valluru on 14 July 2020.

The user has requested enhancement of the downloaded file.

Experimental Validation of Fully Informed Particle Swarm Optimization Tuned Multi-Loop L-PID Controllers for Stabilization of Gantry Crane System



Sudarshan K. Valluru, Madhusudan Singh, Daksh Dobhal, Kumar Kartikeya, Manpreet Kaur and Arnav Goel

Abstract Linear PID controllers are commonly used as an electrical component to decrease the error between anticipated set value and actual measured values for control of various benchmarked systems. The Multi-Loop Linear PID (ML-PID) controller, gives a robust and efficient performance in most of the situations. This paper presents the implementation of linear PID controller to stabilize and control the Gantry Crane System. Optimal performance is obtained for a few specific combinations of the proportional, integral, and derivative gains, which makes it essential to tune these values through Optimization techniques. The Fully Informed Particle Swarm Optimizer (FIPSO) is used to tune the gain values of the ML-PID controlled Gantry Crane System. These values are validated experimentally, and obtained results prove that the FIPSO tuned multi-loop linear PID controller quickly stabilizes the system subjected to external disturbances.

Keywords Multi-Loop linear PID (ML-PID) · Fully Informed Particle Swarm Optimization (FIPSO) · Gantry crane system

S. K. Valluru (✉) · M. Singh · D. Dobhal · K. Kartikeya · M. Kaur · A. Goel
Department of Electrical Engineering, Incubation Center for Control Dynamical Systems and Computation, Delhi Technological University, New Delhi 110042, Delhi, India
e-mail: sudarshan_valluru@dce.ac.in

M. Singh
e-mail: madhusudan@dce.ac.in

D. Dobhal
e-mail: dakshdobhal@gmail.com

K. Kartikeya
e-mail: kkartikeya1998@gmail.com

M. Kaur
e-mail: manpreetdtu.26@gmail.com

A. Goel
e-mail: arnav_bt2k16@dtu.ac.in

© Springer Nature Singapore Pte Ltd. 2020

S. Choudhury et al. (eds.), *Intelligent Communication, Control and Devices*,
Advances in Intelligent Systems and Computing 989,
https://doi.org/10.1007/978-981-13-8618-3_61

1 Introduction

The Linear PID is one of the most widely used electrical component in the systems industry [1, 5]. It provides efficient solutions while being computationally inexpensive and simple to implement and simulate. The Linear PID can control and modulate responses of benchmarked dynamical systems [11, 12]. Though, if the system has nonlinearities, the controller may produce unsatisfactory response including high peak overshoot and large settling time, but still, it has limitations like optimality and tuning [10]. Thus, the difficulty of tuning is to be challenging particularly if the multi PID controllers have more than three objectives. The three terms in the controller namely Proportional, Integral, and Derivative can be varied to get an optimal response which is calculated according to the external disturbances by evolutionary computation [8].

In recent years, Particle Swarm Optimization has gained much attention of researchers as a bio-inspired algorithm for the optimization techniques especially for the multimodal and multi-output functions [6, 13]. Although the standard PSO showed better results and faster convergence time to reach global optima for unimodal objective functions, it suffers from falling of local minima problem for nonlinear, multimodal objective functions [6]. This paper addresses the simultaneous tuning of ML-PID controllers using a new variant of FIPSO [2, 4], and the obtained gain values are tested to control and stabilize the Gantry Crane System [3].

The rest of the paper is organized as following sections. Section 2 describes the design of a Multi-Loop Linear PID Controller. Mathematical modeling of the Gantry Crane system is presented in Sect. 3. The tuning of the Controller using FIPSO is presented in Sect. 4. Experimental results and discussions are given in Sect. 5, and finally, Sect. 6 outlines the conclusion of the paper.

2 Design of Multi-Loop L-PID Controllers

The linear proportional integral derivative controller operates upon the error term $\varepsilon(t)$, the mismatch between the real time output and reference input signals to generate the correction signal $\alpha(t)$. The correction signal is obtained as the sum of the error, derivative of the error, and integral of the error multiplied by their respective coefficients. Usually, the three coefficients are denoted by P, I, and D respectively. Hence the expression for the correction signal is obtained as (1):

$$\alpha(t) = K_p \varepsilon(t) + K_i \int_{t1}^{t2} \varepsilon(t) \cdot dt + K_d \frac{d\varepsilon(t)}{dt} \tag{1}$$

The primary task is determining the optimal combinations of the coefficients which yield the best performance. The term proportional to the error reduces the system response time as it is increased till a maximum allowed value above which

sustained oscillations are introduced in the system response. In reducing the steady state error, the major part is played by the integral term. The derivative term leads to a reduction of the corrected signal if the error is increasing.

3 Gantry Crane System

The crane system, in spite of its simple structure, is a complex system owing to its nonlinear nature [7]. Because of this, it is considered to be one of the under accutated benchmark of control systems. If it is assumed that the payload is a lumped-mass distribution rather than a distributed-mass system, the gantry crane can be modeled as a simple pendulum. Utilizing the assumptions of payload being a point mass distribution, massless cable, the cable of fixed length and not flexible and neglecting the damping coefficient of the cable, the free body diagram of the crane and functional diagrams are shown in Fig. 1 and Fig. 2 respectively. Consider M, m, l, b, and I to be the mass of the trolley, mass of the pendulum bob, length of the cable, the coefficient of friction of cart, and the MOI of the bob respectively.

The variables x and θ are the position of trolley along the belt and angular position of the pendulum. The equations of motion obtained for the system are (2) and (3):

$$F = (M+m)\cdot \acute{x} + b\acute{x} - mlsin\theta \cdot \acute{\theta}^2 + mlcos\theta \cdot \acute{\theta} \tag{2}$$

$$0 = \left(I + ml^2\right)\cdot \acute{\theta} - mglsin\theta + mlcos\theta \cdot \acute{x} \tag{3}$$

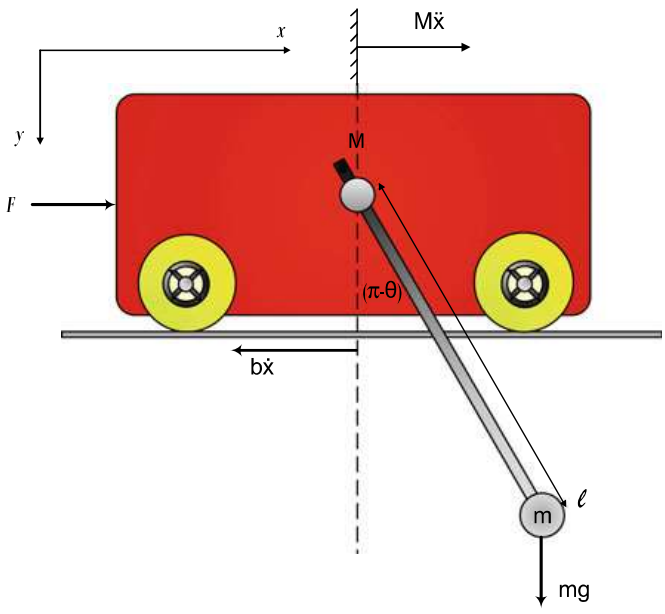


Fig. 1 FBD of gantry crane system

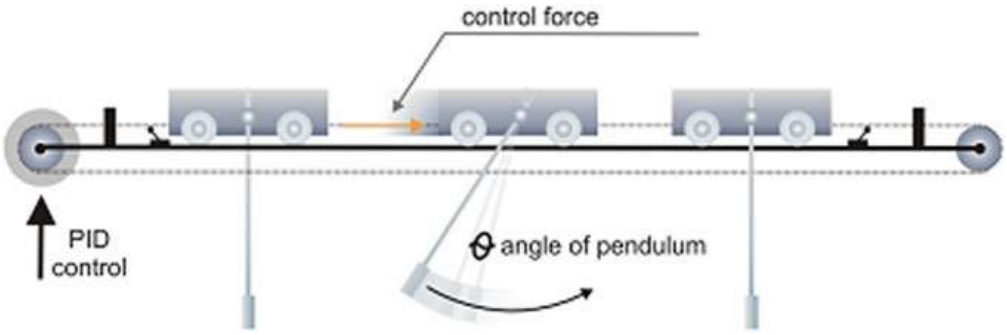


Fig. 2 Functional Diagram

Table 1 Gantry crane system parameters

S. No.	Parameters	Value
1	Mass of the cart (M)	2.4 kg
2	Mass of the pole (m)	0.23 kg
3	Cart friction coefficient (b)	0.05 Ns/m
4	Length of the pole (l)	0.32 m
5	Pendulum moment of inertia (I)	0.099 kg m^2
6	Gravity constant (g)	9.81 ms^{-2}
7	Range of operating voltage of D.C. servo motor (V)	±24 V

For $\theta \rightarrow 0$: $sin\theta \approx \theta$, $cos\theta \approx 1$, $\acute{\theta}^2 \approx 0$

For the simpler understanding of carrying out the stability analysis of the dynamics of the system, the linearized system Eqs. (4) and (5) as

$$F = (M + m) \cdot \acute{x} + b\acute{x} + ml \cdot \acute{\theta} \tag{4}$$

$$0 = \left(I + ml^2\right) \cdot \acute{\theta} - mgl\theta + ml \cdot \acute{x} \tag{5}$$

The system parameters as specified in Eqs. (4) and (5) are given in Table 1.

4 Tuning of ML-PID by FIPSO Algorithm

The PSO algorithm mimics the flocking of birds and the schooling of fishes to find the optimal value of the given target function [9]. The Fully Informed Particle Swarm Optimization takes into account the influence on each particle in a swarm by the mean of its nearest neighbor. Therefore, it has the advantage of fast convergence, escaping

the tendency to fall in local minima and ability to optimize complex functions when compared to Standard PSO algorithm.

The important variables associated with particles in the swarm in FIPSO are x_i (position), v_i (instant velocity), Pb_i (Present Personal best), and Gb_i (Global Position best). Each particle is updated on the basis of above parameters in the above-defined pop size (iteration) for which cost Function (F_{cost}) is to be minimized. For the Gantry Crane Benchmark, x_i can be comprehended as (6)

$$x_i = [K_{P1}, K_{I1}, K_{D1}, K_{P2}, K_{I2}, K_{D2}] \tag{6}$$

where x is the position of the particle, K_{P1}, K_{I1}, K_{D1} are PID constants for the cart's position controller and K_{P2}, K_{I2}, K_{D2} are for the angle controller of the gantry crane system. Particles in a swarm were initialized using the Eq. (7)

$$x_i = unifrnd(PosLL_i, PosUL_i, npop, 1) \tag{7}$$

where $unifrnd$ function generates uniform random numbers in the range ($PosLL_j$,$PosUL_j$). $PosLL$ And $PosUL$ are the lower and upper decision boundary variables respectively, $npop$ is the population (swarm) size and jth represents one of the PID constants shown in above equation.

The updated velocity and position of each particle in an iteration can be calculated as in Eqs. (8) and (9)

$$v_{i+1} = w * v_i + c_1 * (Pb_i - x_i) + c_2 * (Gb_i - x_i) + chi * (x_i - \mathrm{NN}_i) \tag{8}$$

$$x_{i+1} = x_i + v_{i+1} \tag{9}$$

where NN_i is the Nearest Neighbor, defined as the mean of the personal best position of the nearest five particles w is the inertial weight, c_1, c_2 are personal and global learning coefficients respectively and chi is the constriction coefficient. Finally, the innovative Global Best and Personal Best are updated as (10) and (11):

$$Pb_{i+1} = \begin{matrix} \{Pb_i, & Fcost(x_i) > Fcost(Pb_i) \\ x_i, & otherwise\} \end{matrix} \tag{10}$$

$$Gb_{i+1} = \min\{Fcost(Pb_{i+1})\} \tag{11}$$

4.1 Implementation of FIPSO for ML-PID Tuning

The overall FIPSO pseudo code to tune the ML-PID controller for gantry system is summarized as follows:

```
  FCost- call the function whose variables are to be
  tuned
  nVar- no. of variables to be tuned
  NoVar- row vector defining size of the decision
  variable matrix
  PosLL- minimum values of decision variables
  PosUL- maximum value of decision variables
  VelUL-maximum value of velocity
  VelLL- minimum value of velocity
  Setting required PSO parameters:
  Gen- maximum no of iterations to be undergone by the
  code
  PopSize- size of the swarm
  Km- No. of nearest neighbours considered
  NN- Best position closest of neighbours
  Initialisation of swarm parameters
  global best- array of size NoVar containing zeros
  fori=1: PopSize
      initialize position of particle assigning it random
      value;
      set velocity of particle zero;
      evaluate cost function for particle;
      set personal best position and cost equal to
      initial position and cost for each particle;
       if personal best cost of particle i> global best
       cost
          global best=personal best of i;
      end
  end

  Main Loop of PSO
  for it=1: Gen
      fori=1: PopSize
          update velocity using FIPSO innovative vector
          formula;
          apply velocity limits so that velocity remains
          between VelLL and VelUL;
          update position;
      apply position limits so that position remains
      betweenPosLl and PosUL;
          calculate value of cost function;
          if cost of i< personal best cost of i
              personal best position of i= position of i;
              personal best cost of i= cost of i;
          end
          if global best cost < best cost of i
              global best= personal best of i;
          end
      end
end
  apply position limits so that position remains between
  VarMin and VarMax;
  calculate value of cost function;
  if cost of i< personal best cost of i
  personal best position of i= position of i;
  personal best cost of i= cost of i;
  end
  if global best cost < best cost of i
  global best= personal best of i;
  end
  end
end
```

Table 2 ML-PID gain parameters

S. No	Gain co-efficient	Optimal value
1	K_{P1}	4.8590
2	K_{I1}	0.4918
3	K_{D1}	4.0352
4	K_{P2}	4.9101
5	K_{I2}	0.0973
6	K_{D2}	0.3000

The swarm size, iteration, w, c_1, c_2, chi values used in the experiment for the optimization of ML-PID controller for the gantry crane system are 50, 30, 0.7298, 1.4962, 1.4962, 0.7298 respectively. Two main objectives for control of Gantry Crane System, namely, swing minimization of pendulum and tracking of input voltage signal, were achieved using two separate PID controllers. The optimal gain values obtained at 30th iteration with minimum cost function is given in Table 2.

5 Experimental Results

The cart was required to track a sinusoidal trajectory of amplitude 0.3 V and a frequency of 0.1 Hz. A sample time of 100 s was chosen to operate the system. Small manual perturbations of increasing magnitude were given to pendulum for every 20s intervals. Cart Position Response, Pendulum Angle Response, and Control Signal depicting the performance of ML-PID were obtained and validated extensively for stabilizing the Gantry Crane System as shown in Fig. 3. The measure of performance

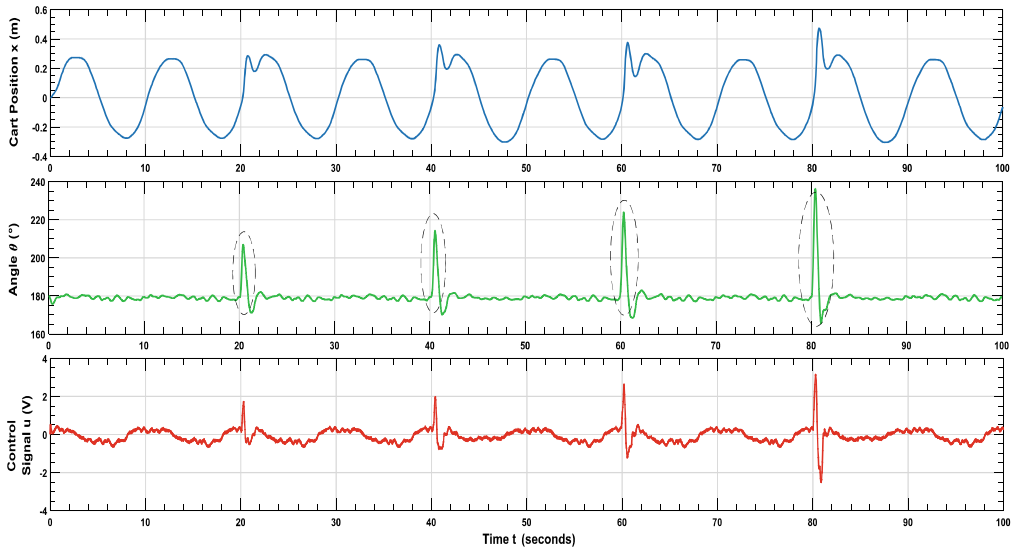


Fig. 3 Performance of Multi-Loop L-PID controller

Table 3 Pendulum angle stabilization response

S. No.	Instant of disturbance initiation (s)	Instant of stabilization (s)	Amplitude of disturbance (°)	Time taken to stabilize (s)
1	20.11	21.57	26.9	1.46
2	40.4	41.7	34.3	1.3
3	59.96	61.58	43.9	1.72
4	80.06	81.67	56.3	1.61

was time taken for stabilization after an external disturbance was given. The instant of stabilization was determined according to the 3% settling time stability condition.

The ML-PID responds to stabilization of the pendulum angle against the external noise/disturbance is given in Table 3. It is observed that as the magnitude of disturbance increases, the time taken for stabilization also increases.

6 Conclusions

In this paper, ML-PID controllers were designed, and the optimal gain parameters are obtained by using the FIPSO algorithm. The controllers were validated to stabilize the benchmarked gantry crane system. It is observed that the ML-PID controller brings the pendulum to the desired position under the perturbations in an average time of 1.522 s.

Acknowledgements The authors would like to thank the Govt. of India and Govt. NCT Delhi for providing funding through Delhi Technological University under TEQIP-II to procure Digital Pendulum System Experimental Setup.

References

1. Astrom, K.J., Hagglund, T.: Benchmark systems for PID control. In: IFAC Proceedings on Digital Control: Past, Present and Future of PlD Control, pp. 165–166 (2000)
2. Gao, Y., et al.: Multi-objective opposition-based learning fully informed particle swarm optimizer with favour ranking. In: Proceedings—2013 IEEE International Conference on Granular Computing, GrC 2013, pp. 114–119 (2013)
3. Jaafar, H.I., Mohamed, Z., Abidin, A.F.Z., Ab Ghani, Z.: PSO-tuned PID controller for a nonlinear gantry crane system. In: IEEE International Conference on Control System, Computing and Engineering, pp. 23–25, Penang, Malaysia (2012)
4. Mendes, R., et al.: The fully informed particle swarm: simpler, maybe better. IEEE Trans. Evol. Comput. **8**(3), 204–210 (2004)
5. Passino, K.M.: Intelligent control for autonomous systems. IEEE Spectr. **32**(6), 55–62 (1995)
6. Poli, R., et al.: Particle swarm optimization an overview. Swarm Intell. **1**(1), 33–57 (2007)

7. Ramli, L., et al.: Control strategies for crane systems: a comprehensive review. Mech. Syst. Signal Process. **95**, 1–23 (2017)
8. Reynoso Meza, G., et al.: Controller tuning by means of multi-objective optimization algorithms: a global tuning framework. IEEE Trans. Control Syst. Technol. **21**(2), 445–458 (2013)
9. Valluru, S.K., et al.: Comparative analysis of PID, NARMA L-2 and PSO tuned PID controllers for nonlinear dynamical system. J. Autom. Syst. Eng. **9**(2), 94–108 (2015)
10. Valluru, S.K.: Introduction to Neural Networks, Fuzzy Logic & Genetic Algorithms. Jaico Publishing House, Mumbai (2011)
11. Valluru, S.K., Singh, M.: Metaheuristic tuning of linear and nonlinear PID controllers to nonlinear mass spring damper system. Int. J. Appl. Eng. Res. **12**(10), 2320–2328 (2017)
12. Valluru, S.K., Singh, M.: Stabilization of nonlinear inverted pendulum system using MOGA and APSO tuned nonlinear PID controller. Cogent Eng. **4**(1), 1–15 (2017)
13. Zhao, J., et al.: Application of particle swarm optimization algorithm on robust PID controller tuning. In: Wang, L., Chen, K., Ong, Y.S. (eds.) Advances in Natural Computation, pp. 948–957. Springer, Berlin, Heidelberg (2005)

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Filter Based Scalable Blockchain for Domestic Internet of Things

Aditya Shyam Bazari
Department of Computer Engineering
Delhi Technological University
New Delhi, India
adityabazari_bt2k16@dtu.ac.in

Aditya Singh
Department of Computer Engineering
Delhi Technological University
New Delhi, India
adityasingh_bt2k16@dtu.ac.in

Abdul Ahad Khan
Department of Computer Engineering
Delhi Technological University
New Delhi, India
abdul_bt2k16@dtu.ac.in

Rajni Jindal
Department of Computer Engineering
Delhi Technological University
New Delhi, India
rajnijindal@dce.ac.in

***Abstract*—With the advancements in technology, the ease of interconnectedness among devices has increased manifold, leading to the widespread usage of Internet of Things. Internet of Things has also reached our homes, often referred to as domestic Internet of Things. However, the security aspect of domestic Internet of Things has largely been under question as the increase in inter-device communication renders the system more vulnerable to adversaries. Largely popular blockchain technology is being extensively researched for integration into the Internet of Things framework in order to improve the security aspect of the framework. Blockchain, being a cryptographically linked set of data, has a few barriers which prevent it from being successfully integrated to Internet of Things. One of the major barrier is the high computational requirements and time latency associated with it. This work tries to address this research gap and proposes a novel scalable blockchain optimization for domestic Internet of Things. The proposed blockchain model uses a flow based filtering technique as an added security layer to facilitate the scenario. This work then evaluates the performance of the proposed model in various scenarios and compares it with that of traditional blockchain. The work presents a largely encompassing evaluation, explanation and assessment of the proposed model.**

***Index Terms*—blockchain, internet of things, flow policies, IoT security, scalable blockchain**

I. Introduction

Blockchain, with the potential to revolutionize modern technological communication, is being rapidly amalgamated into a number of modern architectures such as smart-grid [1], healthcare [2], fintech [3] etc. By primarily providing transparency, security and decentralization to the host architecture, blockchain improves the overall functioning of the network. One such framework in which the applicability of Blockchain is being optimistically considered is the Internet of Things [4]. This paper addresses the scenario of Blockchain in domestically deployed Internet of Things and proposes a feasible novel blockchain communication methodology.

Traditionally, domestic Internet of Things have comprised of a simple permission-based security model, in which, after a user grants the IoT system permission to use a particular data, it can use the data in every possible way [5]. No further insights are usually provided into how the data is being used or the online sinks for data transfer. Such a scenario could possibly result in a massive security breach, compromising a user's private data, such as security camera footage, thermostat readings, schedule notes etc. Such a scenario necessitates additional security mechanisms to secure domestic Internet of Things [6].

Since the popularization of bitcoin and the conceptualization of blockchain technology, extensive research is underway to effectively integrate blockchain communication to Internet of Things [4]. This is due to the possibility that blockchain can potentially improve the communication nuances involved in Internet of Things [20]. Cryptographic links between blocks of data, in blockchain, impart enhanced integrity and immutability [7] to the system, thus adding on to the security layer. The distributed framework of blockchain provide decentralization [7] to the system, thus avoiding single points of failure. Blockchain works in a transparent manner [8], with every transactional record present throughout the system, thus suitable for untrusted environments. However, there are certain hurdles to efficient deployment of blockchain in Internet of Things, which translate to the increased resource consumption and added latency in the system. Domestic IoT devices, often being compact, with certain specialized tasks, do not contain adequate computational power to sustain cryptographic mining, necessitated in traditional blockchain [9]. As the cryptographic hashes grow more and more complex, the temporal aspect of the transactional processes rise, thus causing a large delay and forcing the system to lag. Addressing these issues, this paper proposes a modification for blockchain, in context of domestic Internet of Things, which outperforms the existing blockchain systems in terms of resource consumption and latency, while preserving the security aspects of the system. A detailed analysis of the comparison of performance is also

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:47:47 UTC from IEEE Xplore. Restrictions apply.

presented to illustrate the improvement achieved.

The remainder of the paper is organized as follows: Section II explains the background, Section III presents the proposed approach and Section IV reports the results. Conclusions are presented in Section V.

II. Background

Internet of Things is an interconnected network of devices communicating with one another for the exchange of some sort of information. Various devices can be part of the network such as smartphones, sensors, hubs, etc. These devices exchange information with each other to complete a particular task. Information is transferred, processed and analysed among devices in the network. It is not a necessity that the Internet of Things is always a part of the Internet as the name suggests. A few of the major components which define the overall functioning of the infrastructure are:

- Devices: Connecting devices are the basic building blocks of Internet of Things. These devices can be sensors or actuators. Different devices such as mobile camera, accelerometer, proximity detector, pressure sensors, etc can be the part of IoT.
- Connection: The basic blocks, devices in Internet of Things are connected with each other and with the Internet via some means. They may be connected through standard Wi-Fi, Bluetooth, wired LAN (Local Area Network).
- Data Processing: Data Processing is done by the software. It can be identifying a person using a camera or measuring the temperature of the room. The software processes data and gives appropriate results to the user.
- User Interface: User Interface is responsible to show the results processed by the applications to the user in a very suitable manner. It can be showing continuous results over a mobile application or giving an alert via notification on mobile phone to the user.

The process begins with the sensors and chips that are embedded in the physical things that surround us, each of which transmits valuable data. All these devices having sensors are connected with one another. There are IoT platforms which store the relevant data transmitted by the connected devices. IoT platform analyses the data received from various devices. After analysis, important data is transferred to the real time applications. The application will then show the results of analysis to the user. In this way, IoT helps us to interact with machines and devices.

With sophistication of technology, security has become a cardinal issue in modern day Internet of Things [9]. The devices in IoT are connected with each other and the Internet. The data exchange taking place between these devices over the network must be safe and secure. The inter-connected nature of this setup makes it exceptionally prone to attacks such as:

- Cryptanalysis Attacks [10]: In cryptanalysis attacks, the focus is to find the encryption key and to break the encryption. It refers to the act of breaking ciphers. The attacker will study and analyse the ciphers to break the encryption and will break into the system.
- Software Attacks [11]: In software attacks, viruses, trojan horse programs or worms are used to inject malicious code into the system. The malicious code will then destroy the software. An attack on software of IoT device may lead the attacker to steal the user's personal credentials.
- Network Attacks [11]: IoT is vulnerable to network attacks because the transmission of data takes place in a broadcast manner. As soon as an attacker hacks a device in the network, he can easily break into other devices present in the same network. Hence, it is important to keep the network safe. Network attacks can be active or passive.
  - Passive Attack: In passive network attacks, attacker attacks the functioning components of the network. The attacker attacks the network in a concealed manner and try to gather important information of the user. Traffic analysis, node destruction, node malfunctioning, node outage can be some passive attacks.
  - Active Attack: In active network attacks, the attacker tries to make the network services complex and destroy the function and operation of the network. Jamming, flooding, Denial of Service (DoS), wormhole, blackhole, sinkhole are some active attacks.
- Denial-of-Service (DoS) Attack [12]: In DoS attack, the attacker makes the IoT devices and services unavailable. It is very difficult to detect DoS attacks. The attacker makes unnecessary requests to the services which reduce the capacity of the network. The attacker generally tries to flood the centralized resources with requests such that it is unavailable for the other nodes in the network. In Distributed Denial-of-Service (DDoS) Attack, multiple attackers from multiple sources try to flood the requests to a single resource. It can neither detect the attack beforehand nor can stop it by just blocking a single source as it is originated by multiple sources.

Blockchain is an inter-connected list of time-stamped series of secure and immutable records which is distributed in nature [14]. As highlighted in the previous section, integration of blockchain communication into IoT infrastructures effectively enhances the security aspect of IoT infrastructures and protects the system from the above-mentioned threats [4]. Each and every block in this list or chain uses cryptographic methods to stay connected. All these blocks together provide security to each other and also prevents any form of modification to the data. This chain has no central authority. Due to absence of any central power, this distributed ledger is open to each and everyone to see and add. It is by its very nature a transparent entity and every node involved in it is accountable for its actions. The blockchain is a simple yet ingenious way of passing information from system X to system Y in a highly safe and secure manner. One party creates an initialization

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:47:47 UTC from IEEE Xplore. Restrictions apply.

block, it is then verified by millions of other distributed systems [13]. After the verification, the block is added to the chain which consists of previously added blocks.

The three primary factors that impart this high security to the blockchain setup are:

- Decentralization [7]: The system follows a decentralized computational approach in which every involved entity performs some computation locally. This system eliminates a single point of failure and increases the efficiency.
- Transparency [8]: The system is transparent as the blockchain contains every record in an encrypted manner and a record of it is present in each node. This increases the accountability in the system and enhances the security.
- Immutability [7]: Immutability of this system ensures that the encapsulated data is not tampered with. The hashing ensures that any change in the underlying data renders the entire blockchain invalid.

Despite the benefits of amalgamating blockchain to IoT, the integration is not usually feasible in real world scenarios [19]. This is due to the high latency and computational constraints that accompany the setup [9]. This work, in the following section, attempts to address this research gap and proposes an optimized filtering technique to enhance blockchain performance in domestic Internet of Things.

III. Proposed Model

Modern day Internet of Things are prone to security breaches as the devices store a huge variety of crucial personal and institutional data [9]. Several attacks, such as DDoS (Denial of Service), Side Channel Attacks or Man in the Middle attack can cause immense harm by breach of data or rendering the devices useless. For instance, in 2016, Dyn, an American DNS provider faced a large scale DDoS attack, which originated from millions of IP addresses. Several of these IP addresses belonged to IoT devices such as sensors, printers, routers, monitors etc which were infected with Mirai Malware [15]. Similar circumstances necessitate the improvement of security mechanisms in these devices. Integration of blockchain technology to the Internet of Things may provide a solution, as explained in the earlier sections. However, a major obstacle in the adoption of blockchain by IoT devices is the high computational requirements of blockchain. IoT devices are usually small, with limited resources and computation at their disposal. This work tries to close this gap in the case of domestic Internet of Things by providing a case specific solution.

In traditional blockchain scenarios which have been integrated in the Internet of Things, every transaction initiated in the system, whether genuine or not, is being competed for by miners to validate [14]. The process of validation of a transaction is extremely bulky in nature and thus, utilizes a huge chunk of the computational power, thereby, increasing the processing time and resource usage [16]. This work designs a preliminary selective mechanism to identify the predicted authenticity of a transaction before the mining initiates. This mechanism eliminates a set of transactions before the mining process. The blockchain model uses the selective mechanism for the filtering of transactions, thus eliminating the need for validating a number of transactions which do not fit the selection criteria. This subsequently saves computational time of the system and thus, aids the scalability of blockchain in domestic Internet of Things.

The selectivity of transactions in the proposed blockchain model is based on the idea of flow policies [17] [18]. Flow Policies, in simple words are the underlying principles in a system that govern the flow of data. In domestic Internet of Things, the flow pattern of data files between devices are fairly static in nature, in the sense that any particular device has a predefined set of data flow paths. Under normal circumstances, the devices do not usually defy these predefined pathways and almost always confirm to them. For instance, a thermostat might send its data to a hub processor in the network, but it will never have to send the collected data to a bitmap image processor in the same network under normal circumstances. Similarly, a security camera might send a jpg file to a bitmap processor in the network, but might never have to send it directly to an untrusted online sink, while securely functioning. These predefined path structures in domestic Internet of Things have been researched previously and have also been successfully implemented in traditional security architectures. This paper leverages this attribute of domestic Internet of Things and amalgamates it with blockchain to improve the security architecture of the system, while maintaining the scalability of the associated blockchain framework in terms of time and resource consumption.

Every transaction necessitates the involvement of at least two devices. In the presented model, every device has an associated list of flow policies, defining which devices can send or receive data from them. In order to identify if the transaction is in accordance with the user defined flow policies, an intersection is performed between the associated flow policies of the involved devices and the resulting set of flow policies are used to evaluate the authenticity.

In the presented model, the user defines a set of flow rules for every device in the network, i.e. which devices can send in or receive data from that particular device. Every transaction, upon initiation by a sender is flooded into the blockchain network. In the network, every miner has the flow policies of the network. Upon receiving a transactional data packet, duly stamped with the sender's and receiver's IPs, the miner first checks the fulfilment of the associated flow policies. If the transaction is in accordance to the defined flow policies of both the sender and the receiver, the miner proceeds to the validation process. The miners then compete to validate the transaction and mine a block. This block is then added to the blockchain and the transaction is complete. If the transaction violates the flow policies, the data packet is dropped and the transaction cancelled. The flow policies in this model, thus, act like a primary security layer to filter out invalid transactions which do not follow the rudimentary flow patterns acceptable to the user. For a flow policy violation, no major computation is performed in this model, saving a certain

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:47:47 UTC from IEEE Xplore. Restrictions apply.

amount of computational power and time. In scenarios when the system is under attack, this filter thus considerably reduces the chances of the system being overwhelmed by adversaries generating fake requests in the system.

In real world scenarios, under normal circumstances, devices usually function without any interference, following those user-defined paths. During an attack by an adversary, a burst of faulty transactions of data are initiated with the intention of interrupting the functioning of the model. If processing those faulty transactions consume too much of the system resources or time, the system starts to fall apart and eventually, fails. The proposed model protects an IoT establishment from such scenarios in which a system is under constant attack by a particular device or application, initiating multiple bogus transactions to create bottlenecks in the system and slow it down, ultimately resulting in a system failure.

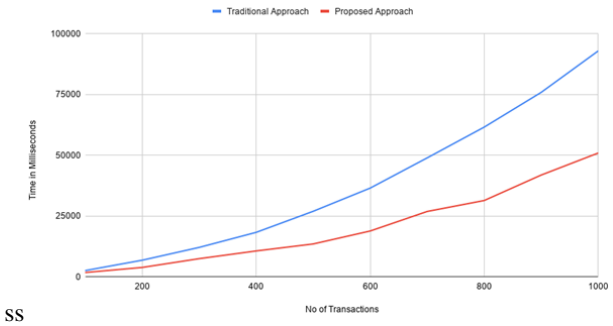
IV. Results

A simulation environment is developed to conduct these runs and only generalized average scenarios are taken into consideration in order to account for the usual trend observed. However, as pointed out earlier, in real world scenarios, the occurrence of fraudulent transactions are often concentrated in short periods of time in which a malicious entity attempts to impair the system. In usual circumstances, the system functions without the presence of any malicious entities without any fraudulent transactions. This work presents scenarios encompassing all possible scenarios.

The first set of simulations are performed on a randomized set of data transactions in the underlying IoT ecosystem in which the flow policies are defined in a probabilistic manner. The flow policy for every pair of device has a defined probability which determines if that particular transactional flow will be valid or not. In real world scenarios, the user directly interacts with the system to explicitly state the flow policies. Thus, the probabilistic method used in this work is merely an attempt to simulate the real world user interaction. The system then generates completely randomized data transactions in the IoT simulation. Blockchain Nodes are subsequently invoked to initiate these transactions. Depending upon the percentage of faulty transactions, the performance evaluation changes.

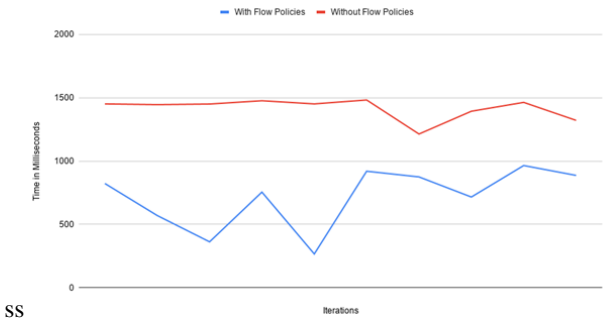
In the particular case of Fig. 1, the probability for every pair of node was set to be 0.5 and since the transactions initiated were randomized, the scenario translated to almost 50% of the transactions being faulty. As visible in the graph presented, the proposed method significantly outperforms the existing approach as almost half the transactions are invalidated by the flow policies, thus, saving the time otherwise required in mining. The presented graph represents an averaged out scenario of multiple iterations of simulation, with data values being reset at every iteration.

The “difficulty” in these blockchain simulations represent a quantized value of the string length of the initial cryptographic challenge which is required to be solved in order to block a mine and validate transactions. In simple words, “difficulty” of the blockchain system is representative of the difficulty of the



SS

Fig. 1. Comparison of traditional and proposed approach



SS

Fig. 2. Difficulty of encryption: 2

cryptographic mining. While for cryptocurrencies, this number is very large, in the case of domestic Internet of Things, this number is relatively smaller. This is due to the disparity in use case, security requirements and time constraints.

The simulations work on “difficulties” ranging from 2 to 5. In other words the length of the initial cryptographic challenge is varied from 2 to 5 in order to give an idea of how this factor affects the time considerations of the implementation. For various difficulty levels, graphs are presented portraying the changes in timing between traditional blockchain approach and the presented approach enhanced with flow policies. It can be seen in Fig. 2, Fig. 3, Fig. 4 and Fig. 5 that as the difficulty in the blockchain method is increased the associated time of transactions also increase. It is evident that the proposed approach provides a huge temporal benefit in optimizing the operation of the Internet of Things Infrastructure of domestic use case scenario. For difficulty levels of 2, 3, 4 and 5, the performance of all the iterations are clearly better for the proposed approach, as compared to the traditional blockchain approach.

This work also encompasses the evaluation of the proposed method in scenarios with different percentages of invalid transactions. Fig. 6 represents the performance of the model in the above-mentioned scenario. It can be observed that as the percentage of invalid transactions increases, the time consumption of the system reduces. This is due to the fact that invalid transactions are rejected in the filtering phase and are thus, not mined at all. This saves the computational time otherwise wasted in their validation. However, even when

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:47:47 UTC from IEEE Xplore. Restrictions apply.

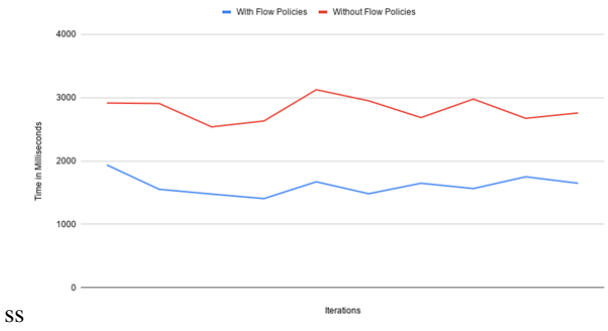


Fig. 3. Difficulty of encryption: 3

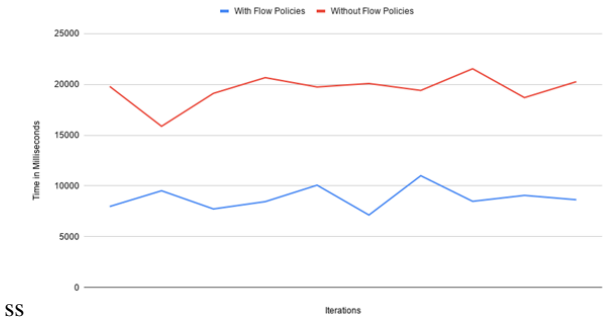


Fig. 4. Difficulty of encryption: 4

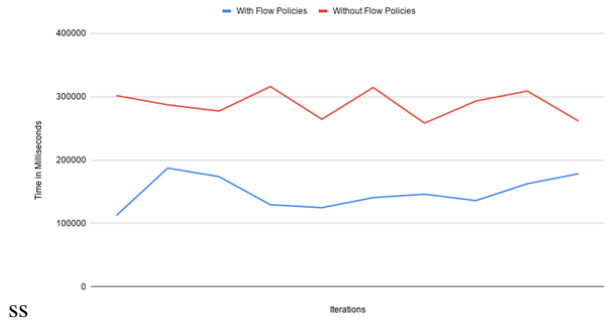


Fig. 5. Difficulty of encryption: 5

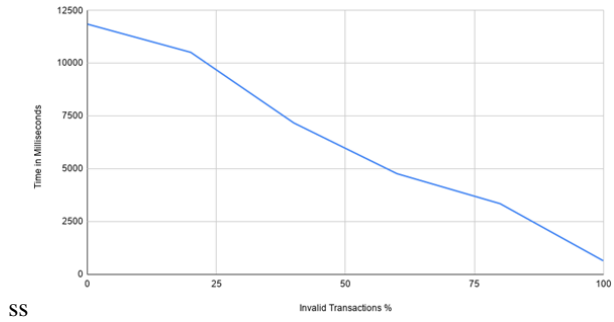


Fig. 6. Variability in Percentage of Invalid Transactions

100% of the transactions are invalid, The time does not reduce to 0. This is because some computation is still required by the system to run processes other than mining. When 0% of the transactions in the system are invalid, the system requires a lot of time as all the transactions are mined as usual and no apparent optimization is achieved.

V. Conclusion

This work navigates through the blockchain mechanism of operation and analyses the scope of integrability to domestic Internet of Things (IoTs). While navigating through the use case scenario of modern Blockchain Technology, the work considers as to how blockchain is suitable for domestic Internet of Things and the benefits that the integration of blockchain provides by enhancing the security and reliability of blockchain by imparting decentralization, immutability and transparency.

One of the major barriers in scaling blockchain into domestic IoTs is the immense latency which comes along with the blockchain technology. Primarily, the Proof of Work (PoW) consensus algorithm associated with blockchain is responsible for the latency caused as the cryptographic challenge presented before miners are time consuming. However, it is also crucial to impart security to the process of validation of transactions. Intensive and inclusive literature review and research led us to an optimizing technique, which when integrated to traditional blockchain led to a significant decline in time considerations. This technique requires installation of entity specific traffic flow policies which are defined for each entity in the IoT infrastructure. This optimization is very specific to the use case of domestic Internet of Things in which data flow paths are relatively static. In other scenarios, such as cryptocurrency, this is not viable owing to the variable data flow paths and the unpredictable nature of transactional packet flow.

This work then conducts an analytical simulation in order to assess the performance metrics by varying the user specific parameters. A few parameters such as percentage of malicious transactions, “difficulty” of the mining and overall number of transactions are varied and the corresponding data sets are used to obtain performance graphs. These graphs indicate the performance improvement achieved and discusses the implication of this improvement in real world scenarios.

References

[1] Mengelkamp, Esther, et al. ”A blockchain-based smart grid: towards sustainable local energy markets.” Computer Science-Research and Development 33.1-2 (2018): 207-214.
[2] Mettler, Matthias. ”Blockchain technology in healthcare: The revolution starts here.” 2016 IEEE 18th international conference on e-health networking, applications and services (Healthcom). IEEE, 2016.
[3] Eyal, Ittay. ”Blockchain technology: Transforming libertarian cryptocurrency dreams to finance and banking realities.” Computer 50.9 (2017): 38-49.
[4] Samaniego, Mayra, and Ralph Deters. ”Blockchain as a Service for IoT.” 2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). IEEE, 2016.
[5] Fernandes, Earlence, et al. ”Flowfence: Practical data protection for emerging iot application frameworks.” 25th USENIX Security Symposium (USENIX Security 16). 2016.
[6] Lin, Huichen, and Neil W. Bergmann. ”IoT privacy and security challenges for smart home environments.” Information 7.3 (2016): 44.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:47:47 UTC from IEEE Xplore. Restrictions apply.

[7] Puthal, Deepak, et al. "The blockchain as a decentralized security framework [future directions]." IEEE Consumer Electronics Magazine 7.2 (2018): 18-21.
[8] Jeppsson, André, and Oskar Olsson. "Blockchains as a solution for traceability and transparency." (2017).
[9] Khan, Minhaj Ahmad, and Khaled Salah. "IoT security: Review, blockchain solutions, and open challenges." Future Generation Computer Systems 82 (2018): 395-411.
[10] Andrea, Ioannis, Chrysostomos Chrysostomou, and George Hadjichristofi. "Internet of Things: Security vulnerabilities and challenges." 2015 IEEE Symposium on Computers and Communication (ISCC). IEEE, 2015.
[11] Deogirikar, Jyoti, and Amarsinh Vidhate. "Security attacks in IoT: A survey." 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC). IEEE, 2017.
[12] Kolias, Constantinos, et al. "DDoS in the IoT: Mirai and other botnets." Computer 50.7 (2017): 80-84.
[13] Zheng, Zibin, et al. "Blockchain challenges and opportunities: A survey." International Journal of Web and Grid Services 14.4 (2018): 352-375.
[14] Bazari, Aditya Shyam, et al. "Node Criticality Assessment in a Blockchain Network." Proceedings of the 2nd Workshop on Blockchain-enabled Networked Sensor. 2019.
[15] Antonakakis, Manos, et al. "Understanding the mirai botnet." 26th USENIX Security Symposium (USENIX Security 17). 2017.
[16] Dorri, Ali, Salil S. Kanhere, and Raja Jurdak. "Towards an optimized blockchain for IoT." 2017 IEEE/ACM Second International Conference on Internet-of-Things Design and Implementation (IoTDI). IEEE, 2017.
[17] Fernandes, Earlence, et al. "Flowfence: Practical data protection for emerging iot application frameworks." 25th USENIX Security Symposium (USENIX Security 16). 2016.
[18] Bull, Peter, et al. "Flow based security for IoT devices using an SDN gateway." 2016 IEEE 4th International Conference on Future Internet of Things and Cloud (FiCloud). IEEE, 2016.
[19] Samaniego, Mayra, and Ralph Deters. "Blockchain as a Service for IoT." 2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). IEEE, 2016.
[20] Suma V. SECURITY AND PRIVACY MECHANISM USING BLOCKCHAIN. Journal of Ubiquitous Computing and Communication Technologies (UCCT). 2019;1(01):45-54.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:47:47 UTC from IEEE Xplore. Restrictions apply.

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 2, ISSN No. 2455-2143, Pages 217-221
Published Online June 2020 in IJEAST (http://www.ijeast.com)



FLEXIBLE SOLAR POWER TREE - A CONCEPT

Mukesh Kumar
Physics Department SSNC
University of Delhi, INDIA

J P Kesari
Mechanical department
Delhi Technological University

Lokesh Rana
Physics & Astrophysics Department
University of Delhi, INDIA

Abstract— **Sunlight is always a great asset for human beings. Among various systems available to access solar power, the most well known innovation is the photovoltaic(PV) transformation of daylight into electricity. Regardless of its benefits, sunlight based PV innovation has issues such as bulky and less efficiency. The most precious process of photosynthesis, the sun catching nature of trees is an inventive thought of light catching cells. The idea of a sun oriented flexible tree is fit for tending to these issues adequately with polish. In this paper efforts have been made to audit the parts of the sun based tree and its energy harvesting capability. The different energy enhancement structures are additionally talked about alongside utilizations of the sun powered tree. The paper additionally addresses the difficulties engaged with this innovation and proposes future research heading.**

Keywords— **Photovoltaic, Solar tree, Solar Energy, Renewable Energy**

I. INTRODUCTION

The journey for green and supportable energy sources has gotten probably the greatest test for our time, because of exhaustion of fossil energy products, environmental challenges, an Earth-wide temperature boost and perpetually developing energy demands. Energy from the sun comprises light and warmth from the Sun, it is being utilized via different advancing technologies. Few of them are like sunlight based heat collectors, sun oriented photovoltaic, sun based heat power, sun based electricity powerhouse, sun-powered building design and counterfeit photosynthesis [1].

The most well-known use of the sun is electric energy generation through the photovoltaic (PV) frameworks. Sun based PV execution is reliant on the nearby climatic conditions and accessibility of sun-powered radiation. Sun based radiation appraisal and estimation is particularly useful for the appropriate structure of sun based energy change frameworks. The trap frequency and intensity of the Sun beams vacillate for the duration of the day and throughout the year. In this manner, sun-powered modules fixed at a specific point may not be completely advanced. Sun following PV frameworks can be structured, however they essentially raise the all-out expense of energy generation as they are exorbitant and require support. Another, encouraging approach to coordinate sun oriented PV into the developed condition is through multidimensional Photovoltaic frameworks [6-8]. One of the essential installation components for the establishment of all such sun based PV modules is land Rooftop sun oriented PV is an alluring choice, however, it has constrained space for modules to be masterminded in a cluster or multistoried building particularly inland limited urban territories. The Sun following PV frameworks can be a possible solution but they fundamentally raise the all out expense of energy as they are exorbitant and require support. Another challenge in this encouraging solar powered PV execution deals with issues like installation space requirement, low PV efficiency, approach to coordinate sun based PV frameworks etc.

Significant cost reductions have been observed in the past few decades along with high efficiency in photovoltaic based solar cells with strong increase in people acceptability. The PV technology now positioned to provide low-cost energy options and have enormous capacity for future energy generation. Improvement in energy-conversion efficiency reduces such costs, with mass manufacturing likely to bring down the additional costs associated for achieving higher efficiencies. This suggests that the industries will evolve beyond the conventional single-junction solar cells that presently dominate commercial production, where energy-conversion efficiencies are constrained by practical values of Shockley–Queisser limits to below 30%. Contrary to that nature's solar capturing technology, plants exploit the Sun more effectively [3]. The low effectiveness is due to associated losses like spectrum loss, reflection loss and heat generation and so forth.

In this article the overall prospects for a range of approaches that can potentially exceed these limits are based on ultimate efficiency prospects, material requirements and installations outlook. The challenges involved in this technique are considered and suggestions are proposed to overcome them. Finally the concept of solar trees is discussed to overcome above losses through law of conservation and minimization process.

II. SIMILARITIES BETWEEN SOLAR CELL AND PHOTOSYNTHESIS

Like plants, solar cells gather their energy from sunlight. Plant leaves collect sunlight and transform this energy into



chemical energy. Similarly, photovoltaic solar cells convert the sun's energy into electricity. They perform the same function but in different senses. The main resemblance between the two is that both photovoltaic and photosynthetic plants are supposed to absorb photons.

This task of absorption of sunlight is carried out by a molecule called chlorophyll in nature (plants). In the case of photovoltaic, scientists have designed materials where the electrons can absorb the energy of the sun's photons. In both cases, photons are absorbed by electrons, which take up the extra energy and get excited to a higher energy state known as excited electron. The process of conversion of the energy absorbed by excited electrons should be fast before they lose their energy and go back to their initial state. In the case of photosynthesis, this issue is resolved by moving the electron from one molecule to another until it finally settles in a molecule that can store this energy for a long time. In photovoltaic, the excited electrons are carried off into a circuit, where they are either used to operate directly or are routed into a battery for storage [4].

There is a type of solar cell that is designed to be as analogous to photosynthesis. This is a non-conventional photovoltaic cell that tries to reproduce the way photosynthesis works. Instead of moving an electron as quickly as possible through a crystal of identical atoms, this dye-sensitized solar cell absorbs energy in a dye molecule, then transfers the excited electron into another material located adjacent to the dye molecule. This technique makes sure that the electron doesn’t lose its energy ineffectively. When connected to a circuit, this electron makes its way through the second material without risking losing its energy.

III. PLANT ARCHITECTURE INSPIRE SOLAR CELL DESIGN

Plants exhibit a variety of systematic structures and structures, which are difficult to explain. Plants are dependent on the light energy they receive from photosynthesis, and strong sunlight is the most important factor in photosynthetic influence, but in addition the leaf temperature, leaf water status, and many other physiological procedures of the plant also influence photosynthesis. There are no fixed configuration rules in all conditions.

Geometrically it is determined by only two parameters: the propagation limit and the branch length. They are undoubtedly more cohesive and consistently attracted to authors of plant conditions and development models. In trees, the leaves outside the hiding place are different from those inside the tree cover. The outer leaves are called "sun leaves," and the inner leaves are "secret leaves." These leaves contain differences such as fiddle, internal structures of life, and science that enable the translation of certain energy to use different types of solar radiation more effectively. The leaves of the sun are generally normal, prolonged and larger than the secreted leaves, which contain many layers of chlorine-containing tissue and are very broad within the arteries. It is believed that the sun's leaves are better adapted to capture and use direct sunlight (if not too different to even think of causing heat and pressure related damage). Its long position is connected in parallel with the most important levels of sunlight reaching the inner cover where the foliage is located. Shaded leaves appear to make good use of sunlight, reaching into the inner shelter after exposure to different substances, such as outer leaves, in the form of direct light. Leaves of similar color can be found in an outdoor shelter such as the color towards the subtle sun. The leaf level of photosynthesis depends on the biomass of the leaf biomass as the emission of light into the leaf area. The scattering and course of leaf functioning within a tree can completely alter the effectiveness of a good harvest. The leaf area all tightened to a given plant is a very important aid that contributes to the partitioning of edible radiation. When the point from the light of the sun decreases from the equator to the highest positions, the trees of different lengths and widths have effectively changed the light capture technique. All kinds in a variety of colors to add an effort to capture a quick glow through the whole tree.

The leaf must "absorb" sunlight, water from the inside and carbon dioxide from the air that combines photosynthesis. Leaf photosynthesis, meanwhile, is influenced by these carefully regulated materials: enough sunlight and carbon dioxide for photosynthesis, yet not all that closely related heat storage or water misery. The leaves found on the Nature Trail trees are expensive and adequate or molded with a needle that is ideal for their water-conserving weather conditions and high maintenance during the day is an important way to persevere.

Plants show a striking assorted variety of structures and structures, which are hard to decipher. Plants rely upon the light energy that they catch by photosynthesis, and sun-powered radiation is the significant main impetus influencing photosynthetic movement, yet in addition leaf temperature, leaf water status, and numerous others. because nobody configuration rules in all conditions. The geometrically dictated by only two parameters: spreading edge and branch length. are undoubtedly multi-fractals and have consistently pulled in mathematicians for plant shape and development models. In trees, the leaves on the outside of the tree shelter contrast from those inside the tree covering. The outside leaves are alluded to as "sun leaves," while the inside leaves are "conceal leaves." These leaves have contrasts fit as a fiddle, interior life structures, and science that make an interpretation of specific capacities to utilize various types of sun based radiation successfully. Sun leaves are normally littler, progressively prolong, and thicker than concealed leaves, with more layers of chlorophyll-containing tissues and increasingly broad inside vascular frameworks. It is believed that sun leaves are better adjusted to catch and utilize direct



sunlight based radiation (when it isn't too exceptional to even think about causing heat and different pressure related harm). Their lengthened shape is likewise connected with more significant levels of sunlight based radiation arriving at the inward covering where the shade leaves are found. Shade leaves appear to successfully utilize diffuse sun oriented radiation, which arrives at the internal shelter in the wake of having been dissipated by different items, similar to the external sun leaves, in the way of direct light. Shade leaves can likewise be found on the outside shelter as an afterthought that faces from the overarching sun. The leaf photosynthesis rate relies upon the biochemical limits of the leaf just as a light appropriation on the leaf surface. The dispersion and course of action of leaves inside a tree can firmly change the light reaping effectiveness. The all-out leaf zone bolstered by a given plant is the most essential auxiliary property that influences the division of consumed radiation. As the sunlight based tendency point diminishes from the equator to higher scopes, trees with contrasting stature to width proportion have inalienably changing efficiencies of light capture. Every single such variety to augment the immediate light capture attempt by the whole tree.

A leaf must "catch" sunlight, water from inside and carbon dioxide from the encompassing air by means of pores for photosynthesis. The leaf photosynthesis, at that point, is influenced by these exercises in careful control: enough daylight and carbon dioxide to run photosynthesis, yet not all that much-related warmth retention or water misfortune. The leaves found on the trees of the Nature Trail are either expensive and level or needle-moulded which appropriate to its climatic conditions for cautious preservation of water and most extreme daylight retention is a significant endurance system. A significant result of variety in a leaf design, its shape and size is the general light gathering and the proficiency factor for photosynthesis.

IV. SOLAR POWER TREE

Nature motivation centres around profoundly requested little bent surface geometry which follows the most extreme snare of daylight and water protection through photosynthesis. Likewise between a sun oriented cell and a tree is that both attempt to catch the most extreme sunlight based energy with least water misfortune or warmth age. At present utilization of PV, frameworks increment quickly because of ceaseless decrease in costs of sun based cells. Still, there are a few obstacles for the appropriation of this innovation like low productivity and enormous space necessities. Different methodologies [6] have been done to improve sun oriented energy factors, for example, low ingestion coefficient, heat loss, reflection coefficient and so forth, The charge bearer age and conduction impact the light conversion proficiency of a sunlight based cell. This is a roundabout way that relies upon the quantity of sun-powered cells introduced and their energy change limit.

Recently inspired by natural heat-trapping technology - photosynthesis, a new concept solar tree is introduced by many researchers [3,4]. Nature centres around profoundly requested little bent surface geometry like tree geometry. So a gathering of cells which are littler might be prolonged, a little bent and have a multilayer structure. The new idea- Solar Tree is a natural venture, an environmental model, a counterfeit sun based structure that seems as though sculptural trees exist from little scope (size of a bonsai tree) to enormous scope (about the size of a breeze turbine) power plant. In sun based trees, PV boards or cells are orchestrated in a Fibonacci arrangement design rather than leaves. It is an independent unit that produces green energy and provides a place of comfort and energy for a wide variety of services.

The structure is a ground-mounted solar system with a pole that supports many individual panels up in the air. The performance of sunlight based trees vary and they have been intended to give the capacity to face various situations. Sunlight based tree has remarkable properties as far as structure and multi-edge direction parameters. By utilizing the tallness parameter, the sun-powered tree requires less space utilization which can lessen the installation cost. Such a mix of the sun oriented cell should be possible on common trees to lessen the expense of mounting. Contrasted with the customary sun based PV homestead which is situated a solitary way, the multi-edge direction parameter from the sun based tree boards yields the potential ability to retain higher daylight force prompting higher yield vitality. Simultaneously, this parameter disposes of the need to introduce a sun oriented tracker which can decrease the operational expense. To build the assortment Generation of 1MW force from PV module framework requires the space of around 5-6 Acres for lodging the boards as it whereas a solar tree would take just 1% of the land territory in contrast with general PV lodging. So it is appropriate for places where there is a shortage of land. It can likewise be bolted at any situation to withstand the wind pressure because of substantial tempest influencing over the fundamental shaft/trunk. The boards will be normally looking with least block from different towards the sun at an edge as required so they can gather the greatest sun-powered energy in a daytime.

Sun-powered trees are actually a functional answer for household use. In this setting utilization of PV innovation for the local necessity as the sun oriented tree is the acceptable option when contrasted with traditional level or housetop mounting PV panels. Solar trees can be introduced close, before the house or on a porch, where there is no concealing consistently. To build assortment efficiency, the nanowire cell can be used rather than present PV cells. The Nanowire cells can gather multiple times progressively sun based vitality

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 2, ISSN No. 2455-2143, Pages 217-221
Published Online June 2020 in IJEAST (http://www.ijeast.com)

when contrasted with regular cells. The idea of nano leaf is a plan to expand the different types of energies, for example, sunlight based, warm and piezoelectric force by stalks, branches, and twigs of a tree.

Solar trees are really a practical solution for domestic use . In this context use of Photovoltaic technology for the domestic applications in the form of the solar tree is the good alternative when compared with conventional rooftop mountings. Solar trees can be installed in front of the houses, at the terrace, in streets or where there is no shading throughout the day.

V. Minimization of Solar Energy Loss

Sunlight increases temperature which affects photovoltaic conversion. This does not significantly affect the light generated current of the cell but reverse saturation current increases exponentially with temperature and also reduces open circuit voltage with temperature. Furthermore, it has been experimentally observed that temperature drops the efficiency of solar cell. It is imperative to reduce the heat build-up in the cell in order to improve conversion efficiencies. Recently, it is seen that solar tree energy conversion efficiency increases through minimizing the chlorophyll antennae [6]. Similarly to keep the heating impact of a solar cell low , various approaches are possible. The approaches are possible such as use of multispectral cells for low heat production, reduction in energy loss through current loss and the use of thermal generators for energy harvesting. The reduction in size expected a reduction in current with less utilization time. Recent quantum level technology developed tuneable band gaps across a wide range of energy levels by changing their size, whereas the bandgap is fixed by the choice of bulk materials. This reduces semiconducting particles below the size of the exciton Bohr radius. The quantum nature of existing particles set electron energies finite alike energies in an atom. The developed Quantum dots methodology is referred to as artificial atoms and energy levels are tuneable by changing their size, which in turn defines the bandgap [1]. The dots can be grown over a range of sizes, allowing them to express a variety of band gaps without changing the construction techniques. This makes it attractive for multi-junction solar cells and use of a variety of materials to improve efficiency by harvesting multiple parts of the solar spectrum. To increase the collecting efficiency of cell, the nanowire cells can be used instead of present Photovoltaic cells. The Nanowire cell can collect 15 times more solar energy as compared to conventional cells [1].

VI. Flexible Solar Tree

From the previous section, it is found that the natural light conversion technology photosynthesis is similar to solar cell technologies. Rather the natural conversion technology is found higher than solar cell light conversion technology [3,6]. These natural processes suggest various different possible solutions to these energy losses in a solar cell. The laws of science put some limits on such power conversion, but also leave some scope for improvement through multiple loss minimizations. Section II shows that sun spectrum absorption varies with nature and configuration of leaves, so give an idea of multi-spectrum material for solar cell construction. Also leaves shape, size effect light inclination and efficiency for light energy conversion. This inspiration focuses on small curved surface geometry to capture more sunlight. The multilevel solar cell - tandem cell can be fabricated to drastically reduce the transmission losses. The efficiency of natural photosynthesis also depends on water evaporations; similarly, heat dissipation play is a key factor in solar cell light conversion. .From the previous section, it is also clear that scaling down also affects efficiency drastically.

All these ideas include a modified version of the solar tree which is discussed in section II. The advancement of technology has reduced the size of solar panels to a size of 0.6 x 2.55 inches . This is comparable to the size of plant leaves and their individual identity power varies from 0.5 to 4 W. The cumulative power output will be quite high for various

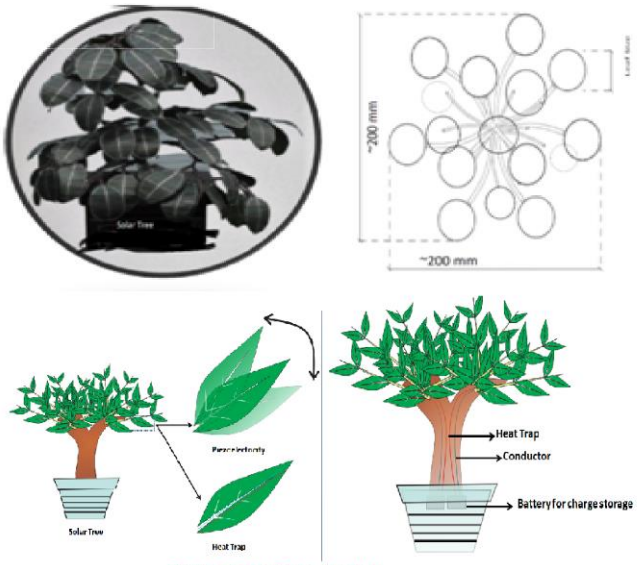


Fig 1 : Solar Tree Model

domestic applications. The concept of nano solar material solar tree leaf that would maximize the light conversion [1,5]. In order to trap maximum solar energy, the use of quantum wire PV boards is orchestrated in a Fibonacci arrangement design. One such plant pot size is proposed is shown in figure1. In this novel solar tree design, multispectral cell junctions can be fabricated to drastically reduce the transmission losses. On top of the leaves, the anti-reflective coating can be used to reduce reflection losses. Leaves are the independent unit that produces green energy and provides a place of comfort and energy for a wide variety of services. This kind of texturing of the rear surface will not be required



to achieve total internal reflection as the surface of flexible materials is already rough.

The leaf photosynthesis is also influenced by these exercises in careful control of carbon dioxide or too much-related water retention. In photosynthesis, this problem of heat generation is solved by molecules energy transfer to another until it settles in a molecule that can store energy for a long time. The conductive branches of photovoltaic solar trees can trap excited electrons into a battery for storage. This leaf-like structure acts as a thermionic generator thereby leading to the enhanced sensitivity. This design ensures the enhanced performance of a solar cell as a reduction in heat-up leads to higher efficiency. In addition to this the flexible nature of leaves also helps in harvesting wind energy. The various forms of energies such as solar, thermal and piezoelectric power by stalks, branches, and twigs of a tree. The amount of energy generation varies with the size/shape of leaves, the number of leaves etc. The flexible nature of coated piezoelectric solar cells helps in harnessing the wind energy throughout day and night. A significant amount of wind power is also generated due to wind-generated movement of small light weighted leaves. The heat energy can be converted simultaneously into electric energy using thermal converters. the energy of radiations. The various designs of flexible trees can be possible depending on geographical conditions and demands. The synchronization of solar energy with heat energy and wind energy supply continuous supply of energy even at night or when the sun is covered with clouds.

VII. CONCLUSIONS

The The sunlight based tree idea is effective to satisfy the expanding energy demands of the individuals, sparing area, and ought to be executed to give power without the issue of intensity cut-off. The concept of nano leaf is an idea to maximize the various forms of energies such as solar, thermal and piezoelectric power by stalks, branches, and twigs of a tree.

All this energy waste harvesting converts it into grid power. The adaptable nature and nano-size leaf fit it into a beautiful small size pot plant and a valuable device. This superb plant that requires no watering and sustains. The beginning venture cost of the sun based tree is likewise equivalent to the same limit PV frameworks as other framework parts are similar. The restitution time of proposed framework is by all accounts high yet because of persistent increment in the expense of network force and decrease in the expense of PV cell because of mechanical progression long recompense period can be compensated. The same photosynthesis move configuration can be reached out to different areas and applications including higher energy utilizations. At the last, Solar Tree front line eco-perfect innovation and should work in agreement with common trees.

VIII. REFERENCE

[1] Baskoutas, Sotirios; Terzis, Andreas F. (2006). "Size-dependent band gap of colloidal quantum dots". Journal of Applied Physics. 99 (1): 013708–013708–4.Bibcode:2006JAP....99a3708B. doi:10.1063/1.2158502

[2] Berny S. and others (2015), Solar Trees: First Large-Scale Demonstration of Fully Solution Coated, Semitransparent, Flexible Organic Photovoltaic Modules, , Advance Science, December 2015, https://doi.org/10.1002/advs.201500342

[3] C. Bhuvaneswari, R Rajeswari, C.Kalaiarasan and K.M.S. Muthukumararajaguru (2013), Idea to Design a Solar Tree Using Nanowire Solar Cell International Journal of Scientific and Research Publications, Volume 3, Issue 12, December 2013 1 ISSN 2250-3153.

[4] Economictimes (2019), Mini Solar Panel: Ideal power source for small devices/portable appliances, Read more at: Economictimes.indiatimes.com

[5] Kumar M. and other (2016), Traffic driven wind energy generator, Technology Letters, Vol.3 No.1, p10-13 (ISSN: 2348-8131).

[6] Melis A. (2009), Plant and Microbial Biology, University of California, Berkeley, CA 94720-3102, USA, Plant Science 177, p 272–280. Gupta S., and others (2015), "The benefits and applications of Solar Tree With Natural Beauty Of Trees", SSRG International Journal of Electrical and Electronics Engineering, Issue April 2015 Vol-4 Issue-2 2018 IJARIIE-ISSN(O)-2395-4396 7653.

[7] Srinivas (2016), A P R, "Design and Development of a SOLAR TREE", International Journal of Scientific & Engineering Research, Volume 7, Issue 10, October-2016 ISSN 2229-5518

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/342881094

Forecasting the Probability of Cost Overrun Risk of Indian Construction Projects using Fuzzy Model

Article · January 2020

CITATIONS
0

READS
30

3 authors, including:



Pradeep K. Goyal
Delhi Technological University
8 PUBLICATIONS **13** CITATIONS
SEE PROFILE



Prof.(Dr.) R.C. Chhipa
Suresh Gyan Vihar Univeristy, Jaipur
64 PUBLICATIONS **123** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

COST OVERRUN ANALYSIS MODEL IN FUZZY ENVIRONMENT FOR CONSTRUCTION PROJECTS View project

Physicochemical and spectroscopic analysis of effluents of textile and paper industries View project

All content following this page was uploaded by Prof.(Dr.) R.C. Chhipa on 12 July 2020.

The user has requested enhancement of the downloaded file.

International Journal on Emerging Technologies **11**(1): 10-22(2020)

ISSN No. (Print): 0975-8364
ISSN No. (Online): 2249-3255

Forecasting the Probability of Cost Overrun Risk of Indian Construction Projects using Fuzzy Model

Savita Sharma[1], Pradeep K. Goyal[2] and R.C. Chhipa[3]
[1]Research Scholar, Department of Civil Engineering, Suresh Gyan Vihar University, Jaipur (Rajasthan), India.
[2]Associate Professor, Department of Civil Engineering, Delhi Technological University, Delhi, India.
[3]Professor, Suresh Gyan Vihar University, Jaipur (Rajasthan), India.

(Corresponding author: Savita Sharma)
(Received 04 October 2019, Revised 11 November 2019, Accepted 10 December 2019)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: A fuzzy model for forecasting the probability of cost overrun risk for Indian construction project is developed by identifying, classifying and ranking the cost overrun factors of the construction industries. Fifty five cost overrun factors are identified, classified into eleven clusters through a detailed literature review process and discussion with experts from the Indian construction industry. A questionnaire survey in Indian construction industry was conducted for data collection to calculate Relative Importance Index (RII) of the factors and clusters. On the basis of the relative importance index of the factors and clusters, a model is developed using the fuzzy logic tool box of MATLAB program software. The model is developed in two stages. First the probability of cost overrun due to the clusters of cost overrun factors is determined and the final cost overrun risk probability of the project is determined taking into consideration the probability of all the clusters. Applicability of the fuzzy model has been tested on a real case study in India. With the help of the proposed model, it is possible to guide project managers and decision makers to make more informative decisions such as contingency estimation, mark-up estimation, bid price, selection of optimum procurement route, evaluation of different projects and preparation of project portfolios.

Keywords: Construction industry, Cost overrun, Probability, MATLAB, Fuzzy logic, India.

I. INTRODUCTION

To complete a construction project within stipulated budget is an important factor among all the norms required for the success of the project [1-3], yet construction projects are found to be suffering from the issue of cost overrun. Cost overrun is defined as the increased costs over estimated cost [4]. This issue of cost overruns is critical both for developed as well as developing countries [5-13]. Indian construction projects are also suffering from the problem of cost overrun. A joint study was conducted by PMI and KPMG with the support of MOSPI in India on infrastructure projects in which a total cost overrun of Rs 2.19 lakh crore for 1304 projects was reported in 2018 [14]. Generally the cost overrun risk analysis is ignored in the projects and the cost of a project is determined just by adding some contingency charges, consequently many projects fail to complete the project in stipulated budget. Therefore it is essential to estimate the probability of cost overrun risk for Indian construction projects to determine a reliable cost of the project.

Several theories and methods have been proposed for risk analysis for construction projects. Zhi, (1995) [15] used qualitative method by using P-I matrix of for risk analysis of overseas construction project, Baccarini and Archer (2001) [16] Computed a risk score for cost, time or quality of project by employing P-I model. Hull (1990) [17] proposed various models by using Monte Carlo Simulation (MSC) and PERT theory for the risk analysis of cost, schedule and technical performance. Özta & Ökmen (2004) [18] presented Monte Carlo simulation technique for determining the risks and uncertainties associated with project cost and duration. Thomas *et al.*, (2006) [19] used a Fault tree method to model risky scenarios by utilizing linguistic variables to evaluate risk likelihood and impact. Dey (2001) [20] presented risk analysis model based on AHP and decision trees. Dikmen and Birgonul (2006) [21] used AHP for risk and opportunity estimation of international construction projects. Nasir *et al.*, (2003) [22] used belief network and MCS for estimating activity duration. Kangari and Riggs (1989) [23] proposed a model for risk analysis by using FST. Tah and Carr (2000) [24] presented a model for hierarchical risk breakdown structure for qualitative risk assessment through the concepts of fuzzy association, fuzzy composition and application of fuzzy logic.

According to a study by Sharma and Goyal [25], it was concluded that each theory has its own advantages, disadvantages and no theory seems to be complete for handling the uncertainty. However, based on the work done earlier it may be suggested that fuzzy theory is more suitable for handling the complex problems in construction industry as the process is based upon experience, assumptions and human judgment. The fuzzy theory has been applied in the different areas of construction projects as decision support tool, enhancing the performance, evaluation and assessment, forecasting and modeling construction risks [26-35, 40].

The previous studies of cost overrun risk analysis in India were very few in which fuzzy techniques were applied. Therefore in this study a fuzzy model for forecasting the probability of cost overrun risk in Indian construction project is developed by identifying, classifying and ranking the cost overrun factors of the construction industries. The construction industry caters to building and other infrastructure development project. Data for the cost overrun factors are collected from medium to large construction industry with turnover varying from 50 crores to 600 crores. Collected data through questionnaire are validated from sample verification technique and expert opinion. Only reliable data are used in the proposed model. With the help of the proposed model, it is possible to guide project managers and decision makers to make more informative decisions such as contingency estimation, mark-up estimation, bid price, selection of optimum procurement route, evaluation of different projects and preparation of project portfolios.

II. METHODOLOGY

The adopted methodology for the development of the model and for obtaining the results of interest is explained diagrammatically in Fig. 1. This model is developed in two stages as shown in Fig.1. First the probability of cost overrun due to each cluster is assessed by the probability level of the factors of respective cluster. The probability of cost overrun of project is assessed by the probability output of each cluster as determined in the previous step. Major components of the methodology are explained in the following sections.

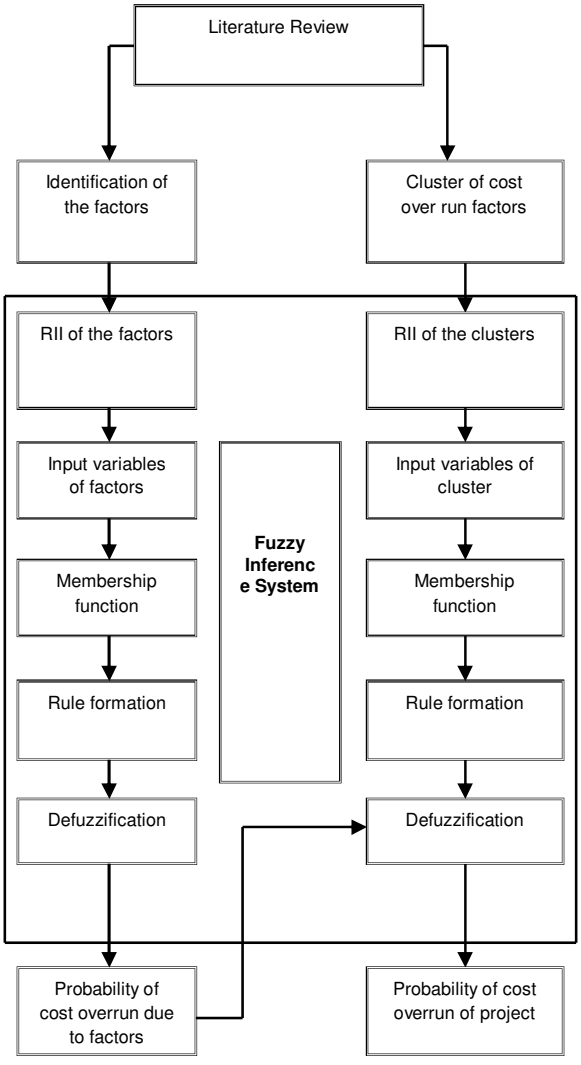


Fig. 1. Methodology.

A. Identification and classification of cost overrun factors

A number of studies conducted on the factors responsible for cost overrun in the construction projects are reviewed [6-14]. The cost escalation causes and their classification are given by Sharma and Goyal (2014) [25] for the construction industry. Fifty five important factors, responsible for cost overrun are identified through intensive literature review. These factors are grouped in eleven clusters named as 'Cluster Owner', 'Cluster Contractor', 'Cluster Consultant', 'Cluster Design', 'Cluster Project', 'Cluster

Material', 'Cluster Labor', 'Cluster Equipment', 'Cluster Contract', 'Cluster Coordination/communication' and 'Cluster External'.

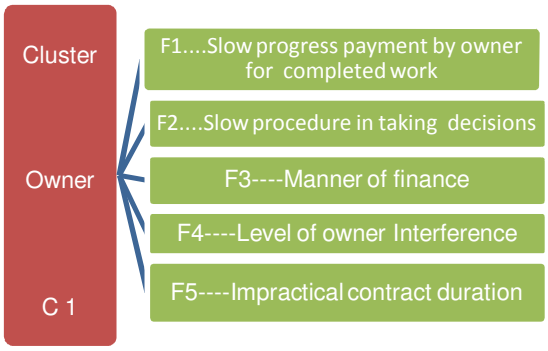


Fig. 2. Cost overrun factors of 'Cluster Owner'.

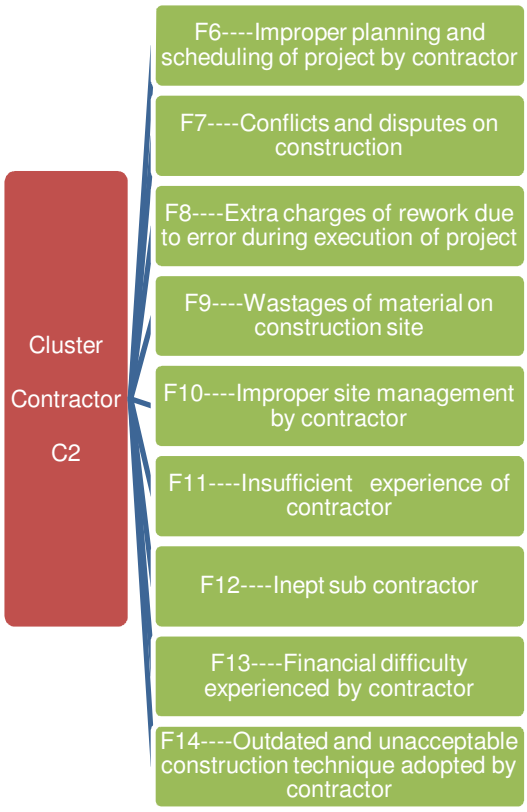


Fig. 3. Cost overrun factors of 'Cluster Contractor'.

All the factors are represented as F1, F2, F3 etc. and clusters are represented as C1, C2, and C3 etc. slow progress payment by owner for completed work, slow procedure in taking decisions, manner of finance, level of owner interference and impracticable contract duration are the factors of 'Cluster Owner' as shown in Fig. 2. improper planning and scheduling of project by contractor, conflicts and disputes on construction sites, extra charges of rework due to error during execution of project, wastages of material on construction site, improper site management by contractor, insufficient experience of contractor, inept sub contractor, financial difficulty experienced by contractor and outdated and unacceptable construction technique adopted by contractor are the factors of 'Cluster Contractor' as shown in Fig. 3.

In correct estimation of time and cost, contract management, delay in inspecting of work, quality assurance & quality control and experience level of consultant are the factors of 'Cluster Consultant' as shown in Fig. 4. In a similar manner all the factors of 'Cluster Design', 'Cluster Project', 'Cluster Material', 'Cluster Labor', 'Cluster Equipment', 'Cluster Contract', 'Cluster Coordination/communication' and 'Cluster External' are shown in Fig. 4-12.

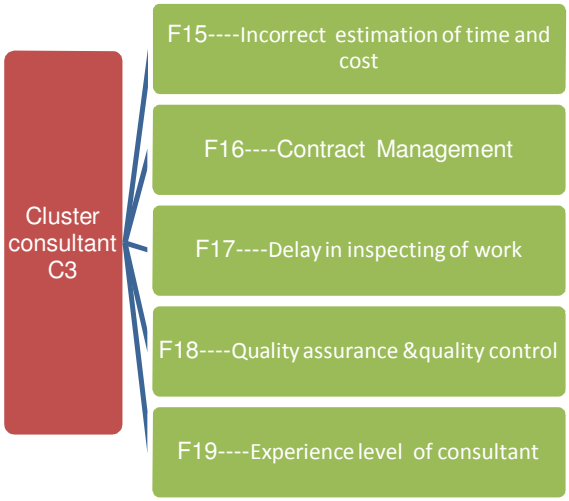


Fig. 4. Cost overrun factors of 'Cluster Consultant'.

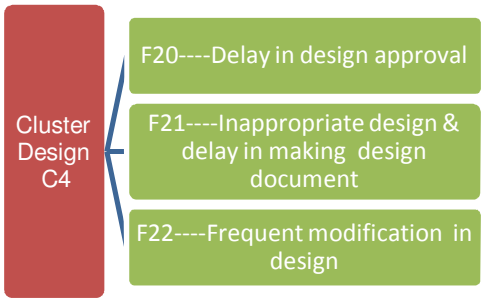


Fig. 5. Cost overrun factors of 'Cluster Design'.

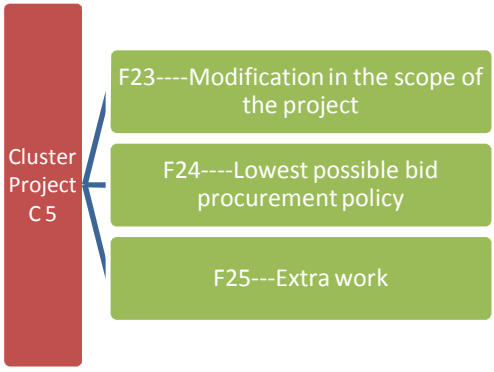


Fig. 6. Cost overrun factors of 'Cluster Project'.

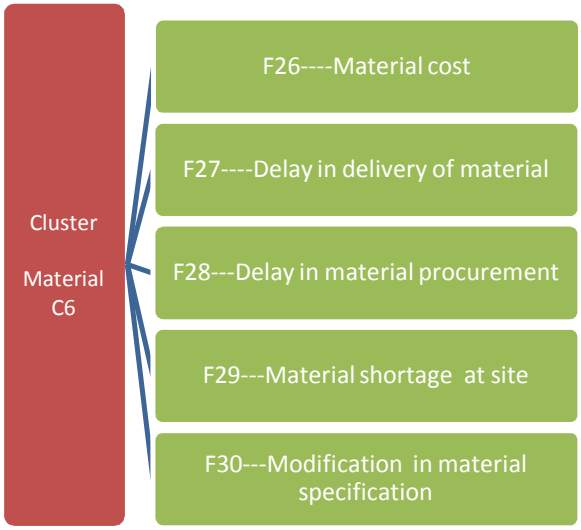


Fig. 7. Cost overrun factors of 'Cluster Material'.

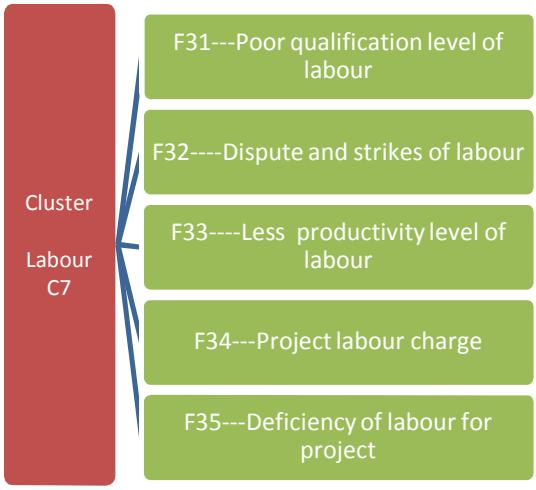


Fig. 8. Cost overrun factors of 'Cluster Labor'.

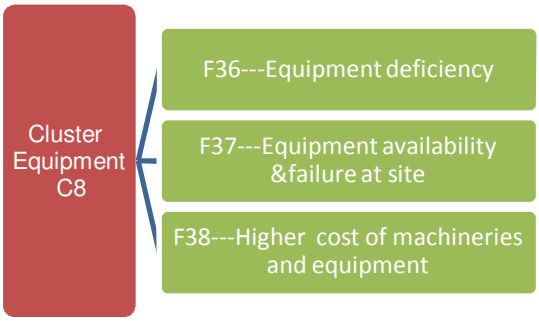


Fig. 9. Cost overrun factors of 'Cluster Equipment'.

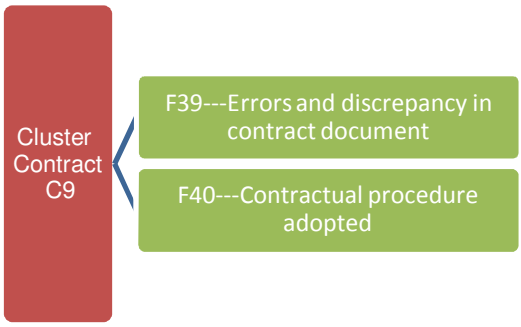


Fig. 10. Cost overrun factors of 'Cluster Contract'.

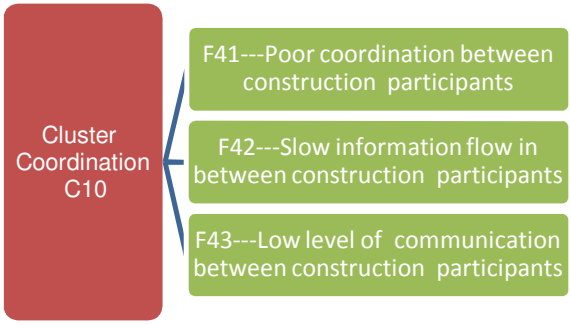


Fig. 11. Cost overrun factors of 'Cluster Coordination'.

B. Relative importance index of factors and clusters

For ranking the cost overrun factors and cluster relative importance index (RII) is calculated by collecting data through a questionnaire survey in Indian construction. The population for survey was drawn from the practitioners of the Indian construction industry, which included owners comprising government sectors (key decision makers), contractors and consultants. A non-probability sampling technique known as convenience and snow ball is used to collect the sample. The questionnaire was distributed through friend and referral network. Personal interviews were also conducted for obtaining the response quickly. Total 250 questionnaires were distributed to the potential respondents within the Indian construction company comprising owner, contractor and consultant. 100 questionnaire set to owner, 100 questionnaire set to contractor and 50 questionnaire set to consultant were distributed. Out of 250 questionnaire set, only 135 responses were returned. Out of total respondents, 62 (45%) of the respondents are working in clients' organization while 47 (34.81%) and 26 (19.2%) respondents are professionals that work in contracting and consulting organizations respectively.

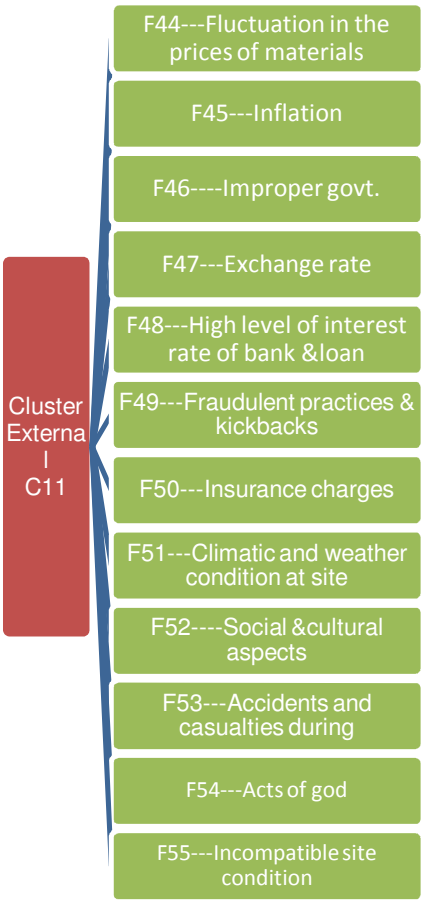


Fig. 12. Cost overrun factors of 'Cluster External'.

The respondents were asked for judging the importance level of identified cost overrun factors. A likert scale of 1 to 5 was used to evaluate the importance level of each factor. These numbers were allocated to the respondents' score as '1 = Very low; 2=Low; 3=Medium; 4=High; 5=Very high'. For ranking these cost overrun factors and clusters the following relation was used to determine the relative importance index (RII)

$$RII = \sum W / (A*N) \quad (1)$$

where W is the weighting for each factor provided by the respondent (it ranges from 1 to 5), A is the maximum weight and N is the number of overall respondents.

C. Model development using fuzzy theory

The fuzzy model to assess the probability of cost overrun is developed in fuzzy inference system (FIS) of MATLAB [36]. Fuzzy inference is a procedure where application of fuzzy logic is employed for mapping from a set of given input to obtain an output. The important features of fuzzy inference system have been explained here in brief.

Fuzzy set. The notion of fuzzy logic was originally pioneered by Zadeh in a seminal article on fuzzy sets theory in 1965 [37]. The concept of the theory is founded on multiple-valued logic; it processes and introduces interpretations which are expressed in vague linguistic terms rather than exact and precise crisp value. The theory is very useful for handling the partial truth issues.

A fuzzy set A of a universe of discourse X can be defined through Eqn. (2):

$$A = \{(x, \mu_A(x)) \mid x \in A, \mu_A(x), \in [0, 1]\} \quad (2)$$

where μ_A (x) is the membership function which provides X a membership values in the interval of 0 to 1.

Fuzzy inference system. As per litreture review Mamdani and Sugeno type of fuzzy inference methods are generally employed. For this research study, Mamdani type fuzzy inference is used because of its wide application [38]. The system of fuzzy inference is basically a rule-based. The crisp input members are fuzzified first to convert them into fuzzy sets by defining the membership functions. After that inference engine of the fuzzy logic system repossesses the knowledge contained in a rule-base where all inputs are combined as per the constructed fuzzy rules. The output fuzzy sets obtained through this process then aggregated and defuzzified where they are converted from a fuzzy set to obtain a crisp number.

The model is developed in two stages as shown in Fig. 2. First the probability of cost overrun due to each cluster is assessed by the probability level of the factors of respective cluster. The probability of cost overrun of project is assessed by the probability output of each cluster as determined in the previous step. The process of fuzzy inference system is applied for both the stages.

The various steps of fuzzy inference process employed for model development are outlined below:

1. The first step of fuzzy inference system is to define all input and output variables for model development. For the first stage of model development the various 'factors' are taken as the input variables and output is the 'probability of cost overrun due to cluster'. For the second stage various clusters are taken as the input variables and 'Probability of cost overrun of project' which is abbreviated as 'PCOP' is defined as the output variable.
2. The input variables are then fuzzified with the help of membership function in this step. The membership function is the fuzziness degree and gives a mathematical value to all input variables. Different shapes of membership functions such as trapezoidal, triangular, bell-shaped, piecewise-linear, Gaussian, g bell-shaped etc. can be used. in this study, g bell-shaped are used as Gaussian and bell membership functions have the advantage of being smooth and nonzero at all points [39]. It can be graphically represented as shown in Fig. 13.

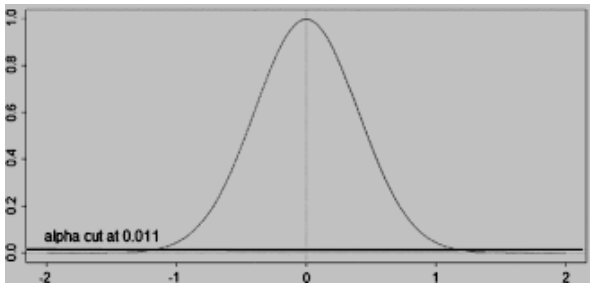


Fig. 13. Gbell-shaped fuzzy membership functions.

3. Fuzzy rules are defined in this step to perform the inference. Fuzzy rules are logical inference and connect

the input variables to output variables. IF-THEN forms of fuzzy rules are used here in this research study. Fuzzy rule for the input linguistic variables x_1, x_2 with their corresponding fuzzy values A_1, A_2 and output linguistic variable y with B as its fuzzy value, can be explained as follows.

If x_1 is A_1 AND x_2 is A_2 then y is B

If part of this rule "x_1 is A_1 AND x_2 is A_2" is called antecedent and then part of the rule "y is B" is known as consequent.

Fuzzy operators "AND", "OR" and "NOT" are also incorporated to construct fuzzy rules.

The estimated values of relative importance for the cost overrun factors are allocated as fuzzy rules weight. An interview was developed with the experts of construction industry in India for the construction of fuzzy rules for forecasting the probability of cost overrun.

4. Defuzzification process takes place finally, where the fuzzy results of the inference engine are converted into a crisp one. Different types of defuzzification methods have been recommended in literature. In this study 'centroid of area' method is applied to forecast 'the probability of cost overrun of the project'. This method of 'centroid of area' calculates the weighted average of a fuzzy set. The outcome of defuzzification to a fuzzy conclusion by using this method can be expressed by the formula

$$Z_{COA} = (\int_Z \mu_A (Z)\, Z\, dz) / (\int_Z \mu_A(Z)\, dz) \quad (3)$$

where Z_{COA} is the crisp output, μ_A (Z) is the aggregated membership function and z is the output variable.

D. Model validation

To test the reliability of the proposed fuzzy cost overrun assessment model, it is tested in a real case study. For this purpose an interview was conducted with a team of the experts of a leading Indian construction company. The panel of the experts included top 10 executive engineers, project managers and site engineers of the company. The experts were requested to examine the cost overrun factors considered for this research and filled in the required information regarding cost overrun factors to assess the proposed model. They were also requested to estimate the cost risk level of the project.

III. NUMERICAL STUDY AND RESULTS

The summary of the calculated, relative importance index (RII) of the cost overrun factors in Indian construction industry and the proposed fuzzy model for these factors has been discussed in the following sections:

A. RII of cost overrun factors and important cost overrun factors in Indian construction industry

The data collected in the questionnaire are analyzed from owner, contractor and consultant's perspective**.** The relative importance index (RII) is computed for each factor causing cost overrun to determine the important factors causing cost overrun in Indian construction industry.

RII of Cluster Owner. RII of the factors slow progress payment by owner for completed work, slow procedure in taking decisions, manner of finance, level of owner interference and impracticable contract duration are 0.498, 0.522, 0.425, 0.357 and 0.676 respectively as shown in Fig. 14.

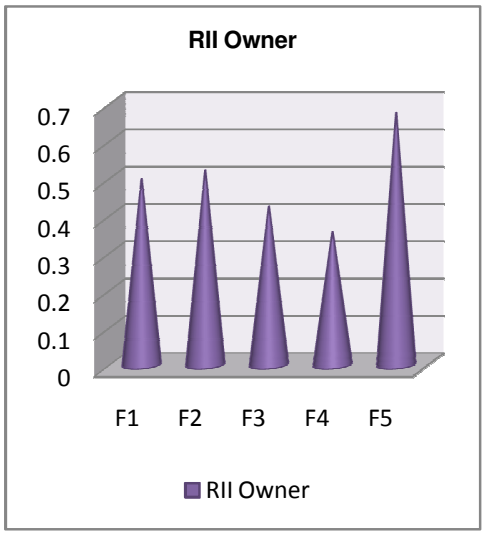


Fig. 14. RII of cluster owner.

RII of Cluster Contractor. RII of the factors improper planning and scheduling of project by contractor, conflicts and disputes on construction sites, extra charges of rework due to error during execution of project, wastages of material on construction site, improper site management by contractor, insufficient experience of contractor, inept sub contractor, financial difficulty experienced by contractor and outdated and unacceptable construction technique adopted by contractor are 0.670, 0.548, 0.495, 0.445, 0.385, 0.352, 0.413, 0.675 and 0.320 respectively as shown in Fig. 15.

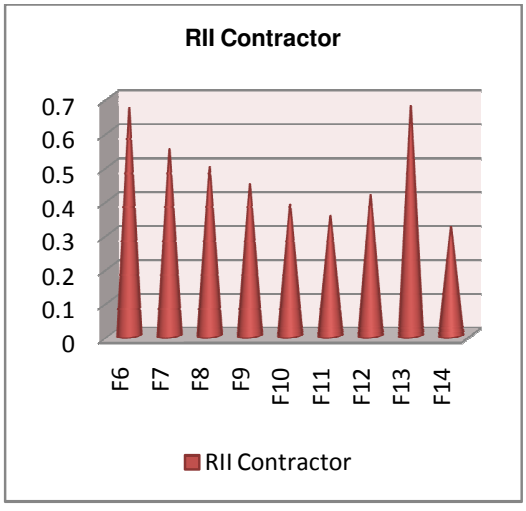


Fig. 15. RII of cluster Contractor.

RII of Cluster Consultant. RII of the factors incorrect estimation of time and cost, contract management, delay in inspecting of work, quality assurance & quality control and experience level of consultant are 0.742, 0.556, 0.391, 0.430 and 0.394 respectively as shown in Fig. 16.

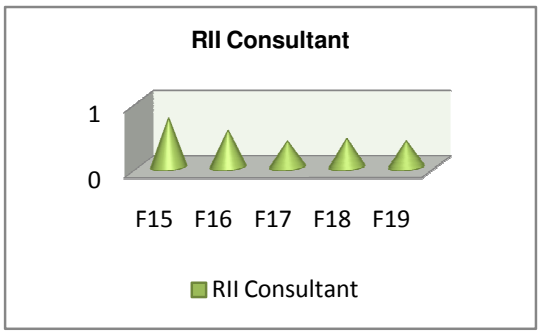


Fig. 16. RII of cluster Consultant

RII of Cluster Design. RII of the factors delay in design approval, inappropriate design & delay in making design document and frequent modification in design are 0.442, 0.478 and 0.704 respectively as shown in Fig. 17.

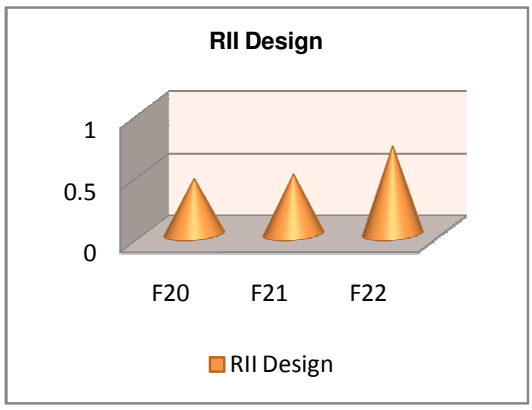


Fig. 17. RII of cluster Design.

RII of cluster Project. RII of the factors modification in the scope of the project, lowest possible bid procurement policy and quantity of extra work done are 0.556, 0.800 and 0.724 respectively as shown in Fig. 18.

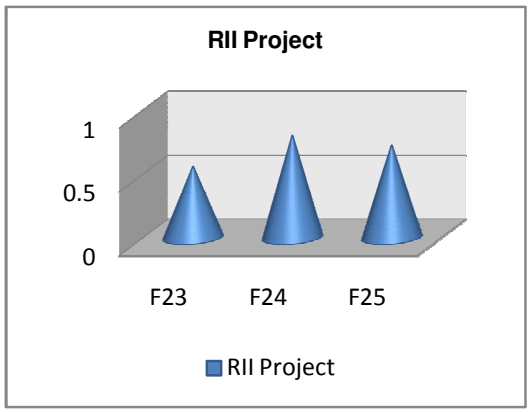


Fig. 18. RII of cluster Project.

RII of cluster Material. RII of the factors material cost, delay in delivery of material, delay in material procurement, material shortage at site and modification in material specification are 0.565, 0.343, 0.317, 0.371 and 0.411 respectively as shown in Fig. 19.

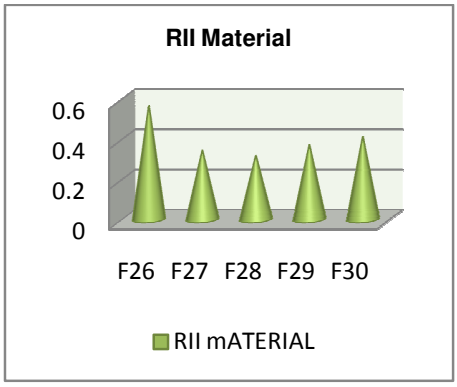


Fig. 19. RII of cluster Material.

RII of cluster Labor. RII of the factors poor qualification level of labor, dispute and strikes of labor, Less productivity level of labor, project labor charge and deficiency of labor for project are 0.341, 0.328, 0.352, 0.628 and 0.370 respectively as shown in Fig. 20.

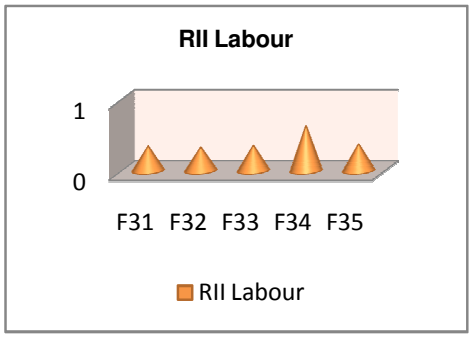


Fig. 20. RII of cluster Labour.

RII of cluster Equipment. RII of the factors equipment deficiency, equipment availability & failure at site and higher cost of machineries and equipment are 0.317, 0.302 and 0.582 respectively as shown in Fig. 21.

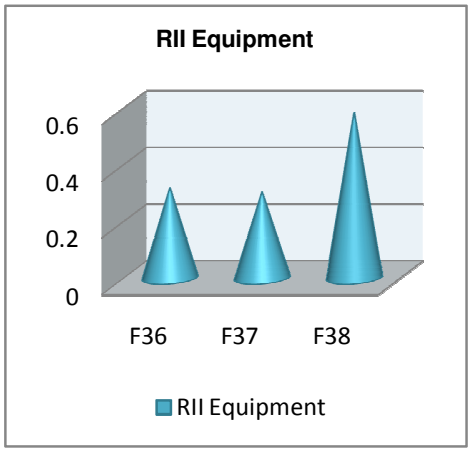


Fig. 21. RII of cluster Equipment.

RII of cluster Contract. RII of the factors errors and discrepancy in contract document and contractual procedure adopted are 0.728 and 0.410 respectively as shown in Fig. 22.

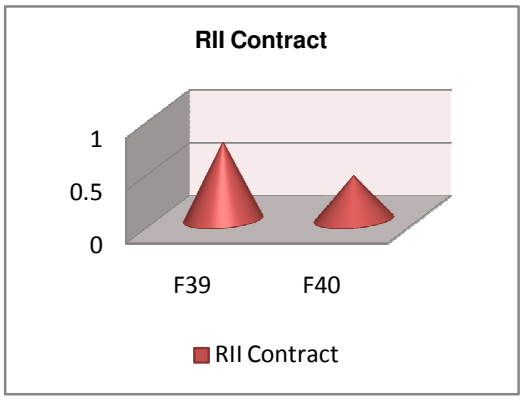


Fig. 22. RII of cluster Contract.

RII of cluster Coordination. RII of the factors are Poor coordination between construction participants, slow information flow in between construction participants and Low level of communication between construction participants are 0.659, 0.440 and 0.430 respectively as shown in Fig. 23.

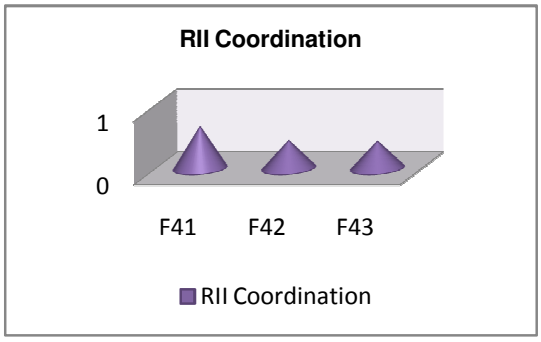


Fig. 23. RII of cluster Coordination.

RII of cluster External. RII of the factors fluctuation in the prices of materials, inflation, improper govt. policy, exchange rate, high level of interest rate of bank &loan, fraudulent practices & kickbacks, insurance charges of project, climatic and weather condition at site, social & cultural aspects, accidents and casualties during construction period, acts of god and incompatible site condition are 0.829, 0.788, 0.780, 0.621, 0.560, 0.520, 0.490, 0.485, 0.423, 0.370, 0.321 and 0.665.

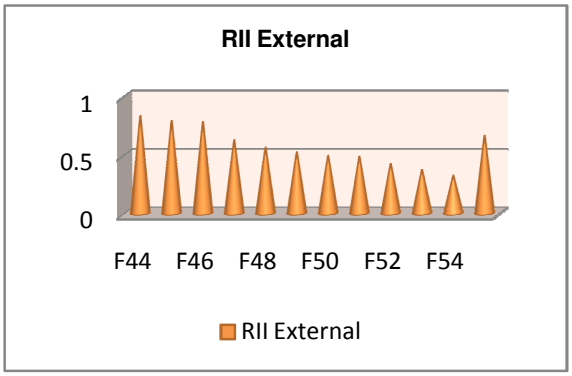


Fig. 24. RII of cluster External.

RII of the clusters. The RII of the clusters is shown in Fig. 25. RII of the clusters of ' Cluster Owner', 'Cluster Contractor', 'Cluster Consultant', 'Cluster Design', 'Cluster Project', 'Cluster Material', 'Cluster Labor', 'Cluster Equipment', 'Cluster Contract', 'Cluster Coordination/communication' and 'Cluster External' are 0.495, 0.477, 0.502, 0.541, 0.693, 0.401, 0.391, 0.4, 0.569, 0.509 and 0.570 respectively.

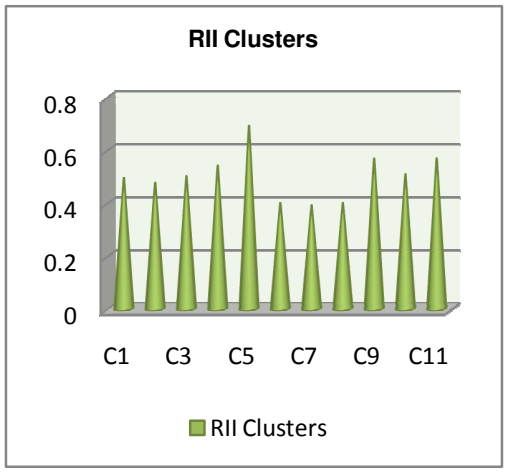


Fig. 25. RII of clusters.

B. Model development

The model is developed in two stages. First the probability of cost overrun due to each cluster is assessed by the probability level of the factors of respective cluster such as the probability of cost overrun due to the 'Cluster Owner' is estimated by probability of cost overrun due to the factors of 'Cluster Owner'.

The probability of cost overrun of project is then assessed by the probability output of each cluster as determined in the previous step.

Define input and output variables. For the first stage of the model development the cost overrun factors are taken as the input variables and output probability of cost overrun due to the cluster is determined from the model. The input and output variables for 'Cluster Owner' are described in Table 1. The various factors of cost overrun of 'Cluster Owner' are Slow progress payment by owner for completed work, Slow procedure in taking decisions, Manner of finance, Level of owner interference and Impracticable contract duration. These factors are considered as input variable for determining the output 'the probability of cost overrun due to the 'Cluster Owner'. During model development these factors are considered as F1, F2, F3, F4, and F5. And the output variable is abbreviated as 'POC1'. Similarly, probability of cost overrun due to the 'Cluster Contractor', 'Cluster Consultant', 'Cluster Design', 'Cluster Project', 'Cluster Material', 'Cluster Labor', 'Cluster Equipment', 'Cluster Contract', 'Cluster Coordination/communication' and 'Cluster External' is determined using factors related to that cluster.

As mentioned earlier the cost overrun probability of the project is determined taking into consideration the probability of all the clusters. Therefore for the second stage of model development clusters such as 'Clusters Owner', 'Cluster Contractor', 'Cluster Consultant', 'Cluster Design', 'Cluster Project', 'Cluster Material',

'Cluster Labor', 'Cluster Equipment', 'Cluster Contract', 'Cluster Coordination/communication' and 'Cluster External' are considered as input variables.

Table 1: Input and output variables for determining 'probability of cost overrun due to cluster 'Owner'.

Input variables		Output variable
F1	Slow progress payment by owner for completed work	probability of cost overrun due to cluster' C1 (Owner) POC1
F2	Slow procedure in taking decisions	
F3	Manner of finance	
F4	Level of owner interference	
F5	Impracticable contract duration	

And the output variable is 'probability of cost overrun of project' as shown in Table 2. During model development these clusters are considered as C1, C2, C3, C4, C5, C6, C7, C8, C9, C10 and C11.and the output variable is 'Probability of cost overrun of project' which is abbreviated as **'PCOP'.**

Table 2: Input and output variables for determining the probability of cost overrun of project.

Input variables		Output variable
C1	Owner	Probability of cost overrun of project (PCOP)
C2	Contractor	
C3	Consultant	
C4	Design	
C5	Project	
C6	Material	
C7	Labour	
C8	Equipment	
C9	Contract	
C10	Coordination	
C11	External	

The input variables in the form of factors and output probability of owner related 'POC1' cluster are shown in the FIS Editor window of Fuzzy Logic tool box of MATLAB Program Software in Fig. 26.

Fuzzy Membership Functions. The membership function of input variables in the form of factors and clusters of factors and output variable Probability of cost overrun of project 'PCOP 'were generated using the membership function editor in the Matlab fuzzy logic toolbox as shown in Fig. 27.

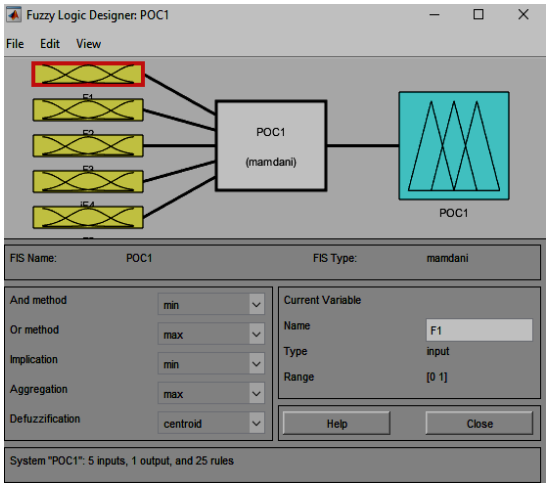


Fig. 26. Defining Input and output variables.

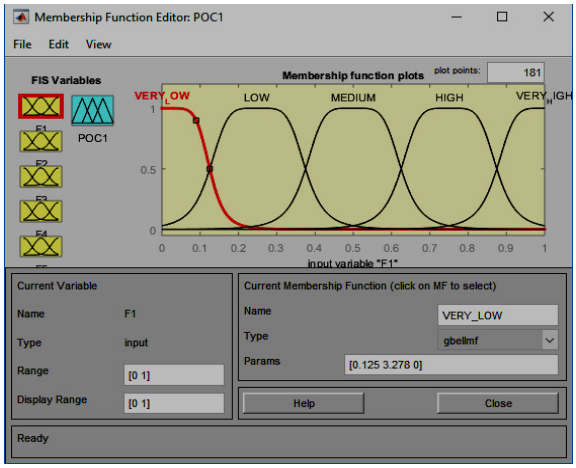


Fig. 27. Defining membership of Input and output variables.

Fuzzy rules and weighting of rules. Considering 55 cost overrun factors, total 275 rules are formed. Samples of the fuzzy rules of cluster 'Owner' are presented in Table 3.

Table 3: Sample fuzzy rules of factors of 'Cluster Owner'.

Rule	Antecedent	consequence	weighting
1	If the probability of cost overrun due to the factor 'Slow progress payment by owner for completed work ' is very low	Then the probability of cost overrun due to Cluster 'owner' is very low	0.498
6	If the probability of cost overrun due to the factor 'Slow procedure in taking decisions' is very low	Then the probability of cost overrun due to Cluster 'owner' is very low	0.522
11	If the probability of cost overrun due to the factor 'Manner of finance ' is very low	Then the probability of cost overrun due to Cluster 'owner' is very low	0.425
16	If the probability of cost overrun due to the factor 'Level of owner interference' is very low	Then the probability of cost overrun due to Cluster 'owner' is very low	0.357
21	If the probability of cost overrun due to the factor 'Impracticable contract duration' is very low	Then the probability of cost overrun due to Cluster 'owner' is very low	0.676

Table 4: Sample Fuzzy rules for estimating the probability of cost overrun of the project.

Rule	Antecedent	Consequence	Weighting
1	If the probability of the cost overrun due to 'Cluster Owner' is very low	Then the 'probability of cost overrun of the project' is very low	0.495
6	If the probability of the cost overrun due to the 'Cluster Contractor' is very low	Then the 'probability of cost overrun of the project' is very low	0.477
11	If the probability of the cost overrun due to the 'Cluster Consultan' is very low	Then the 'probability of cost overrun of the project' is very low	0.502
16	If the probability of the cost overrun due to the 'Cluster Design' is very low	Then the 'probability of cost overrun of the project' is very low	0.541
21	If the probability of the cost overrun due to the 'Cluster Project' is very low	Then the 'probability of cost overrun of the project' is very low	0.693
26	If the probability of the cost overrun due to the 'Cluster Material' is very low	Then the 'probability of cost overrun of the project' is very low	0.401
31	If the probability of the cost overrun due to the 'Cluster Labour' is very low	Then the 'probability of cost overrun of the project' is very low	0.391
36	If the probability of the cost overrun due to the 'Cluster Equipment' is very low	Then the 'probability of cost overrun of the project' is very low	0.4
41	If the probability of the cost overrun due to the 'Cluster Contract' is very low	Then the 'probability of cost overrun of the project' is very low	0.569

Since the probability of cost overrun due to the cluster 'Owner' is directly related to the probability of cost overrun due to slow progress payment by owner for completed work, slow procedure in taking decisions, Manner of finance, Level of owner interference and Impracticable contract duration, therefore the probability of cost overrun of these factors will influence the probability of cost overrun due to the cluster 'Owner'. The calculated RII of the factors are the weighting of the rules.

Considering 11 clusters of cost overrun factors, total 55 rules are formed. Since the probability of cost overrun of the project is directly related to the probability of cost overrun due to the 'Clusters Owner', 'Cluster Contractor', 'Cluster Consultant', 'Cluster Design', 'Cluster Project', 'Cluster Material', 'Cluster Labor', 'Cluster Equipment', 'Cluster Contract', 'Cluster Coordination/communication' and 'Cluster External'. Therefore the probability of cost overrun due to these clusters will influence the probability of cost overrun of the project. Fuzzy rules for estimating the probability of cost overrun of the project are presented in Table 4.

For fuzzy inference the rules are made as follows:

If F1 is 'very low' then POC1 is 'very low': Which means that if the probability of cost overrun due to the factor 'slow payment of complete work by owner' is 'very low' then the probability of cost overrun due to Cluster 'Owner' is 'very low'. The rule editor window of Fuzzy Logic tool box of MATLAB Program Software is shown in Fig. 28.

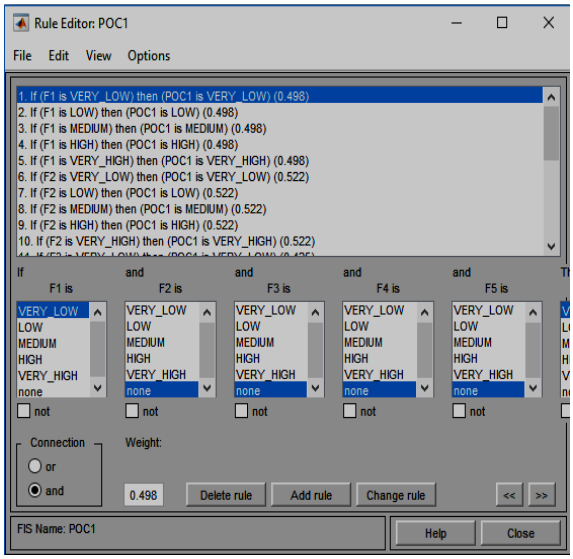


Fig. 28. Rule formation.

Defuzzification. Finally the 'probability of cost overrun due to Cluster 'Owner' PCO1 and 'probability of cost overrun' of project' is estimated by defuzzification by 'centroid of area' method as shown in Fig. 29.

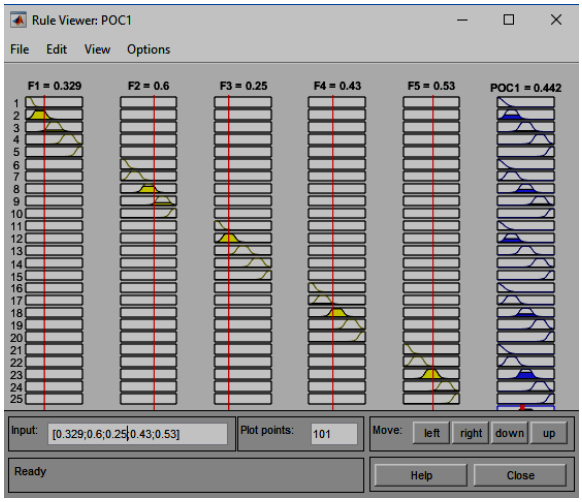


Fig. 29. Defuzzification process.

C. Testing of the model

To test the reliability of the proposed fuzzy cost overrun assessment model, an interview has been conducted with a team of the experts of a leading Indian construction company. The panel of the experts consists of the top 10 executive engineers, project managers and site engineers of the company. The experts were requested to examine the cost overrun factors considered for this research and filled in the required information regarding cost overrun factors to assess the proposed model. They were also requested to estimate the probability of cost overrun of the project. The information provided by the group of the experts are presented as shown in the following table in the Table 5.

Table 5: Information provided by the group of the expert.

S.No.	Factors	Probability of occurrence of factor
1	F1	0.329
2	F2	0.6
3	F3	0.25
4	F4	0.43
5	F5	0.53
6	F6	0.4
7	F7	0.3
8	F8	0.25
9	F9	0.3
10	F10	0.5
11	F11	0.35
12	F12	0.25
13	F13	0.55
14	F14	0.25
15	F15	0.42
16	F16	0.36
17	F17	0.27
18	F18	0.38
19	F19	0.23
20	F20	0.35
21	F21	0.62
22	F22	0.58
23	F23	0.22
24	F24	0.8
25	F25	0.5
26	F26	0.6
27	F27	0.33
28	F28	0.28
29	F29	0.15
30	F30	0.24
31	F31	0.27
32	F32	0.16
33	F33	0.23
34	F34	0.32
35	F35	0.21
36	F36	0.18
37	F37	0.21
38	F38	0.23
39	F39	0.25
40	F40	0.27
41	F41	0.56
42	F42	0.42
43	F43	0.38
44	F44	0.87
45	F45	0.82
46	F46	0.93
47	F47	0.62
48	F48	0.72
49	F49	0.4
50	F50	0.43
51	F51	0.85
52	F52	0.35
53	F53	0.27
54	F54	0.23
55	F55	0.74

The determined probabilities of cost overrun due to clusters by model are given in Table 6.

Table 6: Probability level of clusters.

S.No.	Cluster	Name of cluster	Probability of cost overrun due to cluster	Probability of cost overrun of the project
1	C1	Owner	0.442	0.435
2	C2	Contractor	0.405	
3	C3	Consultant	0.406	
4	C4	Design	0.476	
5	C5	Project	0.524	
6	C6	Material	0.436	
7	C7	Labour	0.289	
8	C8	Equipment	0.26	
9	C9	Contract	0.263	
10	C10	Coordination	0.465	
11	C11	External	0.609	

The probability of cost overrun of the project was calculated by the experts was in the range of 0.35-0.50. Provided information was tested using the proposed model and The probability of cost overrun of the project obtained was 0.435 as shown in Fig. 30, which is much closer to the estimated cost risk level of the project estimated by the panel of the experts of the company.

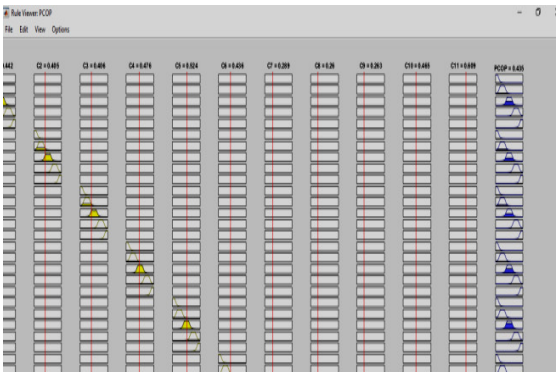


Fig. 30. Probability of cost overrun of the project.

V. CONCLUSION

It is observed that very few projects get completed within stipulated costs in India, despite the importance of economy in construction projects. The cost of a project is calculated just by adding some contingency charges and proper analysis of cost overrun risk is ignored, consequently many projects fail to complete the project in stipulated budget. Therefore in this study a fuzzy model for forecasting the probability of cost overrun risk in Indian construction project is developed by identifying, classifying and ranking of the factors responsible for the cost overrun in the construction industries. The fuzzy theory is chosen for developing the model as the theory is found suitable for quantifying the probability of cost overrun risk of the construction project when the variables regarding the cost overrun factors are expressed in linguistic terms. The fuzzy set theory (FST) is equipped to take care of imprecise, incomplete and uncertain data expressed in linguistic form. Through an extensive literature survey and discussion with the experts of the Indian construction industry, fifty five cost overrun factors were identified and then classified into eleven clusters. A questionnaire survey was conducted in Indian construction industry for data collection to calculate relative importance index (RII) of the recognized factors causing cost overrun and clusters. Total 250 questionnaires were distributed to the potential respondents within the Indian construction company comprising owner, contractor and consultant. Out of 250 questionnaire set, only 135 responses were returned. The calculated relative importance index of the factors and clusters was used to develop model using the fuzzy logic tool box of MATLAB program software. First the probability of cost overrun due to the clusters of cost overrun factors is determined. The final cost overrun risk probability of the project is then determined taking into consideration the probability of all the clusters. For all input and output linguistic variables five membership functions are defined using g-bell shaped fuzzy numbers. Fuzzy rules for estimating the probability of cost overrun of the project are constructed with the discussion of experts of the Indian construction industry in the simple forms of the Mamdani-style. The calculated RII of the factors and clusters are taken as the weighting of the rules. Centre of area (COA) method is chosen for defuzzification. The model has been tested by taking into consideration a real case study in India. To test the model an interview was conducted with a team of the experts of a leading Indian construction company. The panel of the experts consisted of the top 10 executive engineers, project managers and site engineers of the company. The experts were requested to examine the cost overrun factors considered for this research and filled in the required information regarding cost overrun factors to assess the proposed model. They were also requested to estimate the cost risk level of the project. The probability of cost overrun of the project was calculated by the experts was in the range of 0.35-0.50. Provided information was tested using the proposed model and The probability of cost overrun of the project obtained was 0.435, which is much closer to the estimated cost risk level of the project estimated by the panel of the experts of the company. With the help of the proposed model, it is possible to guide project managers and decision makers to make more informative decisions such as contingency estimation, mark-up estimation, bid price, selection of optimum procurement route, evaluation of different projects and preparation of project portfolios.

VI. LIMITATION OF APPROACH AND FUTURE SCOPE

The following are the limitations of the proposed model: The sample size of population for the identification cost overrun factors is very small as compared to the size of construction industry. Therefore the calculated the relative importance index (RII) does not represent a true picture of actual cost overrun scenario. The model can differ as per the shapes and number of membership function adopted in fuzzy logic tool. The constructed rules for fuzzy model are based on expert opinion. It is based on subjectivity. For different attitudes of the experts rules can also be changed and in this condition the model will produce different result. Results can vary according to the defuzzification method adopted for obtaining output. Fuzzy rules are based on literature survey findings expert judgment and, it is clear that a completely different model may be proposed by other researchers based on different expert opinions. The model could also be developed for estimating the performance and delay probability in construction and other industrial projects using fuzzy logic theory.

Conflict of Interest. No.

REFERENCES

[1]. Hajarat, A.D., & Smith J. N. (1993). Exposure envelopes: an assessment of the exposure to time and cost overruns during construction projects. *Int. J. Proj. Manag., 11*(4), 227–231.

[2]. Azhar, N., Farooqui R.U., & Ahmed, S.M. (2008). Cost overrun factors in construction industry of Pakistan. In: Paper presented at the first international conference on construction in developing countries (ICCIDC-I). Advancing and integrating construction education, research & practice.

[3]. Memon, A.H., Rahman, I.A., Asmi, A., & Azis, A. (2010). Factors Affecting Construction Cost in Mara Large Construction Project: Perspective of Project Management Consultant. *International Journal of Sustainable Construction Engineering & Technology, 1*(2), 41-54.

[4]. Akinci, B., & Fischer, M. (1998). Factors affecting contractors' risk of cost overburden, *Journal of Management in Engineering, 14*(1), 67-76.
[5]. Reina, P., & Angelo, W. J. (2002). Megaprojects need more study up front to avoid cost overruns. *ENR*, *249*(3).
[6]. Frame, J. D. (1997). Establishing project risk assessment teams. In: K. Kähkönen & K. Artto (eds.), Managing risks in projects, London: E & FN Spon, 17-19.
[7]. Jackson, S. (2002). Project cost overrun and risk management in Greenwood. In: Proceedings of Association of Researchers in Construction Management, 18th Annual ARCOM Conference, Newcastle, Northumber University, UK, 2-4.
[8]. Chang, A.S.T. (2002). Reasons for Cost and Schedule Increase for Engineering Design Projects. *Journal of Management in Engineering, 18*(1), 29–36
[9]. Olawale, Y.A., & Sun, M. (2010). Cost and time control of construction projects: Inhibiting factors and mitigating measures in practice. *Construction Management and Economics, 28*(5), 509–526.
[10]. Cantarelli, C.C., Flyvbjerg B., Molin, E.J.E., & Wee, B. V. (2010). Cost *Overruns* in Large-Scale Transportation Infrastructure Projects: Explanations and Their *Theoretical* Embeddednes. *European Journal of Transport and Infrastructure Research*, *10*(1), 5-18.
[11]. Koushki P.A., Al-Rashid, K., & Kartam, N. (2005). Delays and cost increases in the construction of private residential projects in Kuwait. *Construction Management and Economics*, *23*(3), 285–294.
[12]. Gupta, N. (2009). Avoiding Time and Cost Overruns in the Construction of Rohtang Tunnel. Institute for Defence Studies and Analyses.
[13]. Frimpong, Y., Oluwoye, J., & Crawford, L. (2003). Causes of delays and cost overruns in construction of groundwater projects in developing countries; Ghana as a case study. *International Journal of Project Management, 21*, 321-326.
[14]. KPMG (2019). Study on Assessment of infrastructure projects and corrective recommendations for performance. A joint study conducted by PMI and KPMG in India on infrastructure projects in India, supported by Ministry of Statistics and Programme Implementation (MoSPI).
[15]. Zhi, H. (1995). Risk management for overseas construction projects. *International Journal of Project Management, 13*(4), 231-237.
[16]. Baccarini D., & Archer, R. (2001). The risk ranking of projects: a methodology. *International Journal of Project Management 19*(3), 139-145.
[17]. Hull, J. (1990). Application of risk analysis techniques in proposal assessment. *International Journal of Project Management, 8*(3), 152-157.
[18]. Özta, A, & Ökmen, Ö (2004). Risk analysis in fixed-price design-build construction projects. *Building and environment* 39(2)**,** 229-237.
[19]. Thomas, A., Kalidindi, S., & Ganesh, L. (2006). Modelling and assessment of critical risks in BOT road projects. *Construction Management and Economics, 24*(4)**,** 407-424
[20]. Dey P K (2001). Decision support system for risk management: a case study. *Management Decision, 39*(8): 634-649.
[21]. Dikmen, I., Birgonul, & M. T. (2006). An analytic hierarchy process based model for risk and opportunity assessment of international construction projects. *Canadian Journal of Civil Engineering, 33*(1): 58-68.
[22]. Nasir, D., McCabe, B., & Hartono, L. (2003). Evaluating Risk in Construction–Schedule Model (ERIC–S): Construction Schedule Risk Model. *Journal of Construction Engineering and Management, 129*, 518.
[23]. Kangari, R., & Riggs, L. S. (1989). Construction risk assessment by linguistics. *IEEE transactions on engineering management*, *36*(2), 126-131.
[24]. Tah, J., & Carr, V. (2000). A proposal for construction project risk assessment using fuzzy logic. *Construction Management and Economics, 18*(4), 491-500.
[25]. Sharma, S., & Goyal, P. K. (2014). Cost overrun factors and project cost risk assessment in construction industry- a state of the art review. *International journal of civil engineering, 3*(3), 139-154.
[26]. Oliveros, A. V. O., & Fayek, A. R. (2005). Fuzzy logic approach for activity delay analysis and schedule updating. *J. Constr. Eng. Manage*, *131*(1), 42–51.
[27]. Dikmen, I., Birgonul, M.T., & Han, S. (2007b). Using fuzzy risk assessment to rate cost overrun risk in international construction projects. *International Journal of Project Management, 25*(5), 494-505.
[28]. KarimiAzari, A., Mousavi, N., Mousavi, S. F., & Hosseini, S. (2011). Risk assessment model selection in construction industry. *Expert Systems with Applications*, *38*(8), 9105-9111.
[29]. Nieto-Morote, A., & Ruz-Vila, F. (2011). A fuzzy approach to construction project risk assessment. *International Journal of Project Management, 29*, 220–231.
[30]. Yeung, J., Chan, A., & Chan, D. (2012). Fuzzy set theory approach for measuring the performance of relationship-based construction projects in Australia. *J. Manage. Eng.,* 10.1061/(ASCE)ME.1943-5479 .0000083, 181–192.
[31]. San Cristobal, J. (2013). Critical path definition using multi criteria decision making: The PROMETHEE method. *J. Manage. Eng.,* 10.1061/(ASCE)ME.1943-5479.0000135, 158–163.
[32]. Gunduz, M., Nielsen, Y., & Ozdemir, M. (2015). Fuzzy Assessment Model to Estimate the Probability of Delay in Turkish Construction Projects. *Journal of Management in Engineering, ASCE,* DOI: 10.1061/(ASCE)ME.1943-5479.0000261
[33]. Shaktawat, A., & Vadhera, S. (2016, December). Fuzzy logic based determination of cost overrun of hydro power plant. In *2016 International Conference on Electrical Power and Energy Systems (ICEPES)* (pp. 301-304). IEEE.
[34]. Khalek, H. A. E., Aziz, R. F. & Kamel, H. M. (2016). Risk and Uncertainty Assessment Model in Construction Projects Using Fuzzy Logic. *American journal of civil engineering, 4*(1), 24-39.
[35]. Manoliadis, O. (2018). The Assessment of Construction Project Risks with the Use of Fuzzy Delphi Methodology- Case Study Photovoltaic Project in Greece. *International Journal of Management and Fuzzy Systems, 4*(1), 7-14.
[36]. Fuzzy Logic Toolbox™ 2 User's Guide (2008), MATLAB, The MathWorks, Inc.
[37]. Zadeh, L.A. (1965). Fuzzy sets. *Information and control, 8*(3), 338-353.
[38]. Mamdani, E. H., & Assilian, S. (1975). An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, *7*, 1-13.
[39]. Mittal, H., & Bhatia, P. (2007). Optimization criteria for effort estimation using fuzzy technique. *CLEI Electronic Journal*, *10*(1), 1-11.
[40]. Sharma, S. and Goyal, P. K. (2019). Applying "Fuzzy Techniques" in construction project management. *International Journal on Emerging Technologies*, *Research Trend, 10*(2): 384-391.

How to cite this article: Sharma, Savita, Goyal, Pradeep K. and Chhipa, R. C. (2020). Forecasting the Probability of Cost Overrun Risk of Indian Construction Projects using Fuzzy Model. *International Journal on Emerging Technologies*, *11*(1): 10–22.

View publication stats

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Fuzzy Logic based Automatic Generation Control

MD Shadan Salik
Electrical Engineering
Delhi Technological University
Delhi, India
shadansaik@gmail.com

Shaurya Priyal
Electrical Engineering
Delhi Technological University
Delhi, India
shauryapriyal1312@gmail.com

Neha Khokher
Electrical and Electronics Engineering
Delhi Technological University
Delhi, India
nehakhokher9@gmail.com

DR. Narendra Kumar
Electrical Engineering
Delhi Technological University
Delhi, India
narendra.kumar@dce.edu

***Abstract*— Automatic Generation Control (AGC) is a controlling technique used in electrical power generation to maintain a constant nominal frequency and power interchange with other linked systems by creating a balance of the generation of the system with system load losses. The frequency of the system deviates from its desired value if there is any sudden load change. If the frequency deviates in a large amount the system may get collapsed. Hence, a fast and accurate controller action is required to maintain stability and nominal frequency. For AGC many controlling techniques have been studied. In this paper, a PI controller is compared with a fuzzy controller. A two area system is considered both being thermal. Different step load changes have been considered and the performance of both the controllers has been observed in different load conditions. Simulation results have been brought up to show the better working of the fuzzy controller than the PI controller.**

Keywords— AGC, area control error, fuzzy logic controller, pi controller, two area system

I. INTRODUCTION

The purpose of electrical power generating units is to generate a certain amount of power that fulfills the required amount of load in a specific region. These generating units are referred to as power plants. There are different types of plants like a thermal plant, hydro plant, nuclear plant, etc. The load conditions do not remain the same throughout the time. It changes very rapidly and not only in one direction but in both directions that is it may go higher or lower than the normal demand. It depends on the consumers how much they are utilizing the load. Due to these variations in load demand the system frequency gets affected. It deviates from the nominal frequency. The deviation of frequency from its nominal value is undesirable as it affects the speed of the machine. The speed of a synchronous machine which is used as in generating units directly depends on the frequency. Therefore, as the frequency fluctuates, the speed of machines also fluctuates due to which the output of the machine is inconsistent. Hence the frequency of the system must remain constant.

To maintain a constant frequency of the system a control scheme is known as Automatic Generation Control (AGC) is used that controls the amount of the steam entering the turbine [1]. The primary purpose of the controller is to set the frequency of the system uniform whatever the changes in load demand may be [7]. As the load demand changes the frequency suddenly increases or decreases, immediately the current frequency is compared with a reference frequency or nominal frequency and according to it an error signal is sent to the controller which the takes adjust the amount of steam valve to be open or closed. Certain limits of load fluctuation are controlled by the controller or AGC but if load fluctuations are very high then the controllers need to be reset either manually or automatically [1]. The controller has a very vital role in AGC, and its proper and fast functioning is a must.

There are a lot of controllers that are used like P, PI, PID, fuzzy, genetic algorithm, PSO, etc [8-10]. Different controllers have different advantages. Many works have been followed out for inspecting different controller's behavior by considering a certain load change in a certain area. In this paper different performance of conventional PI controller has been contrasted with the fuzzy logic controller. A two area system is simulated in MATLAB Simulink. Different load changes are given to area 1 of the two area system. First taking the PI controller and then fuzzy controller the system is simulated and frequency variations and power variations are plotted. These plots are presented to show a comparative analysis of both the controllers.

II. CONTROL AREA CONCEPT

A power system in which a common generation control technique is used is known as a control area [2],[9]. Many control areas are attached through tie-lines [3]. In a control area, the action of a controller is to regulate the amount of inlet of the steam valve so that the amount of steam input to the turbine is regulated. The controller uses Area Control Error to bring the frequency at a nominal value [4].

Area Control Error (ACE) is the variation between actual and scheduled power generation in a control area considering frequency bias setting [1],[2]. In this paper two area thermal-thermal system is considered.

Mathematically ACE is given as:

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:53:05 UTC from IEEE Xplore. Restrictions apply.

For area 1: $ACE_1 = \Delta P_{12} + B_1 \Delta f_1$
For area 2: $ACE_2 = \Delta P_{21} + B_2 \Delta f_2$

where:

ΔP_{12} and ΔP_{21} denotes the departures from the scheduled interchange
B_1 and B_2 denotes frequency bias coefficient of area 1 and area 2 individually.
Δf_1 and Δf_2 area frequency deviation of area 1 and area 2 individually.

III. TWO AREA SYSTEM

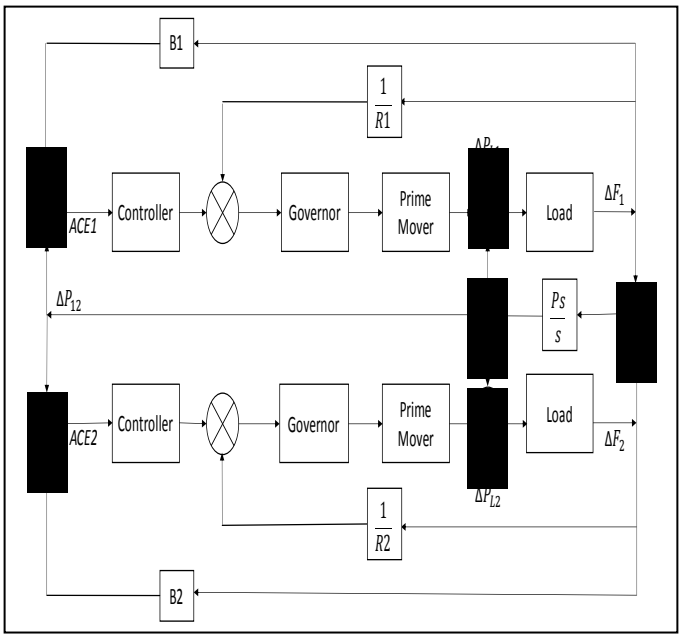


Fig 1: Two area thermal-thermal system block diagram

As shown in the above figure each area of the two-area system has its own controller and both are connected through a tie-line. Load changes (ΔP_L) may occur in any of the area which results in the frequency variations of area 1 i.e. ΔF_1 and in area 2 i.e. ΔF_2.
In this paper load changes of 0.1 per unit (pu), 0.2 pu and 0.3 pu have been considered in area 1. The controllers are first taken as PI controller and then fuzzy logic controller and then the system is simulated in MATLAB Simulink.

A. Parameters used

Time constant of turbine= 0.5 s
Time constant of governor = 0.2 s
Inertia constant of governor = 5 s
Speed regulation of governor= 0.05
Rated output of turbine = 250 MW

IV. FUZZY LOGIC CONTROLLERS

The fuzzy system is one of the key agents of computational intelligence. Prof. A Zadeh proposed the idea of a fuzzy system in 1965. A Fuzzy system differs from conventional computing as conventional computing involves Boolean logic whereas a fuzzy system involves multivalued logic. It deals with uncertainties due to ambiguity, imprecision, and vagueness. Fuzzy logic is much more general than conventional logic systems. It is mostly used in the systems with uncertainties due to ambiguity, imprecision, vagueness ambiguity, randomness, partial truth, and approximation.
The fuzzy logic controller has three steps [5],[6]:

1. Fuzzification
2. Inference rule
3. Defuzzification

It takes input variables in the form of membership functions. Membership functions can be of various types like triangular, Gaussian, trapezoidal, etc. Here, a triangular membership function is used. Every input variable is assigned with a specific number of membership functions that contain values of certain limits. These values are assigned with some linguistic terms [6]. This step is known as fuzzification.
After fuzzification there comes the rule base of the fuzzy logic controller. In the rule base, rules are assigned formulated on the knowledge of the system. Outputs are also given linguistic terms and input variable mapping is done to assign a particular output linguistic variable following the given matching of input linguistic variables. The full rule base considered in the system is mentioned in the following section.
In the final step that is defuzzification, the output linguistic variables are changed to corresponding numbers which decide the controlling action accordingly.

A. Membership functions

The following variables are created using MATLAB fuzzy command with five membership functions each.

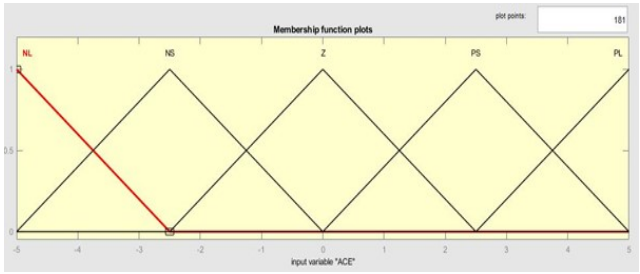


Fig 2: Input variable ACE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:53:05 UTC from IEEE Xplore. Restrictions apply.

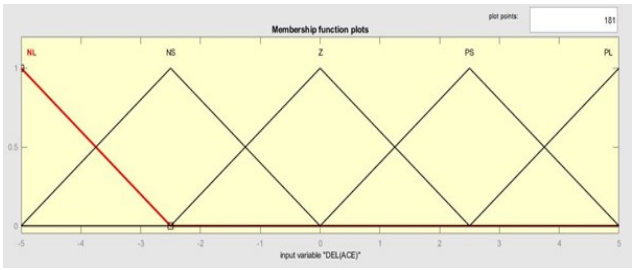


Fig 3: Input variable del(ACE)

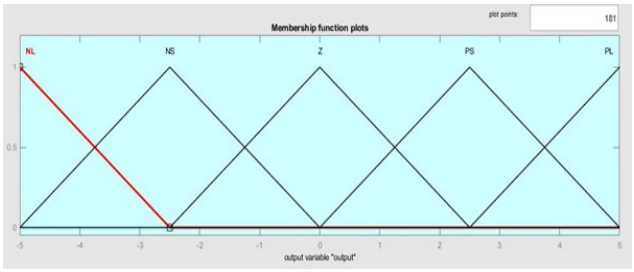


Fig 4: Output variable

B. Fuzzy Rule Base

The rules are assigned with the help of MATLAB fuzzy toolbox. Following table displays the rules that are set for input variables ACE and del(ACE) in order to give necessary output. Here the ACE and del(ACE) are used as input variables and assigned with values from -5 to 5 and given linguistic terms : More negative (NL), Less negative (NS), Zero (Z), Less positive(PS), More positive (PL).

TABLE I. RULE BASE [4]

		ACE				
		NL	**NS**	**Z**	**PS**	**PL**
DEL(ACE)	**NL**	NL	NL	NS	Z	Z
	NS	NL	NL	NS	PS	Z
	Z	NS	Z	Z	PS	PS
	PS	NS	Z	PS	PL	PL
	PL	Z	Z	PS	PL	PL

V. SIMULATION AND RESULTS

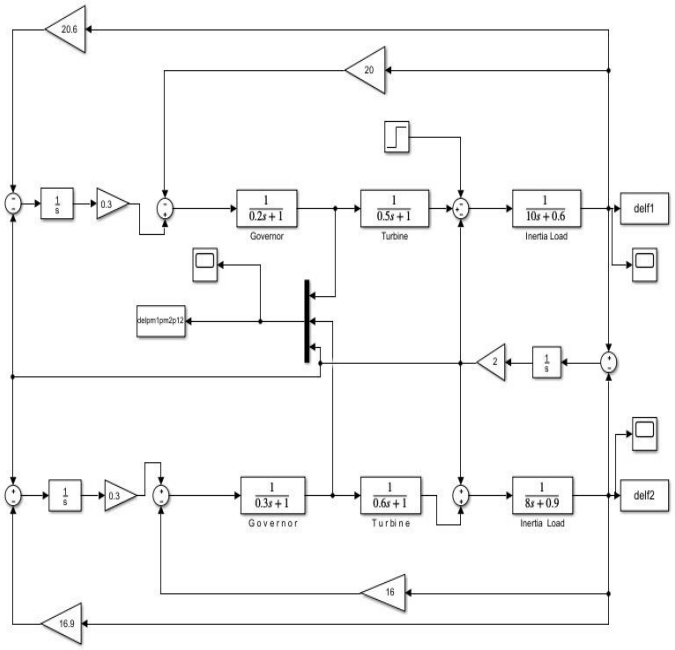


Fig 5: Two area system Simulink model by using PI control

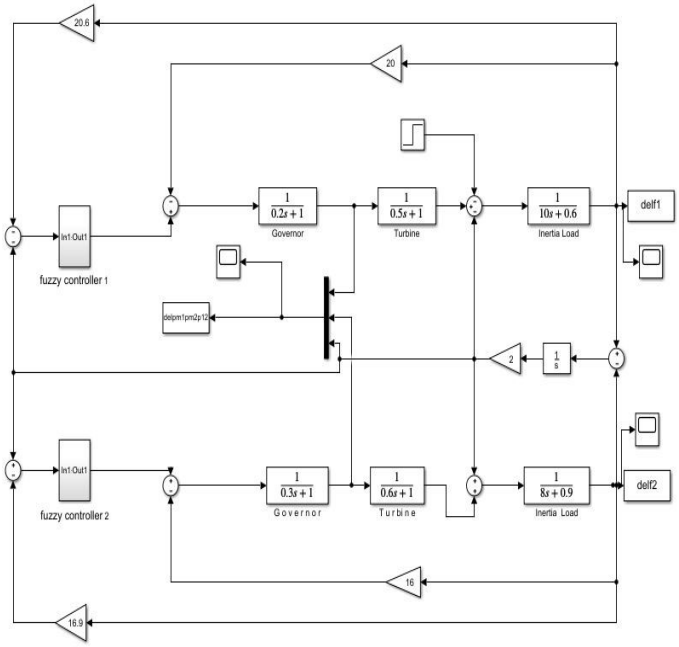


Fig 6: Two area system Simulink model by using fuzzy control

The following plots show the behavior of frequency as well as power fluctuations in both areas using a fuzzy controller and PI controller. For each load change, four plots are shown in the following sections.

A. Results for 0.1 pu load change

A change of 0.1 pu load is applied in area1 and the following results have been obtained.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:53:05 UTC from IEEE Xplore. Restrictions apply.

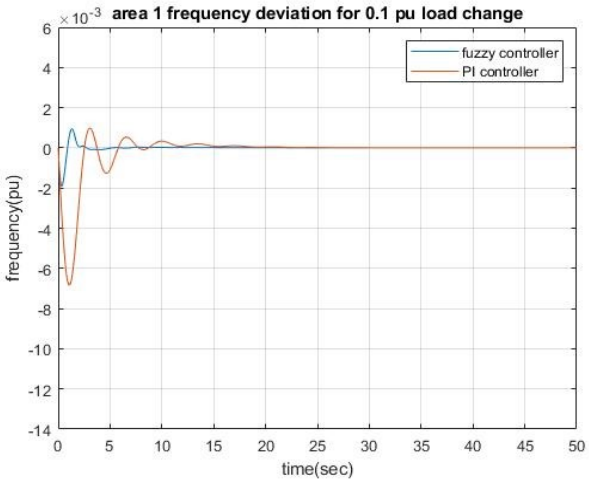


Fig 7: Area 1 frequency variations

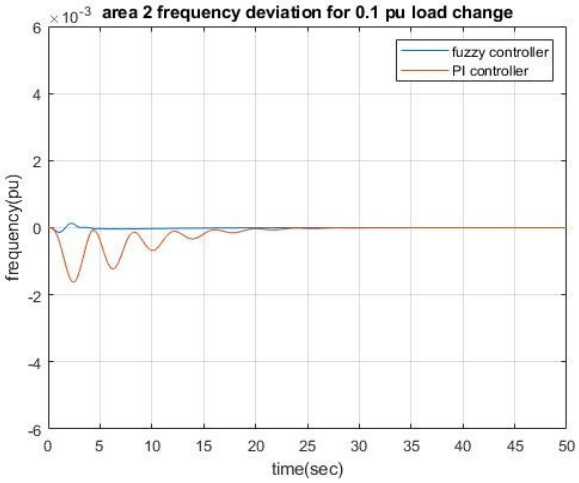


Fig 8: Area 2 frequency variations

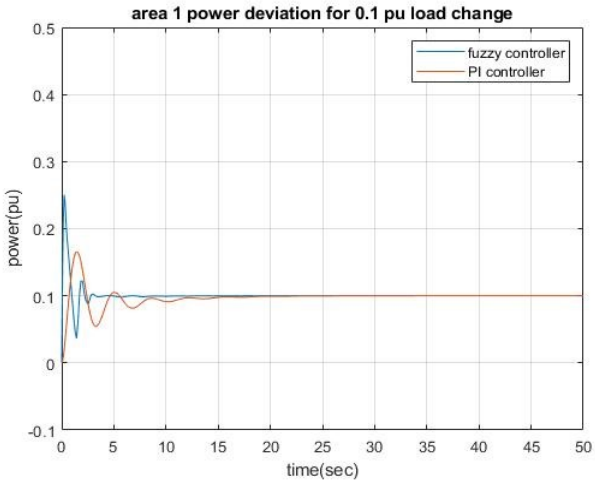


Fig 9: Area 1 power variations

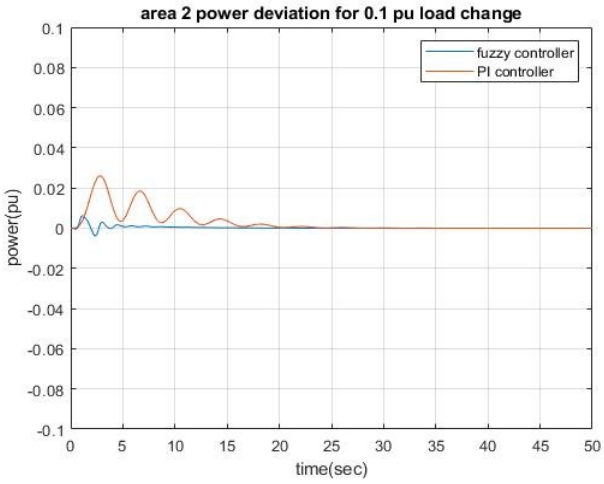


Fig 10: Area 2 power variations

B. Results for 0.2 pu load change

A change of 0.2 pu load is applied in area1 and the following results have been obtained.

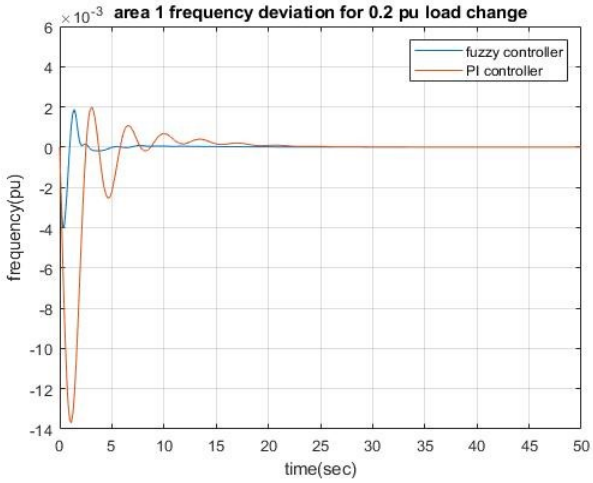


Fig 11: Area 1 frequency variations

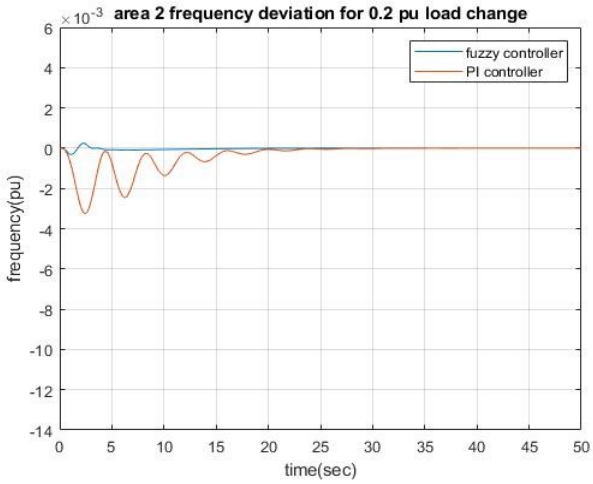


Fig 12: Area 2 frequency variations

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:53:05 UTC from IEEE Xplore. Restrictions apply.

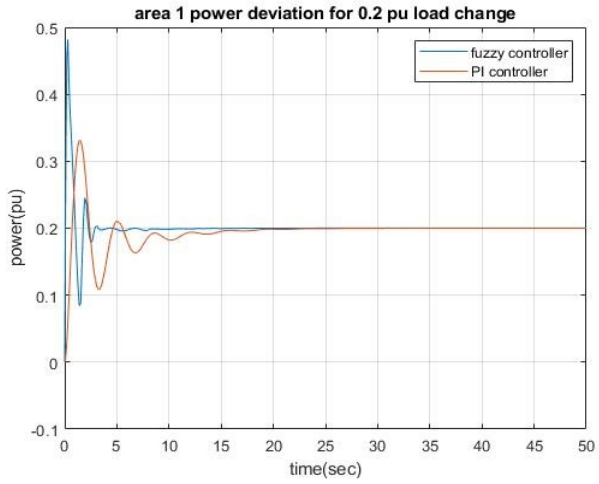


Fig 13: Area 1 power variations

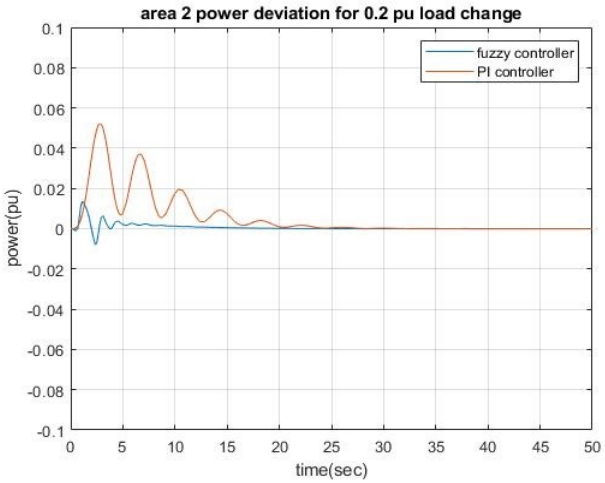


Fig 14: Area 2 power variations

C. *Results for 0.3 pu load change*

A change of 0.3 pu load is applied in area1 and the following results have been obtained.

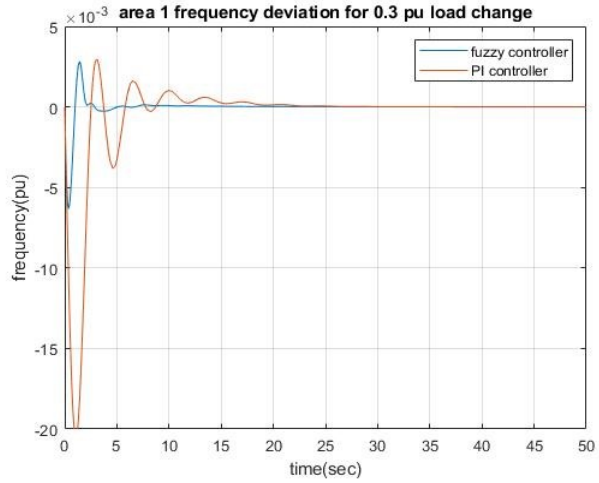


Fig 15: Area 1 frequency variations

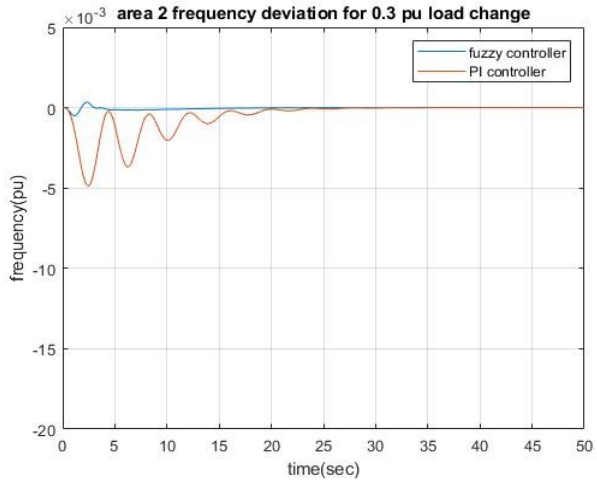


Fig 16: Area 2 frequency variations

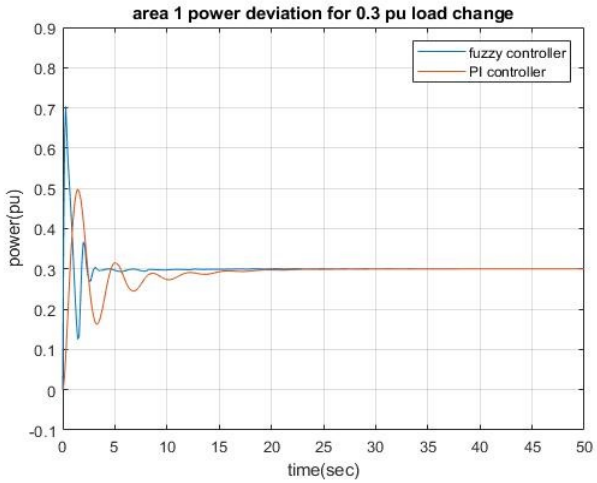


Fig 17: Area 1 power variations

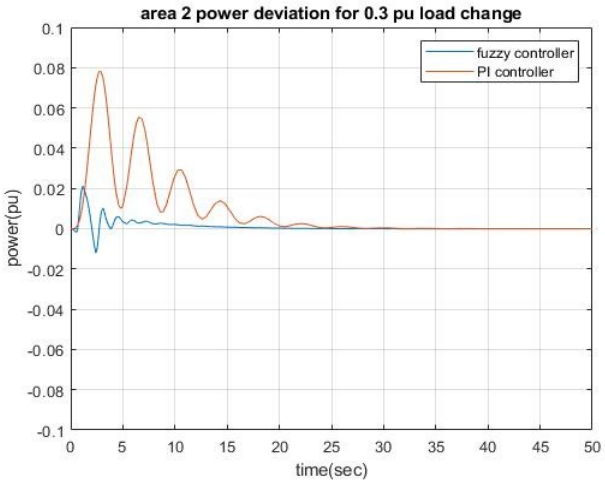


Fig 18: Area 2 power variations

From all the above plots the settling time and maximum deviation have been calculated with the help of MATLAB and following table is prepared which shows settling time and maximum fluctuation of frequency change in area 1 (ΔF1), frequency change in area 2 (ΔF2), power change in area 1 (ΔP1), power

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:53:05 UTC from IEEE Xplore. Restrictions apply.

change in area 2 (ΔP2) for all three load changes.

TABLE II. COMPARATIVE RESULT OF SETTLING TIME AND MAXIMUM DEVIATION OF FREQUENCY AND POWER

1% change	PI		Fuzzy	
	Settling time(sec)	max fluctuation (pu)	Settling time(sec)	max fluctuation (pu)
ΔF1	18.3083	0.0068	9.1211	0.0037
ΔF2	23.3675	0.0016	14.6863	0.000345
ΔP1	18.124	0.1647	11.7476	0.2463
ΔP2	23.8816	0.0261	13.8683	0.002
2% change	PI		Fuzzy	
	Settling time(sec)	max fluctuation (pu)	Settling time(pu)	max fluctuation (pu)
ΔF1	22.2716	0.0136	10.2409	0.0078
ΔF2	27.3657	0.0032	17.8777	0.0012
ΔP1	23.041	0.3282	12.7378	0.4925
ΔP2	26.8495	0.0501	16.4937	0.0039
3% change	PI		Fuzzy	
	Settling time(sec)	max fluctuation (pu)	Settling time(pu)	max fluctuation (pu)
ΔF1	23.3151	0.0201	13.4343	0.0075
ΔF2	31.3705	0.0049	18.7375	0.0017
ΔP1	22.012	0.4907	13.7400	0.7087
ΔP2	27.8761	0.0778	19.6394	0.0065

VI. CONCLUSION

By observing all the plots of frequency and power deviations of both the areas using two controlling techniques for a particular load change the frequency returns to its nominal value in lesser time in case of the fuzzy logic controller in contrast to the traditional PI controller and also power deviations are less in case of fuzzy logic controller. Moreover, due to the increased the amount of step load change, the deviation of transients of frequency and power in the conventional controller are increased much before coming to its nominal value whereas in the case of the fuzzy logic controller the deviations are not much increased by increasing the load changes. Hence fuzzy logic is a more reliable and effective controller as compared to the traditional PI controller.

LIMITATIONS AND FUTURE SCOPE

In the fuzzy controller, a triangular membership function is used. In place of triangular other membership functions can be considered like trapezoidal, Gaussian, etc. In place of fuzzy control more advanced artificial intelligence techniques can be used.

REFERENCES

[1] Saadat, H., 1999. *Power system analysis*. WCB/McGraw-Hill

[2] Wood, A.J., Wollenberg, B.F. and Sheblé, G.B., 2013. *Power generation, operation, and control*. John Wiley & Sons..

[3] Subbaraj, P. and Manickavasagam, K., 2008. Automatic generation control of multi- area power system using fuzzy logic controller. *European transactions on electrical power*, *18*(3), pp.266-280.

[4] Chandrashekar, M.J. and Jayapal, R., 2013, December. Design and comparison of I, PI, PID and Fuzzy logic controller on AGC deregulated power system with HVDC link. In *2013 International conference on Circuits, Controls and Communications (CCUBE)* (pp. 1-6). IEEE.

[5] Jain, S.K., Bhargava, A. and Pal, R.K., 2015, September. Three area power system load frequency control using fuzzy logic controller. In *2015 International Conference on Computer, Communication and Control (IC4)* (pp. 1-6). IEEE.

[6] Prakash, S., Sinha, S.K., Pandey, A.S. and Singh, B., 2009. Impact of slider gain on load frequency control using fuzzy logic controller. *ARPN J. Eng. Appl. Sci*, *4*(7), pp.20-7.

[7] Çam, E. and Kocaarslan, I., 2005. Load frequency control in two area power systems using fuzzy logic controller. *Energy conversion and Management*, *46*(2), pp.233-243.

[8] Mathur HD, Manjunath HV. Extended fuzzy logic based integral controller for three area power system with generation rate constraint. In2006 IEEE International Conference on Industrial Technology 2006 Dec 15 (pp. 917-921). IEEE.

[9] AnilKumar A, Srikanth NV. Load frequency control for diverse sources of interconnected two area power system: An adaptive fuzzy approach. In2016 International Conference on Control, Computing, Communication and Materials (ICCCCM) 2016 Oct 21 (pp. 1-4). IEEE.

[10] Moon YH, Ryu HS, Kim B, Kim S, Park SC. Fuzzy logic based extended integral control for load frequency control. In2001 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No. 01CH37194) 2001 Jan 28 (Vol. 3, pp. 1289-1293). IEEE.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:53:05 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Generating Visible Spectrum Images from Thermal Infrared using Conditional Generative Adversarial Networks

1st Neeraj Bhat
Dept. of Computer Science and Engineering
Delhi Technological University
New Delhi, India
neerajbhat98@gmail.com

2nd Navneet Saggu
Dept. of Computer Science and Engineering)
Delhi Technological University
New Delhi, India
navneetbt2k16@dtu.ac.in

3rd Pragati
Dept. of Computer Science and Engineering
Delhi Technological University
New Delhi, India
pragatisingh7890@gmail.com

4th Sanjay Kumar
Dept. of Computer Science and Engineering
Delhi Technological University
New Delhi, India
sanjay.kumar@dtu.ac.in

***Abstract*—Thermal Infrared cameras have a unique set of capabilities that enables them to see and capture images in situations where vision is severely obscured, including dark, shade, and foggy conditions. Conversion of TIR image to visible spectrum makes them more interpretable, but automating transformation is a challenging task as it requires estimation of both chrominance and luminance for each pixel, posing the need for a more complex model that is capable of capturing higher-level semantics of a TIR image to generate a perceptually realistic RGB image.The use of Conditional Generative Adversarial Networks (cGANs) is proposed with a combination of content and adversarial losses to tackle this problem. The proposed cGAN uses a U-Net auto-encoder network as generator, and a patch based discriminator for adversarial training. The cGAN learns the most effective loss function during training, producing state-of-the-art perceptually realistic outputs. Qualitative and quantitative analyses show that our approach outperforms existing methods.**

***Index Terms*—Conditional Generative Adversarial Networks, Deep Learning, Image to Image Transformation, Thermal Infrared Imaging**

I. Introduction

Thermal infrared imaging technology originally found applications in the fields of Military and Intelligence [8]. Thermal infrared cameras capture images and videos in the thermal spectrum, which is invisible to the human eye. Infrared radiation is reflected and emitted by all objects, even in the dark. Thus, thermal infrared cameras can capture images in the dark, and in conditions where visibility is obscured e.g. foggy conditions. In recent years, the technology has significantly improved. Resolution of thermal sensors has increased, and thermal infrared cameras have become less expensive. This has led to the application of thermal infrared cameras in entirely new fields like; automotive computer vision systems, surveillance systems, medical imaging and specialized equipment for various scientific and industrial applications including archaeology, geology, mining etc. Other significant advantages of thermal cameras are that they can capture images even in very low light, they are robust to extreme changes in light and shade conditions, and they preserve the anonymity of the subjects being captured in surveillance systems. As the use of thermal infrared cameras becomes more common, there is a growing need to develop efficient and accurate methods to transform thermal infrared (TIR) images to visible spectrum (RGB) images. Perceptually realistic visible spectrum (RGB) images are more intuitive to interpret and easier to understand for the untrained human eye, and will help improve the viewing performance of the human operator that operates a thermal infrared imaging system.

Transformation of black and white (grayscale) images into colored (RGB) images is a similar problem that has been well researched before [2]–[6]. Autoencoder networks have been used to obtain impressive results in grayscale colorization e.g. the "Colorful Image Colorization" [19] and "Let there be color!" [18]. When colorizing grayscale inputs, a given pixel of the input grayscale image can correspond to many different colors in the colorized output. As each pixel in a grayscale image only captures brightness (luminance), the color (chrominance) of the output remains ambiguous. An input grayscale pixel intensity can map to an infinite number of combinations of red, green and blue intensity values that produce the same luminance as the input but output very different colors. This is known as chrominance ambiguity in colorization methods. The problem of transforming thermal infrared images to visible spectrum images poses an entirely new challenge. A thermal infrared image captures intensities of infrared radiation only. There is no way to estimate visible spectrum intensities knowing only the thermal infrared intensi-

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:49:24 UTC from IEEE Xplore. Restrictions apply.

ties, as they both lie in different radiation spectra. Therefore, it has been required to resolve not only chrominance ambiguity, but also luminance ambiguity for generating a perceptually realistic visual spectrum image from thermal infrared input. Colorization methods have previously been researched that utilise autoencoder Convolutional Neural Networks (CNN), as they are able to extract a semantic representation of input and then generate an appropriate output. This approach was used by Berg et al. (TIR2Lab) [17] for thermal infrared colorization and Limmer et al. [20] for near infrared colorization. However, the images generated using this approach were distorted and blurred Therefore, a more complex system is needed that is able to generate the visible spectrum image from a more abstract representation of the input image.

II. Method

The use of a Conditional Generative Adversarial Network (cGAN) [16] is proposed for generating visible spectrum images from thermal infrared input. This is an image to image transformation problem, a class of problems which has previously been approached using autoencoder networks. Such networks must train on carefully hand engineered loss functions to work effectively. An ill-engineered loss will eventually result in blurry and unrealistic output, which is a common problem in autoencoder models. The proposed Conditional Generative Adversarial Network architecture not only learns the input-output mapping, but also learns the loss function that trains the complete network resulting in clearer and more realistic looking output image.

A. Conditional Generative Adversarial Networks

Conditional Generative Adversarial Networks provide an architecture for effective training of generator networks to learn a conditional mapping from given input image, to a target output image. The generator, a deep convolutional neural network, first extracts the relevant semantic features from the input infrared image, and generates colorized output from that abstract semantic representation. This generator network trains on a combination of adversarial and content losses, in parallel with a discriminator, using the cGAN architecture. Our cGAN was trained on a large dataset (50,184 images) of pairs of thermal infrared and visible spectrum images.

A cGAN consists of two primary components, a generator network (G), and a discriminator network (D). The generator model is an autoencoder network that maps the input image to its corresponding output. With each training iteration, it adjusts it's weights to produce a more realistic image, that is closer to the actual output provided in the dataset. The discriminator model is typically a convolutional neural network that is trained to correctly identify original output images (output images sampled from the dataset) from fake output images (output images generated using the generator model). Unlike in traditional GANs, in cGANs, the discriminator receives both the input image and the target image as input, and it predicts whether the output is a valid transformation of the given input. This aids in learning a conditional mapping between the input and the output. The training process of a GAN is a min-max game, with an objective function as given in eq. 1

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data(x)}}[\log D(x)] + \mathbb{E}_{x\sim p_z(x)}[\log(1 - D(G(x))] \quad (1)$$

Here, G and D are the generator and discriminator network respectively. $D(x)$ represents the discriminator prediction for the given input image (x), $D(G(x))$ represents the discriminator prediction for the generated image ($G(x)$) for the given input image. With the objective of averting detection by generating more realistic outputs, the generator aims to increase the discriminator's classification error, while the discriminator trains to better classify real and fake samples, decreasing the classification error. This min-max game drives the training of a GAN, and adversarial competition trains the generator to produce more genuine looking images.

B. Network Architecture

The generator of the cGAN architecture proposed is an autoencoder network with skip connections that has previously been used for image segmentation and image synthesis (U-Net) [15]. Our discriminator is a convolutional neural network that classifies 70x70 sized patches of the image as real or fake (PatchGAN), producing a tensor as output.

1) Generator Architecture: U-Net consists of a contracting encoder network (left) and an expanding decoder network (right)(see Fig. 1). The encoder network uses convolution operation to downsample the image and obtain an abstract representation of the image at the bottleneck. The decoder network then generates the output image from this feature map by means of upsampling.

2) Encoder: The encoder resembles most Convolutional Neural Networks, with added skip connections to aid in direct transfer of low level information directly from initial layers towards the output layers. The entire encoder section consists of 8 stages. With each encoding stage, two convolutions using a 3x3 filter is applied. The activation function used is Rectified Linear Unit (ReLU). The convolution layer is followed by a 2x2 max pooling layer that has a stride of 2 for downsampling. This downsampling is what lends the encoder network it's contracting path shape. With each stage, the number of channels is doubled at the downsampling step.

3) Decoder: The decoder network consists of 8 decoding stages. Each stage of the decoder applies a simple 2D up sampling operation to the feature map, expanding the map. After that, a 2x2 up convolution is applied, that halves the number of channels in the feature map. A skip connection then concatenates the resulting feature tensor with the output from a corresponding encoding stage. This is followed by application of two 3x3 convolutions, using Rectified Linear

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:49:24 UTC from IEEE Xplore. Restrictions apply.

Unit (ReLU) activation.

4) Discriminator Architecture: The PatchGAN discriminator [9] is another unique element of the proposed architecture (Fig. 2). The function of the discriminator is to correctly classify the given images as being original or fake. Our discriminator classifies patches of the image of NxN size, instead of making predictions for the entire image at once. The primary benefit of using a patch based discriminator is that there are fewer parameters to train in the discriminator network, resulting in faster training. Using a patch based discriminator also encourages sharper detail in generated images [10], [11]. When using a Patch discriminator, the image is essentially modeled as a Markov field, where pixel values are separated by more than a span of N can be considered to be conditionally independent. Through extensive parameter tuning, it has been discovered that a patch size of 70x70 pixels is optimal, making fast discriminator predictions without much loss in image quality. With a patch size of 70x70, there are 900 patches of an output image to predict, resulting in an output tensor of 30x30.

C. Objective Functions

The use of a combination of per-pixel content loss, and the cGAN adversarial loss, are used to train the network. Through the adversarial loss, the high level semantic structure of the input image can be captured, while content loss aids in the capture of low frequency details in the input.

1) Content Loss: This is a per-pixel loss. It evaluates the average difference between each pixel of the target image and the generated image, thus effectively minimizing the per-pixel error in generated images. The L1 content loss is used as it has been observed that use of L2 content loss tends to generate blurry output images.

$$L_{content} = \frac{1}{WH}\sum_{i=1}^{W}\sum_{j=1}^{H}||y_{i,j} - G(x)_{i,j}||_1 \quad (2)$$

where W and H denote the height and width of the image respectively, and $||.||_1$ denotes the L1 norm. $y_{i,j}$ and $G(x)_{i,j}$ represent the pixel values for the sample output image(y) and the generated image ($G(x)$) respectively.

2) Adversarial Loss: Adversarial loss in a cGAN can be expressed as:

$$L_{adversarial}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(G(x), z))] \quad (3)$$

Equation 3 represents a quantitative measure of the classification error of the discriminator. G, the generator network, and D, the discriminator network are involved in a two player min-max game, with the objectives of respectively maximising and minimising this classification error.

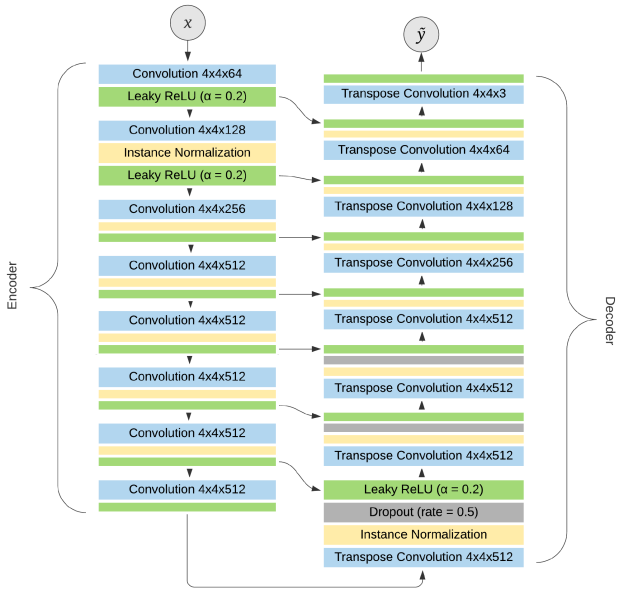


Fig. 1. Generator Architecture

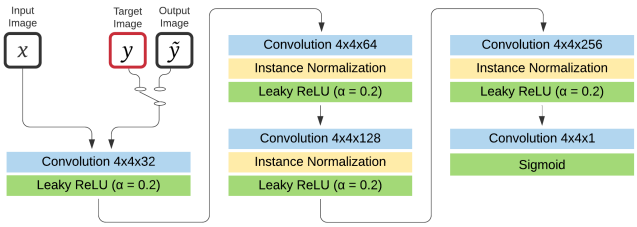


Fig. 2. Discriminator Architecture

3) Final Objective: The final objective function is a weighted sum of the content loss and the adversarial loss.

$$L_{total} = \lambda_{content} L_{content} + \lambda_{adversarial} L_{adversarial} \quad (4)$$

After experimenting and fine tuning parameters on the training data-sets, the optimal values of the weights are obtained as: $\lambda_{content} = 1, \lambda_{adversarial} = 100$

III. Experiments

The model is trained on the publicly available KAIST-MS dataset [7], which is benchmark dataset used for training and testing multi-spectral image transformation models. The dataset consists of 50,184 sample mappings of visible spectrum RGB images and their corresponding TIR images. All images were resized to 256x256 three channel RGB with a train-test split ratio of 80:20. The generator network was initialized with random weights, with a normal distribution $N(0, 0.02)$. The dimensions of the patch discriminator were set to 70x70 after fine-tuning. Training was done with a

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:49:24 UTC from IEEE Xplore. Restrictions apply.

learning rate of 0.0002, $\beta_1 = 0.5$ and $\beta_2 = 0.9$, using the Adam Optimizer. Publicly available Nvidia Tesla P100 (15.9 GB) GPUs were used to train the model on Kaggle.

IV. Results

There is no single widely accepted technique to evaluate the quality of computer generated images, and a variety of multiple methods and measures have been used by researchers working in this field. The quality of our results are evaluated using a number of such quantitative measures in IV-A, and also compared them with baseline results for a subjective qualitative analysis in IV-B.

A. Quantitative Analysis

The model is also compared to its contemporaries on the basis of relevant quantitative metrics. The metrics used will also include the structure similarity index (SSIM) [14], multiscale structure similarity index (MSSIM) [13] and peak signal-to-noise ratio (PSNR). Higher values indicate better quality of generated images. Comparing our method with the Naive and TIR2Lab methods, both trained on the same KAIST dataset, it has been observed that our model outperforms these baselines. Average scores for the compared methods are presented in table I.

1) Structure Similarity Index (SSIM): Structure Similarity Index is a measure used to estimate the perceived quality of cinematic movies and television picture, as well as various types of digital images and videos. The index is calculated on a number of windows of the image. The SSIM for two image windows p and q, having size $N \times N$ is:

$$SSIM(x,y) = \frac{(2\mu_p\mu_q + c_1)(2\sigma_{pq+c_2})}{(\mu_p^2 + \mu_q^2 + c_1)(\sigma_p^2 + \sigma_q^2 + c_2)} \quad (5)$$

Where μ_p is mean, σ_q is standard devation, and σ_{pq} is co-variance.

2) Peak Signal-to-Noise Ratio (PSNR): PSNR is a mathematical ratio between the maximum power of a given signal and the power of noise error. It is commonly used to measure the image reconstruction quality of lossy compression methods. PSNR is used to compare the quality of the generated RGB image with the actual visual spectrum image sample. A simple PSNR can be defined by mean squared error.

$$\frac{1}{WH}\sum_{x=1}^{W-1}\sum_{y=1}^{H-1}(P(x,y) - Q(x,y))^2 \quad (6)$$

Where the images P and Q are $W \text{x} H$ sized.

3) Multi-scale SSIM (MSSIM): Multi-scale SSIM calculates the structural similarity index by calculating the SSIM index for several differently scaled versions of an image, and then combining them. The MS-SSIM index is more robust than SSIM index.

TABLE I
AVERAGE RESULTS OF ALL THE TEST IMAGES OF THE KAIST DATA-SET.

Method	PSNR	SSIM	MSSIM
Naive	9.15	0.32	0.13
TIR2Lab	14.6	0.52	0.51
Proposed	**16.12**	**0.55**	**0.53**

B. Qualitative Analysis

Fig. 3 shows the results generated using our proposed model. Comparing them with the TIR2Lab baseline in Fig. 4, It can also be seen that the obtained results are highly detailed, and perceptually realistic. Our model is robust and correctly generates outputs for both day time, and night time inputs. This was ensured by training the cGAN on day and night images separately.

V. Conclusion

In this paper, a deep learning approach is proposed to solve the colorization problem of thermal infrared images. In particular, a multi-term objective function associated with conditional generative adversarial networks is used to learn the mapping for this image to image translation task. Our method has been evaluated on a publicly available dataset. Our qualitative results produce perceptually realistic images, while retaining the high level information of the given input. When tested on quantitative measures, our model gave a comparable SSIM score, as well as a better PSNR value. Both qualitative and quantitative analysis methods demonstrate that our proposed work outperforms all previous traditional approaches such as the use of a Convolutional Neural Network as an auto-encoder network.

References

[1] Ali, E. A., Qadir, H., Kozaitis, S. P. (2014, June). Color night vision system for ground vehicle navigation. In Infrared Technology and Applications XL (Vol. 9070, p. 90700I). International Society for Optics and Photonics.

[2] Cao, Y., Zhou, Z., Zhang, W., Yu, Y. (2017, September). Unsupervised diverse colorization via generative adversarial networks. In Joint European conference on machine learning and knowledge discovery in databases (pp. 151-166). Springer, Cham.

[3] Cheng, Z., Yang, Q., Sheng, B. (2015). Deep colorization. In Proceedings of the IEEE International Conference on Computer Vision (pp. 415-423).

[4] Deshpande, A., Lu, J., Yeh, M. C., Jin Chong, M., Forsyth, D. (2017). Learning diverse image colorization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 6837-6845).

[5] Gupta, R. K., Chia, A. Y. S., Rajan, D., Ng, E. S., Zhiyong, H. (2012, October). Image colorization using similar images. In Proceedings of the 20th ACM international conference on Multimedia (pp. 369-378).

[6] Huang, Y. C., Tung, Y. S., Chen, J. C., Wang, S. W., Wu, J. L. (2005, November). An adaptive edge detection based colorization algorithm and its applications. In Proceedings of the 13th annual ACM international conference on Multimedia (pp. 351-354).

[7] Hwang, S., Park, J., Kim, N., Choi, Y., So Kweon, I. (2015). Multispectral pedestrian detection: Benchmark dataset and baseline. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1037-1045).

[8] Gade, R., Moeslund, T. B. (2014). Thermal cameras and applications: a survey. Machine vision and applications, 25(1), 245-262.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:49:24 UTC from IEEE Xplore. Restrictions apply.

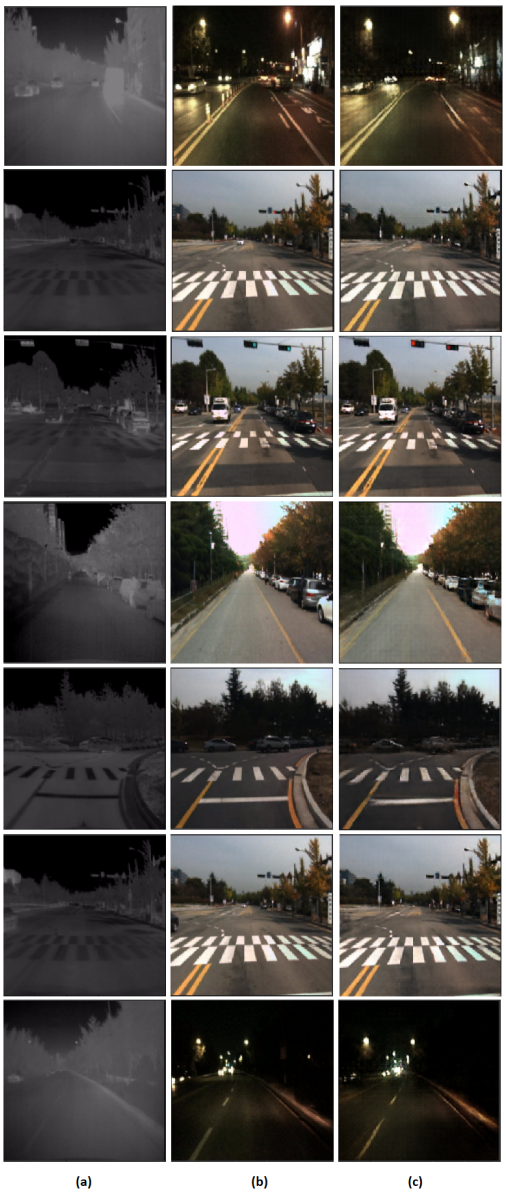


Fig. 3. Our generated results (a)Thermal Infrared Images, (b) Ground Truth(True RGB images), (c)Proposed



Fig. 4. Our results compared with TIR2Lab (a)Thermal Infrared Images, (b) Ground Truth(True RGB images), (c)TIR2Lab, (d)Proposed

[9] Li, C., Wand, M. (2016, October). Precomputed real-time texture synthesis with markovian generative adversarial networks. In European conference on computer vision (pp. 702-716). Springer, Cham.
[10] Gatys, L. A., Ecker, A. S., Bethge, M. (2016). Image style transfer using convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2414-2423).
[11] Li, C., Wand, M. (2016). Combining markov random fields and convolutional neural networks for image synthesis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2479-2486).
[12] Damera-Venkata, N., Kite, T. D., Geisler, W. S., Evans, B. L., Bovik, A. C. (2000). Image quality assessment based on a degradation model. IEEE transactions on image processing, 9(4), 636-650.
[13] Wang, Z., Simoncelli, E. P., Bovik, A. C. (2003, November). Multiscale structural similarity for image quality assessment. In The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003 (Vol. 2, pp. 1398-1402). Ieee.
[14] Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, 13(4), 600-612.
[15] Ronneberger, O., Fischer, P., Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.
[16] Mirza, M., Osindero, S. (2014). Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.
[17] Berg, A., Ahlberg, J., Felsberg, M. (2018). Generating visible spectrum images from thermal infrared. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 1143-1152).
[18] Iizuka, S., Simo-Serra, E., Ishikawa, H. (2016). Let there be color! Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. ACM Transactions on Graphics (ToG), 35(4), 1-11.
[19] Zhang, R., Isola, P., Efros, A. A. (2016, October). Colorful image colorization. In European conference on computer vision (pp. 649-666). Springer, Cham.
[20] Limmer, M., Lensch, H. P. (2016, December). Infrared colorization using deep convolutional neural networks. In 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 61-68). IEEE.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:49:24 UTC from IEEE Xplore. Restrictions apply.

GloVeInit at SemEval-2020 Task 1: Using GloVe Vector Initialization for Unsupervised Lexical Semantic Change Detection

Vaibhav Jain
Delhi Technological University
Delhi, India
vaibhav29498@gmail.com

arXiv:2007.05618v1 [cs.CL] 10 Jul 2020

Abstract

This paper presents a vector initialization approach for the SemEval2020 Task 1: Unsupervised Lexical Semantic Change Detection. Given two corpora belonging to different time periods and a set of target words, this task requires us to classify whether a word gained or lost a sense over time (subtask 1) and to rank them on the basis of the changes in their word senses (subtask 2). The proposed approach is based on using Vector Initialization method to align GloVe embeddings. The idea is to consecutively train GloVe embeddings for both corpora, while using the first model to initialize the second one. This paper is based on the hypothesis that GloVe embeddings are more suited for the Vector Initialization method than SGNS embeddings. It presents an intuitive reasoning behind this hypothesis, and also talks about the impact of various factors and hyperparameters on the performance of the proposed approach. Our model ranks 13th and 10th among 33 teams in the two subtasks. The implementation has been shared publicly.[1]

1 Introduction and Background

Lexical Semantic Change (LSC) Detection is an active research topic in the field of natural language processing, and has been applied for diachronic (across time) and synchronic (across domains) tasks (Schlechtweg et al., 2019). Previously limited to manual "close-reading" approaches, the availability of large-scale corpora have allowed the use of computational methods for this task (Tahmasebi et al., 2018). This topic has found various applications in various disciplines such as improving information retrieval from historical documents (Morsy and Karypis, 2016), preventing cross-domain ambiguity in requirements elicitation interviews (Jain et al., 2020), and studying the impact of societal and cultural changes on word meanings and usage (Tahmasebi and Risse, 2017).

LSC detection involves the use of two corpora C_1 and C_2 which, in the diachronic case, belong to different time periods t_1 and t_2 respectively. The various approaches found in the literature usually involve the construction of a word embedding space specific to each corpus. The embeddings can be constructed through count-based methods such as Positive Pointwise Mutual Information and Random Indexing, or predictive methods such as Skip-Gram with Negative Sampling (SGNS). Most of these models are stochastic in nature which means that separately trained embedding models live in their own space. In order to project them onto a *unified space*, alignment techniques such as vector initialization and orthogonal Procrustes are used. The LSC of a word is then quantitatively determined by measuring the contextual dissimilarity between the word's representations (Tahmasebi et al., 2018).

Recently, there have been efforts to evaluate these various methods by comparing their results with manually-annotated data (Schlechtweg et al., 2019; Ahmad et al., 2020). The SemEval-2020 Task 1 is one such effort which is based on LSC detection in corpora of English, German, Latin, and Swedish languages (Schlechtweg et al., 2020). Its aim is to provide an evaluation framework for unsupervised LSC detection systems by comparing their results against a ground truth, as annotated by native speakers or scholars. It consists of two subtasks: *Given two corpora C_1 and C_2 (for time periods t_1 and t_2), for a set of target words,*

[1] github.com/vaibhav29498/GloVeInit

This work is licensed under a Creative Commons Attribution 4.0 International License. License details: http://creativecommons.org/licenses/by/4.0/.

1. *decide which words lost or gained senses between t_1 and t_2, and which ones did not; as annotated by human judges.*

2. *rank them according to their degree of LSC between t_1 and t_2 as annotated by human judges. A higher rank means stronger change.*

2 System Overview

2.1 Vector Initialization

The Vector Initialization (VI) alignment method, which was first used by Kim et al. (2014), involves the embedding space for t_1 to be trained independently on C_1. It is then used to initialize the embedding space for t_2 which is then subsequently trained from C_2. The underlying idea is that the embedding for a word w will get considerably updated if it is used within different contexts in C_1 and C_2, otherwise it will receive only a slight update. A study by Schlechtweg et al. (2019), which applied the VI method on SGNS embeddings, found it to perform significantly weaker on LSC detection tasks than other methods such as orthogonal Procrustes (OP). They attributed this to the sensitivity of the VI method to the frequency of a word in C_2. The high frequency of a word in C_2 will result into the word's embedding getting frequent updates away from its initial state. This leads to a large divergence between the word's representations even if it does not undergo any LSC from t_1 to t_2.

In a recent shared task on LSC detection in the German language, a modified VI approach achieved the third-best result and outperformed certain OP approaches (Ahmad et al., 2020). Instead of only initializing on the word embeddings obtained from C_1, the modified approach initializes on the complete SGNS model which includes the hidden layer. However, it also theoretically suffers from sensitivity to high frequency.

2.2 Comparison of GloVe and SGNS

GloVe (Global Vectors) and SGNS are unsupervised algorithms for obtaining word embedding spaces (Pennington et al., 2014; Mikolov et al., 2013). GloVe makes use of co-occurrence matrix X; its (i, j) entry, X_{ij} is the number of times the word w_j appears in the context of the word w_i (as defined by the window-size L). It trains the word embeddings by minimizing the cost function

$$J = \sum_{i=1}^{V} \sum_{j=1; j \neq i}^{V} f(X_{ij})(u_j^T v_i - \log X_{ij})^2, \tag{1}$$

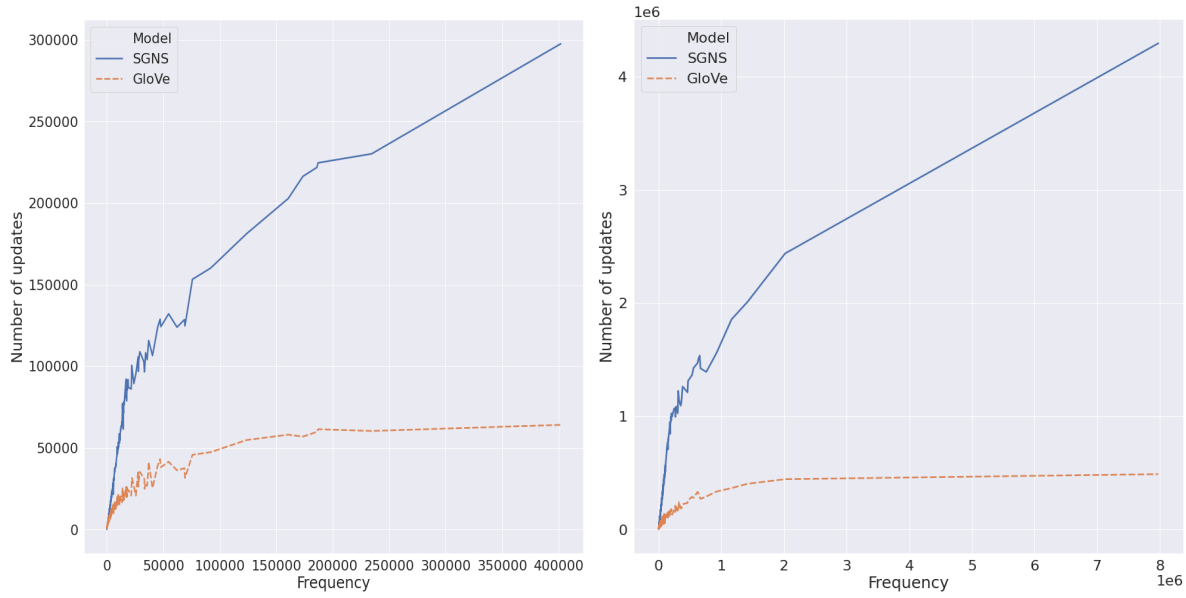
where V is the vocabulary size and $u, v \in \mathbb{R}^D$ are the word and context word vectors respectively. The final word embeddings can be obtained by summing or averaging the two.

On the other hand, SGNS is a predictive model which relies on a shallow two-layer neural network which, given a word, predicts the set of its context words. To avoid the expensive softmax function in the training objective, negative sampling is used by drawing a few negative samples from a noise distribution.

An important distinction between GloVe and SGNS from this paper's point of view is the number of updates that are made to a word embedding during training. In a single epoch, the number of updates to the SGNS embedding of a word is roughly equal to the number of words that appear in its context throughout the corpus. Hence, the number of updates is proportional to the frequency of the word which has an upper-bound of the total number of words in the corpus. However, this relationship is not exactly linear because of downsampling of frequent words and negative sampling. In the GloVe model, the number of updates received in a single epoch is equal to the number of *distinct* context words. This number is limited by an upper-bound of vocabulary size V, which is usually much less than the total number of words in the corpus.

As discussed in Section 2.1, the VI method can falsely give a high LSC score to words with high frequency. Our hypothesis is that the GloVe model is more suited for the VI method due to its lesser sensitivity to high frequency. To give an indication of the extent of this difference, the relationship between a word's frequency and the number of updates to its embedding is depicted in Figure 1 for both GloVe and

SGNS models trained on the C_2 corpora of all the four languages. These results are based on the official GloVe implementation[2] and the `gensim`[3] implementation of SGNS. The models were trained for a single epoch and only words with a frequency of greater than or equal to 5 were considered. It is clear that the proportion of updates for frequent and rare words is comparatively more balanced in the case of GloVe. On the other hand, there is a high rate of growth in the number of updates with respect to the frequency in the case of SGNS, which indicates that it is biased towards estimating a high LSC for frequent words.



(a) CCOHA 1960-2010 (English)

(b) BZ and ND 1946-1990 (German)

(c) LatinISE 0-2000 (Latin)

(d) KubHist 1895-1903 (Swedish)

Figure 1: Plot of frequency of a word (x-axis) vs the number of updates to its embedding (y-axis). The subcaptions include the name of the corpora C_2, the corresponding time-period t_2, and its language.

[2] `github.com/stanfordnlp/GloVe`
[3] `radimrehurek.com/gensim`

3 Results and Experimental Setup

The dataset provided by the competition organizers consists of the corpora C_1 and C_2, and the list of target words for four languages: English, German, Latin, and Swedish (Schlechtweg et al., 2020). All of the corpora were already prepossessed: they were in tokenized and lemmatized form with punctuation marks and one-word sentences removed.

Subtask 1 is concerned with identifying the loss or gain of one or more word sense(s), where subtask 2 tests a model's ability to detect fine-grained changes in the two sense frequency distributions. For example, consider the word *cell* which historically referred to either a chamber or the smallest unit of an organism. However, the relative usages of both these senses have decreased with time, with mobile phone being the predominant meaning now. For subtask 2, the ground-truth ranking of the target words is determined by calculating the Jensen-Shannon divergence between their normalized sense frequency distributions from t_1 and t_2 (Donoso and Sánchez, 2017). A submission is scored by its Spearman's rank-order correlation coefficient against the ground-truth ranking.

We used cosine distance as the metric to calculate the distance between a word's vectors in diachronic spaces, and every word with a distance of more than 0.55 (for German) or 0.45 (for other languages) were classified to have gained of lost word sense(s).

In all of our experiments, only words having a frequency of at least 5 were considered. Our best-performing solution during the evaluation phase achieved an accuracy of 60% in subtask 1 and a score of 0.352. It was obtained by training word embeddings with a dimensionality of 50 on both C_1 and C_2 for 60 epochs each with a window-size of 10. The subtask-2 score was close to but less than the modified SGNS-based VI approach discussed in Section 2.1, which was proposed by the team *IMS* and received a score of 0.372. We ranked 13th and 10th out of 33 participants in the two subtasks respectively.

Further experiments on subtask-2 were conducted after the competition for analyzing the impact of hyperparameters like window-size L and embedding dimensionality d and are reported in Table 1. Number of training epochs was limited to 20 for models with window-size 10 because of their high computational requirements. The results suggest that training the models for higher number of epochs can produce better results if the embedding dimensionality is high, but can backfire in the opposite case. Using a larger window-size improves the average result in most cases.

L	**Epochs**	***d***	**Scores**				
			English	**German**	**Latin**	**Swedish**	**Average**
5	20	5	0.012	0.366	0.391	0.15	0.23
		10	-0.144	0.362	0.394	0.114	0.182
		20	0.193	0.402	0.368	0.244	0.302
		50	**0.278**	0.446	0.315	0.229	0.317
		100	0.226	0.312	0.254	0.186	0.244
	60	5	-0.018	0.354	0.432	0.121	0.222
		10	-0.095	0.283	0.379	0.108	0.169
		20	0.212	0.41	0.377	0.324	0.331
		50	0.228	0.464	0.329	**0.366**	**0.347**
		100	0.269	0.312	0.276	0.309	0.291
10	20	5	0.011	0.334	**0.485**	0.218	0.262
		10	-0.036	0.351	0.409	0.146	0.218
		20	0.144	**0.479**	0.397	0.187	0.302
		50	0.198	0.477	0.275	0.267	0.304
		100	0.199	0.349	0.257	0.227	0.258

Table 1: Results of the post-evaluation experiments

3.1 Error Analysis

We defined the ranking error for each target word as the difference between its predicted rank and its true rank divided by the total number of target words. It lies in the range $(-1, 1)$ and has an ideal value of zero. A positive value indicates that the model overestimated the LSC for a word, and a negative value indicates otherwise. A regression plot between the relative frequency of the target words and their ranking error as per our best performing model is depicted in Figure 2. A word's relative frequency is defined as its frequency divided by the total frequency of all words in the corpus.

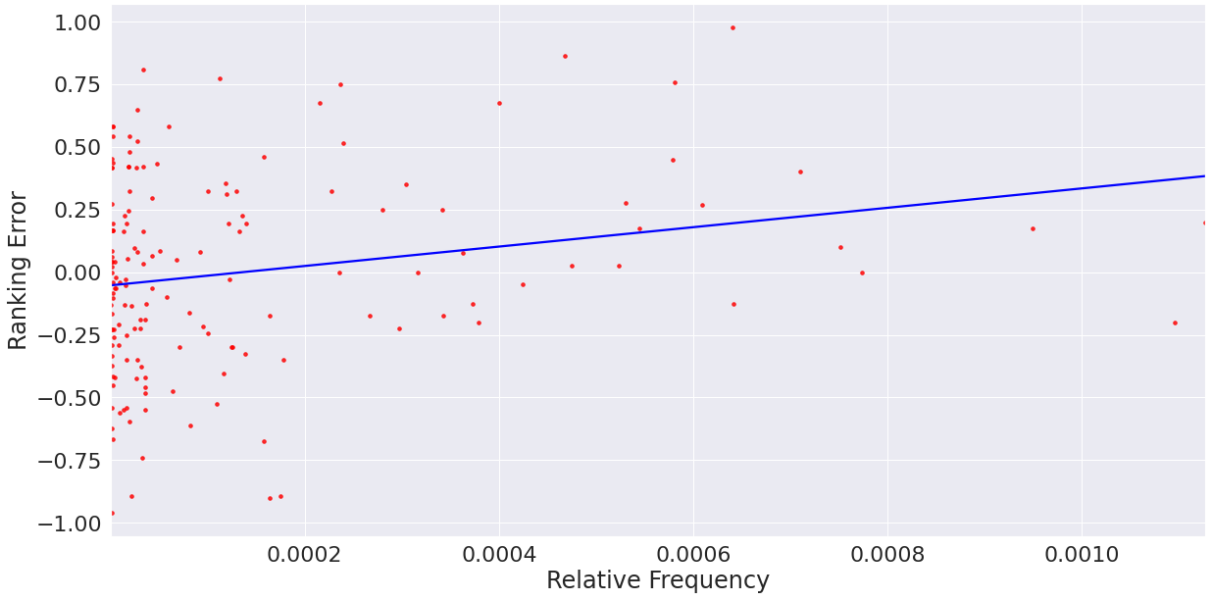


Figure 2: Regression plot between the frequency of the target words and their ranking error

There is a statistically significant correlation between the ranking error and the relative frequency, with the result of the Spearman's rank-order correlation test being $\rho = 0.186$ and $p = 0.020$. This indicates that the GloVe model's relative insensitivity to high frequency can lead to it assigning a rank lower than the true rank to such words. We believe that the frequency of the target words is not large enough for SGNS' sensitivity to high frequency to have a major impact on the results. This explains our model's inferior performance as compared to SGNS-based model in this task.

4 Conclusion and Future Work

In this paper, we reported our work in the SemEval2020 Task 1. We proposed a GloVe-based VI approach which achieved the 11th rank out of 32 participating teams. We gave a theoretical reasoning behind why GloVe is less sensitive towards high frequency than SGNS and thus more suited for VI method, and empirically showed the magnitude of the difference between the number of updates to word embeddings in the two models. We believe that despite a lower-than-expected performance in this competition, our work presents a good case for the suitability of GloVe model when corpora of larger size are involved. However, proving this quantitatively is a challenging task because of the limitations associated with manual annotation-based evaluation.

Planned future work includes the study of how techniques such as dimension-wise mean-centering and length normalization, which have proved beneficial in OP-based approaches, can be applied for VI method.

Acknowledgments

We would like to thank the task organizers for preparing the dataset and organizing the competition, Jens Kaiser (member of the team *IMS*) from University of Stuttgart for sharing the results from his work on the SGNS-based VI alignment method, and the anonymous reviewers for their detailed and helpful feedback.

References

Adnan Ahmad, Kiflom Desta, Fabian Lang, and Dominik Schlechtweg. 2020. Shared task: Lexical semantic change detection in german. *CoRR*, abs/2001.07786.

Gonzalo Donoso and David Sánchez. 2017. Dialectometric analysis of language variation in twitter. In *Proceedings of the Fourth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial)*, pages 16–25, Valencia, Spain, April. Association for Computational Linguistics.

Vaibhav Jain, Ruchika Malhotra, Sanskar Jain, and Nishant Tanwar. 2020. Cross-domain ambiguity detection using linear transformation of word embedding spaces. In *Joint Proceedings of REFSQ-2020 Workshops, Doctoral Symposium, Live Studies Track, and Poster Track co-located with the 26th International Conference on Requirements Engineering: Foundation for Software Quality (REFSQ 2020)*, volume 2584 of *CEUR Workshop Proceedings*, Pisa, Italy. CEUR-WS.org.

Yoon Kim, Yi-I Chiu, Kentaro Hanaki, Darshan Hegde, and Slav Petrov. 2014. Temporal analysis of language through neural language models. In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, pages 61–65, Baltimore, MD, USA, June. Association for Computational Linguistics.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS13, page 31113119, Red Hook, NY, USA. Curran Associates Inc.

Sara Morsy and George Karypis. 2016. Accounting for language changes over time in document similarity search. *ACM Trans. Inf. Syst.*, 35(1), September.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar, October. Association for Computational Linguistics.

Dominik Schlechtweg, Anna Hätty, Marco Del Tredici, and Sabine Schulte im Walde. 2019. A wind of change: Detecting and evaluating lexical semantic change across times and domains. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 732–746, Florence, Italy, July. Association for Computational Linguistics.

Dominik Schlechtweg, Barbara McGillivray, Simon Hengchen, Haim Dubossarsky, and Nina Tahmasebi. 2020. SemEval-2020 Task 1: Unsupervised Lexical Semantic Change Detection. In *To appear in Proceedings of the 14th International Workshop on Semantic Evaluation*, Barcelona, Spain. Association for Computational Linguistics.

Nina Tahmasebi and Thomas Risse. 2017. On the uses of word sense change for research in the digital humanities. In *Research and Advanced Technology for Digital Libraries*, pages 246–257, Cham. Springer International Publishing.

Nina Tahmasebi, Lars Borin, and Adam Jatowt. 2018. Survey of computational approaches to lexical semantic change. *CoRR*, abs/1811.06278.

The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/1934-8835.htm

Gratitude, social intelligence, and leadership among university teachers: mediation and moderation analysis

Naval Garg
University School of Management and Entrepreneurship, Delhi Technological University, Delhi, India

Anuradha Jain
Vivekanand School of Economics, Vivekananda Institute of Professional Studies, Pitampura, India, and

B.K. Punia
Haryana School of Business, Guru Jambheshwar University of Science and Technology, Hisar, India

Received 18 January 2020
Revised 8 April 2020
20 May 2020
Accepted 24 May 2020

Abstract

Purpose – The purpose of this study is to investigate the linkage between gratitude and leadership among university teachers. It also tended to examine mediating and moderating effect of social intelligence amidst the relationship between gratitude and leadership. As social intelligence is a multi-dimensional concept, this paper explored mediating and moderating effect of three dimensions of social intelligence, i.e. social awareness, social skills and social information processing.

Design/methodology/approach – Mediation effect was examined with help of Baron and Kenny's (1986) technique, Sobel test and bootstrapping estimates. Significance of interaction between independent variable (gratitude) and moderator (social intelligence) was investigated for moderation analysis.

Findings – The findings reported partial mediating effect of all three dimensions of social intelligence. No significant moderation effect was observed.

Originality/value – This is one of the pioneer studies that explored interrelationship between gratitude, social intelligence and leadership among Indian University teachers.

Keywords Gratitude, Social intelligence, Leadership, Mediation, Moderation

Paper type Research paper

Introduction

A growing number of researchers in positive psychology have proved that gratitude improves well-being (Emmons and McCullough, 2003; Watkins *et al.*, 2014; Wood *et al.*, 2010), facilitates life satisfaction (McCullough *et al.*, 2002; Garg, 2017a), elicits prosocial behavior (Bartlett and DeSteno, 2006) and reduces aggression, envy and narcissism (Froh *et al.*, 2011). However, researchers still believe that the domain of gratitude is yet to be explored adequately. Emmons and Crumpler (2000) highlighted the scarcity of studies on gratefulness in the sphere of social science and organizational psychology. Only handfuls of researchers have investigated the role of gratefulness in modern organizations (Waters, 2012; Grant and Wrzesniewski, 2010).

International Journal of Organizational Analysis
© Emerald Publishing Limited
1934-8835
DOI 10.1108/IJOA-01-2020-2016

According to Emmons and McCullough (2003), gratitude directly influences organizations through improving organizational climate. It also enhances employee's well-being and reduces negative emotions like resentment and envy at the workplace. Also, Fehr *et al.* (2017) proposed that gratitude may play an essential role in various prominent organizational phenomena such as leadership, organizational citizenship and customer service. The feeling of thankfulness may also promote employee efficiency, success, productivity and loyalty (Grant and Wrzesniewski, 2010; Garg, 2018). Previous studies support the positive association of gratitude with organizational citizenship behaviors (Dik *et al.*, 2014; Spence *et al.*, 2013; Garg *et al.*, 2019), prosocial work behaviors (Grant and Gino, 2010), altruism (Spence *et al.*, 2013) and social support at the workplace (Kaplan *et al.*, 2014). In a critical study, Edmondson (2002) reported that gratitude facilitates psychological safety at the workplace, which promotes trust, faith and courage to take interpersonal risks. It influences the attitude of employees toward their colleagues, superiors and management, which enhances their job satisfaction by facilitating better interpersonal relationships (Hu and Kaplan, 2014; Giles and Yates, 2014). Gratuitous organizational culture ensures the recognition of employees' contributions to the workplace. Authors also reported that grateful employees approach their colleagues in a more positive note, which leads to more magnificent team spirit and altruistic behavior.

Although these studies have highlighted the potential benefits of gratitude for modern organizations, they are often criticized for three main reasons. Firstly, most of the studies are theoretical; thereby, the domain calls for empirical investigation of these hypothesized relationships. Secondly, a theoretical explanation of gratitude (see broaden-and-build theory of positive emotions, social-cognitive theory of gratitude and find, remind and bind theory of gratitude) is primarily based on western context and researchers advocate the application of these theories in eastern countries for better acceptability and research rigor. Then finally, research studies are yet to explore the direct impact of gratitude on individual and organizational outcomes such as motivation, morale, intelligence, customer satisfaction and profitability (Fredrickson, 2004; Fehr *et al.*, 2017). Fredrickson (2004) highlighted the need to investigate the role of gratitude in organizations. "In organizations, does gratitude lead to lower employee turnover, more customer loyalty, higher net sales or more profitable outcomes?" (Fredrickson, 2004, p. 160). Unfortunately, the research gap highlighted by Fredrickson is yet to be fully answered. Other researchers, including Grant and Wrzesniewski (2010), Watkins *et al.* (2003), Garg (2020) and Garg and Gera (2019), have also called for more empirical researches on gratitude, especially in the modern organization.

While explaining the broaden-and-build theory of positive emotions in the context of gratefulness, Fredrickson (2004) suggested an investigation of the mechanism through which gratitude strengthens organizations. Probably, the author was aspiring for future researchers to examine moderating and mediating variables amidst the relationship between gratitude and proposed organizational outcome variables. Algoe and Stanton (2012) find, remind and bind the theory of gratitude and Watkins's *et al.* (2014) amplification theory of gratitude also reiterated the same research gap. Based on these theories, this study proposes a model that tends to examine mediating and moderating social of social intelligence in the relationship between gratitude and leadership. The mediating variable explains the mechanism through which the effect of independent variables transmits to the dependent variable. In other words, the independent variable influences the mediating variable, which ultimately influences the dependent variable. Then, the moderating variable influences the relationship between predictor and criterion variable. The moderators act as a catalyst that either increases or decreases the intensity of association between two variables. Following a section of this paper reviews previous literature on gratitude, social intelligence

and leadership. Next, several hypotheses regarding the relationship between gratitude, three dimensions of social intelligence (social awareness (SA), social skills (SS) and social information processing), and leadership are proposed in Section 2. Propositions of mediating and moderating effects of three dimensions of social intelligence are also introduced here. Then, the methodology, sample, questionnaires and findings of the study are discussed in Section 3. This paper ends with a discussion on the theoretical and practical implications of the study in Section 4. The limitations of the study, along with scope for future researchers, were also discussed in Section 5 of the paper.

Theory and conceptual framework

Gratitude

Gratitude has been conceptualized in several forms, including emotion, trait, virtue, habit and coping mechanism (Emmons and McCullough, 2003). "In various contexts, gratitude can refer to an emotional state, an emotional expression, a character trait or even a virtue" (Watkins *et al.*, 2014, p. 13). Researchers suggested that Rosenberg's (1998) "level of analysis" is helpful for an adequate understanding of various facets of gratitude (Watkins *et al.*, 2014). Rosenberg (1998) argued that an affective phenomenon could be divided into three levels of analysis, namely, emotion, mood and trait. Similarly, gratitude could also be approached using this typology, which assists researchers to conceptualize gratefulness appropriately.

As emotions, gratitude may be experienced when people feel "that something good has happened to them, and they recognize that someone else is largely responsible for this benefit" (Watkins *et al.*, 2014, p. 17). Moral affect theory of gratitude (McCullough *et al.*, 2001) reported that gratitude is an honest emotion in three ways. Firstly, it acts as a moral barometer. A barometer is an instrument that indicates a change in atmospheric pressure. In the same way, gratitude represents the emotional response of beneficiary toward benefactor's intentional act of benevolence. Secondly, gratitude acts as a moral motivator. A grateful individual is more likely to act morally and pro-socially. Previous researches (Peterson and Stewart, 1996; Graham, 1998) specially elaborated that a beneficiary possesses a higher probability of contributing to the welfare of the benefactor or even to the third party. Then finally, it can reinforce moral behaviors. A benefactor who receives sincere gratitude is more likely to help their beneficiary again, and thereby, gratuitous behaviors are reinforced (Watkins *et al.*, 2014).

As an affective trait, it is defined as a "general tendency to recognize and respond with grateful emotions to the roles of other people's benevolence in the positive experiences and outcomes that one obtains" (McCullough *et al.*, 2002, p. 112). This approach considers gratitude as a personality trait of an individual. "Trait gratitude may be viewed more as the being of gratitude, whereas gratitude emotion is related to the feelings of gratitude" (Watkins *et al.*, 2014, p. 22). Watkins (2009) defined three essential components of trait gratitude and termed them as "three pillars of gratitude." Firstly, grateful people are poised to have a strong *sense of abundance* or lack of *sense of deprivation*. A sense of abundance or lack of sense of deprivation allows an individual to acknowledge and appreciate all physical and material things that a person possess. Such persons start appreciating previously unnoticed blessings such as a human body, sound mind, three square meal, family, friend, job and home. Although these items seem quite reasonable and insignificant, a mere realization that a large number of people still dreams of having a life similar to them does the miracle. This realization helps a grateful person to acknowledge the fact that he is far more blessed than most of the others. Appreciation of simple pleasure is expressing thankfulness for mundane aspects of life such as a pleasant meal, excellent weather, sudden and

unplanned meeting with an old friend. Secondly, grateful people have a strong sense of *appreciation for the simple pleasures* of life. Even mundane, routine, and day-to-day benefits are cheered by people possessing trait gratitude. Appreciation of simple pleasure is expressing thankfulness for all aspects of life such as a pleasant meal, excellent weather, sudden and unplanned meeting with an old friend, etc. Then finally, grateful people are also characterized by *social appreciation* or *appreciation of others*: they appreciate the contribution of others in their life. This component of trait gratitude also values the importance of expression of appreciation to others.

In addition to these pillars of gratitude, McCullough *et al.* (2002) elaborated four facets of trait gratitude, namely, *intensity*, *frequency*, *span* and *density*. Frequency is the number of times an individual experiences gratitude in a given set of time. Span refers to several events for which one feels gratefulness in a given period. Density involves the number of persons for which an individual experiences gratefulness regarding one particular positive outcome. Then finally, intensity refers to a degree of experience of gratefulness. To summarize, people high in trait gratitude should show facets of intensity, span, frequency and density of gratitude. Little work has investigated gratitude as mood probably because emotion and mood states are not appreciatory distinguished in previous researches (Watkins *et al.*, 2014). However, Rosenberg's approach clarified that moods are longer in duration and are more rooted in awareness than emotions. Grateful moods are not about something particular. Then they are assumed to facilitate grateful emotions. In a nutshell, researchers hailed gratitude as a multidimensional concept that can take several forms (Alkozei *et al.*, 2017). Although researchers have discussed several perspectives of gratitude, they broadly agree on four facets of gratefulness. These facets are, firstly, benefactor favors beneficiary intentionally without being bound of duty; secondly, the gift has certain utility for the receiver; thirdly, the recipient recognizes the goodness of favor/gift and finally, the receiver aspires to reciprocate the gift received from the giver.

Social intelligence

Social intelligence is defined as "the capability to understand and manage humanity to act wisely in human relations (Thorndike, 1920, p. 228). According to Marlowe (1986), social intelligence is one's capability to understand the feelings, thoughts and behaviors of self and others in social circumstances and consequently act according to that understanding. It is accessed with the help of the Tromso Social intelligence scale (TSIS) that investigates three dimensions of social intelligence, namely, SA, SS and Social Information Processing (SIP). SA refers to one's ability to sense social cues about another person's thoughts, aspirations, and emotions. It helps an individual to understand the social situation appropriately (Goleman, 2006). Goleman (2006) empathized that the empathic acknowledgment of other's feelings and roles is the basis of SA. Further, SS are competencies that promote interaction and communication with others. They are cumulative inherent and learned behaviors that influence one's relationship with others (Agran *et al.*, 2016; Davies *et al.*, 2015). Examples of such behaviors are praising and appreciating others, cooperation, helping and seeking help from others, collaborating with others. Then finally, SIP describes the process of cognitive and behavioral response generation is affected by the SIP mechanism. It refers to the ability to understand social cues and other's feelings and aspirations. It helps an individual to predict one's beliefs, expectations, wishes and behavior through observing their facial expression, gestures, eye movement and body language. In a nutshell, it sensitizes employees for their and other feelings.

Leadership

Northouse (2007) defines leadership as "a process whereby an individual influences a group of individuals to achieve a common goal" (p.75). Also, Greer ad Plunkett (2000) stated that "leadership is an art that can be acquired and developed by anyone with the motivation to do so." (p. 157). Researchers have propagated various theories to explain how people become leaders. One of the most eminent scholars, Bass (1990) contributed greatly in the development of three leadership theories. Firstly, trait theory acknowledged the contribution of certain desirable personality traits in evolution of a leader. Secondly, situational theory gave credit to a situation or event that provided an opportunity to rise above the occasion. Such situations are so stimulating and challenging that they elicit extraordinary leadership skills from an ordinary-looking man. Then finally and most importantly, transformational leadership theory emphasized on four essential dimensions of leadership i.e. individualized consideration, intellectual stimulation, inspirational motivation and idealized influence. The theory stated that "transformational leaders are people with a vision who demonstrate the ability to accomplish greatness, are capable of obtaining results through the injection of energy and enthusiasm, and are generally trusted, respected and admired by their followers" (pp. 23–25). The underlying premise of the present study is based on the transformational theory of leadership. Here, gratitude is proposed to have influence over four inherent dimensions of transformational leadership. Garg (2020) also supported the association between gratitude and different constructs of leadership. The theoretical premise of the interrelationship between gratitude, social intelligence and leadership are explained in the following paragraphs.

Gratitude, social intelligence and leadership

Proposed associations between gratitude, social intelligence and leadership are theoretically based on Fredrickson's (1998) *broaden-and-build theory of positive emotions*, and Algoe and Stanton (2012) *find, remind and bind argument of gratitude*. According to Fredrickson, positive emotions like gratitude "broaden people's momentary thought-action repertoires and build their enduring social resources" (Fredrickson, 2004, p. 147). She elaborated that positive emotion broadens thought-action tendencies in three broad ways, namely, through broadening the scope of attention, the range of cognition and scope of action. It is reported that positive emotions elicit a well-thought response from beneficiary instead of a narrow tit-for-tat reaction. Grateful people creatively explore several activities as probable reflections of their gratitude. They tend to think consciously about how to meet the needs of their benefactor, which broadens their attention toward the giver's needs (Watkins *et al.*, 2014). Several studies, including Peterson and Stewart (1996), Graham (1998), McCullough *et al.* (2001), also favored the broad perspective of the theory.

Fredrickson also argued that gratitude builds resources in at least four ways. Firstly, gratitude is assumed to develop and strengthen social bonds. Secondly, in addition to establishing a connection between benefactor and beneficiary, gratitude also strengthens one's relationship with the divine. Several researchers, including Uhder *et al.* (2010), Watkins *et al.* (2014), also supports the fact that gratitude builds and strengthens spirituality. Thirdly, gratitude promoted civil communities and organizations. Then finally, gratitude is poised to expand one's "more general skills for loving" (Fredrickson, 2004, p. 152). In a nutshell, gratitude is poised to transform both at individual and organizational levels. Following the same analogy, this study proposed that gratitude broadens one's social intelligence, and it builds leadership skills.

Now, this study proposes that gratitude improves SA, SS, and social information processing. Broaden perspective of positive emotion elaborated expansion of thought-action

repertoires, the scope of attention, cognition and action and awareness of other's needs. Then, these expanded repertoires seem to have a direct bearing on the social intelligence of the employees. Few researchers, including Covey (1996), Boyatzis and Saatcioglu (2008) reported a positive relationship between gratitude and some elements of social intelligence. Further, banking on Fredrickson's explanation, it is proposed that gratitude broadens social intelligence, which builds leadership resources (skills) for a manager.

The proposed association between gratitude and leadership could also be hypothesized based on Algoe and Stanton (2012) *find, remind and bind theory of gratefulnes*s. Algoe's work is an extension of Fredrickson's philosophy, and it tends to explain the mechanism through which gratitude builds resources for an individual. According to this theory, gratitude helps one find new relationships in social and organizational settings. A grateful person creates and avails more opportunities to initiate a social exchange with others, which facilitates the establishment of a new connection. It also maintains an ongoing relationship by reminding the blessing of association with others. A grateful person cherishes the social context, and thus, makes conscious and consistent efforts to manage them. This effort tends to bind these social affiliations and makes them long-standing. Again, based on this explanation of Find, Remind and Bind theory of gratitude, it could be proposed that gratitude works on one's social intelligence, which bestows them with permanent and favorable social affiliation. Then, good social association with colleagues and followers is the hallmark of success for a leader.

Garg and Gera (2019) explored the mediating effect of social intelligence amidst the relationship between gratitude and leadership. They explained that social intelligence partially mediates the intended association. The current study aspired to further strengthen the findings of the authors by exploring the mediating effect of three constructs of social intelligence. It has also been argued that an individual's personality traits like social intelligence may influence the relationship between gratitude and leadership. Accordingly, the present tends to investigate both mediation and moderation effect with the help of the following theoretical models (Figures 1 and 2) and hypotheses.

H1. Gratitude is positively related to leadership.

H2. Gratitude is positively related to SA.

H3. Gratitude is positively related to SS.

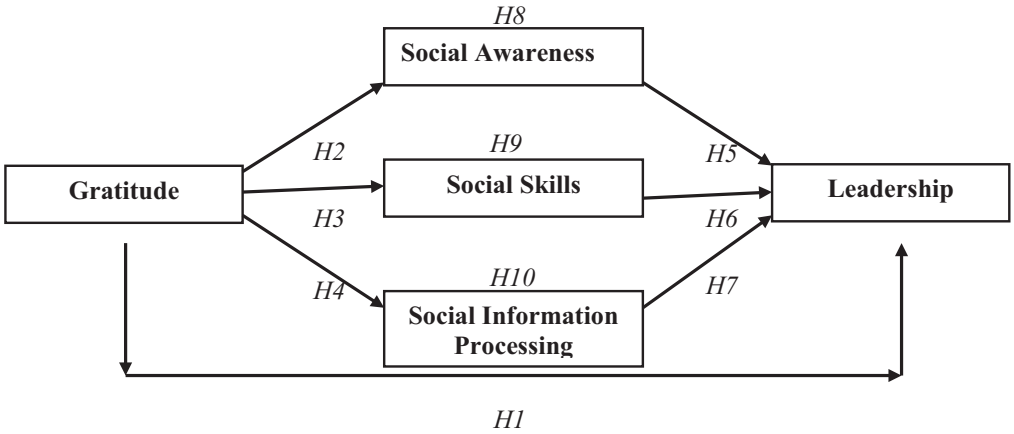


Figure 1.
Mediation model

H4. Gratitude is positively related to social information processing.

H5. SA is positively related to leadership.

H6. SS is positively related to leadership.

H7. SIP is positively related to leadership.

H8. SA mediates the relationship between gratitude and leadership.

H9. SS mediates the relationship between gratitude and leadership.

H10. SIP mediates the relationship between gratitude and leadership.

H11. SA moderates the relationship between gratitude and leadership.

H12. SS moderates the relationship between gratitude and leadership.

H13. SIP moderates the relationship between gratitude and leadership.

Methodological framework

Objectives, sample and procedures

The principal aim of this study was to investigate the linkage between gratitude and leadership among university teachers. It also tended to examine the mediating and moderating effect of social intelligence amidst the relationship between gratitude and leadership. As social intelligence is a multi-dimensional concept, this paper explored the mediating and moderating effect of three dimensions of social intelligence, i.e. SA, SS and social information processing. The sample comprised of 252 university professors teaching across various private, government, and autonomous institutes across Delhi- National Capital Region (NCR), Haryana and Punjab. Respondents were approached either personally or through emails. A total of 300 University teachers were approached and 255 teachers reverted. Three questionnaires were discarded for being incomplete. Now, colleagues of these 252 teachers were approached. Respondent was asked to respond to gratitude questionnaire (GQ-6) and TSIS and one of the colleagues of the respondent was approached to access leadership skills of respondent through multifactor leadership questionnaire (MLQ-5X). The convenience sampling method was used to reach out to these 300 respondents. Respondents were assured regarding the anonymity of responses and for strictly academic usage of collected data. Out of 252 respondents, 124 (49%) were male and 128 (51%) were female. Also, 19% of teachers were below 25 years, 26% were between 26-35 years, 21% were between 36-45 years, 16% were between 46-55 years and the remaining 18% were above the age of 55 years. Finally, the majority of respondents (51%) were from Delhi-NCR.

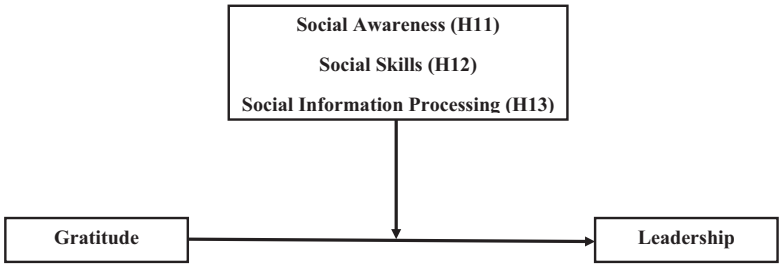


Figure 2.
Moderation model

Measures

Gratitude. It was measured with the help of a GQ-6, a self-report instrument developed by McCullough *et al.* (2002). This scale has six items rated on a seven-point scale ranging from 1 (strongly disagree) to 7 (strongly agree). Out of six items, two items are reverse scored on this scale. GQ-6 is one of the most practiced gratitude scales in academic researches (Froh *et al.*, 2008; Yüksel and Oguz Duran, 2012; Sun and Kong, 2013; Kong *et al.*, 2015; Caputo, 2016; Langer *et al.*, 2016). The six items include "If I had to list everything, I felt grateful, it would be a very long list," "I have so much in life to be thankful for" and "Long amounts of time can go by before I feel grateful to something or someone" (reverse-scored).

Social intelligence. Social Intelligence was measured with the 21 items TSIS developed by Silvera *et al.* (2001). TSIS accesses three dimensions of social intelligence, namely, SA (SA), SS (SS) and SIP. Seven statements are reverse scored, and each statement is measured on the seven-point rating scale. SA is measured with items like "Other people become angry with me without me being able to explain why" and "I have often hurt others without realizing it." Examples of items included to measure SS are "I fit in easily in social situations" and "I have a hard time getting along with other people." SIP is accessed with items like "I know how my actions will make others feel" and "I understand others' wishes."

Leadership. Leadership is measured with the help of the 21 items MLQ-5X developed by Rowold (2004). It accesses seven dimensions of leadership (inspirational motivation, idealized influence (attributed), idealized influence (behavior), individualized consideration, intellectual stimulation, contingent reward and management by exception) rated on a five-point rating scale ranging from 1 (strongly disagree) to 5 (strongly agree).

Reliability and validity

Reliability refers to the consistency of a measure. It is the degree to which similar results are obtained under consistent conditions (Malhotra and Dash, 2018). In this paper, internal consistency reliability, which is the consistency of the respondent's response across different items of the scale, is accessed Cronbach's alpha. It measures split-half correlation through splitting items into two sets, and Cronbach's alpha represents mean of all possible split-half correlations for a set of items. According to George and Mallery (2003), Cronbach's alpha should be greater than 0.70 to confirm reliability. All measures which confirm internal consistency reliability are reported in Table 1 (Cronbach's alpha value for gratitude = 0.877, SA = 0.748, SS = 0.709, SIP = 0.736 and leadership = 0.778). It is suggested to calculate composite reliability along with Cronbach's alpha (Cronbach and Shavelson, 2004). Composite reliability estimates should also be greater than 0.70 to reaffirm reliability (Hair *et al.*, 2010). Also, Convergent and Discriminant validity is determined with average variance extracted (AVE), which was calculated by summation of square of each factor

Variable	Mean	S.D.	α	CR	AVE	VIF
Gratitude	4.68	1.231	0.888	1.270	0.877	1.396
Social awareness	3.83	0.565	0.800	0.584	0.748	2.062
Social skills	3.51	0.643	0.782	0.633	0.709	2.727
Social information processing	3.34	0.761	0.794	0.764	0.736	2.475
Leadership	3.65	0.732	0.773	0.713	0.778	1.231

Notes: S.D. – Standard deviation; α – Cronbach's alpha; CR – Composite reliability; AVE – Average variance extracted; VIF – Variance inflation index
Source: Primary data

Table 1.
Descriptive statistics

loading divided by the number of indicators. AVE must be greater than 0.50 for convergent validity (Fornell and Larcker, 1981). AVE values are reported in Table 1. Discriminant validity indicates distinctiveness. It reflects the degree to which one construct is different from another construct (Bryman and Bell, 2014). Fornell and Larcker (1981) suggested a simple method to ensure Discriminant validity i.e. AVE's value should be higher than the squared inter-correlations of constructs. Table 3 depicts the results of Fornell and Larcker's (1981) results. AVEs were compared (highlighted in bold in the diagonal) with the squared intercorrelations (non-bold entries and below the diagonal). It was found that AVE was greater than the squared correlations, and thereby, it supports the existence of discriminant validity.

Multicollinearity and common method bias

According to Kline (2009), the problem of multi-collinearity surfaces when the correlations among variables are high. In this study, multi-collinearity was explored with the variance inflation factor (VIF), which should be less than five (Kline, 2009). The present study did not have an issue with multi-collinearity, as all VIF values are less than five (Table 1). Detrimental effects of method variance include biased estimates of reliability and validity, inflated or deflated estimates of the relationship between two constructs, affect hypothesis testing, lead to Type I or Type II error and deflated interaction effect (Podsakoff *et al.*, 2012). To minimize common method bias, this study adopted both procedural and statistical remedies. Firstly, pre-validated measurement scales with a different rating scale for predictor (seven-point) and criterion (five-point) were used. Secondly, the anonymity of the respondents was assured and strictly maintained. Also, respondents (university teachers) were explained how the study and its findings might be useful for them. It motivates respondents to respond more accurately and positively (Podsakoff *et al.*, 2012). Also, the independent variable and dependent variables were measured from two different sources. Respondent was asked to respond to GQ-6 and TSIS, and one of the colleagues of the respondent was approached to access the leadership skills of the respondent through MLQ-5X. Apart from these procedural methods, the Harman one-factor was also used to check common method bias. More than one factor was reported when items of different measures were subjected to factor analysis.

Statistical tools

The collected data is analyzed with the help of different statistical tools such as the measure of central tendency, standard deviation, correlation, hierarchical regression, Sobel test and bootstrapping.

Data analysis and interpretation

Table 1 shows means, standard deviations, and reliability, validity and multicollinearity estimates of variables of the study (Tables 1 and 2). Table 2 elucidates the correlation among variables of the study. Results suggest that gratitude is positively associated with the leadership of the teacher and with three dimensions of social intelligence. Hence, *H1–H4* are supported. Also, SA, SS, and SIP are positively correlated with leadership. Thus, *H5–H7* are confirmed (Tables 3 to 5).

Now, mediating effect of SA, SS and SIP on the relationship between gratitude and transformational leadership is explored with Baron and Kenny's (1986) four-step approach, in a combination of "Sobel test" and "bootstrap of the indirect effect" as suggested by Zhao *et al.* (2010). As per Baron and Kenny (1986), the first three conditions for mediation are significant regression between an independent variable and dependent variable,

independent variable and mediator and mediator and dependent variable. In the fourth step, a mediator is controlled, and the independent variable is again regressed on the dependent variable. Complete mediation is present when the predictor variable no longer predicts criterion variable and partial mediation is present when the influence of the independent variable on the dependent variable is reduced on the controlling effect of the mediating variable.

Table 4 presents path coefficients and significance levels for the proposed mediation hypotheses using hierarchical regression. It is argued that Baron and Kenny's (1986) method provides a fair idea about mediation, but the statistical significance of the mediator could be examined using a Sobel test (Yay, 2017). It is also argued that both these methods assume a normal distribution of data. Thus, researchers advocate the bootstrapping process, which is resampling with a replacement for large times such as 5,000 or 10,000 times, to confirm the mediation effect. Bootstrapping method is increasingly gaining acceptance among researchers (Shrout and Bolger, 2002; Fritz *et al.*, 2012). Table 5 shows the result of the Sobel and bootstrapping method. Results of Tables 3 and 4 suggest that SA, SS and SIP partially mediate the relationship between gratitude and transformational leadership. Hence, *H8–H10* are supported.

Finally, the moderation effect of three dimensions of social intelligence was analyzed by calculating the interaction effect of an independent variable (gratitude) and proposed moderators. It is being suggested that significant interaction confirms the moderation effect. Results of Table 6 show that none of the three dimensions of social intelligence (SA, SS and social information processing) moderates the relationship between gratitude and leadership. Hence, *H11*, *H12* and *H13* are rejected. Findings are discussed in the next section of the paper.

Discussion

In one of the most popular articles, 'grateful leaders make the best leaders,' Nieuwhof (2017) discussed five essential reasons for leaders to cultivate and cherish the virtue of gratitude.

Variables	1	2	3	4	5
Gratitude	1				
Social awareness	0.507**	1			
Social skills	0.454**	0.652**	1		
Social information processing	0.376**	0.620**	0.753**	1	
Leadership	0.549**	0.629**	0.736**	0.654**	1

Notes: **Correlation is significant at the 0.01 level (two-tailed)
Source: Primary data

Table 2. Correlation matrix

Variables	1	2	3	4	5
Gratitude	0.877				
Social awareness	0.257	0.748			
Social skills	0.206	0.426	0.709		
Social information processing	0.141	0.384	0.567	0.736	
Leadership	0.301	0.395	0.541	0.427	0.778

Source: Primary data

Table 3. Discriminant validity results

Firstly, he argued that gratitude impacts the overall attitude of a human being. It makes a person more compassionate, kinder, less suspicious and less resentful. Then all these qualities are essential traits of a good leader. Secondly, a sense of abundance allows a grateful leader to see and use an opportunity where other misses. According to Nieuwhof, this abundance thinking develops an abundance mentality that enables leaders to realize immense prospects of growth and development. Thirdly, gratitude may fuel generosity among leaders, which wins the hearts and souls of the followers. Also, people desperately wish to work with a grateful leader. In other words, gratitude attracts people to work in a team for larger organizational goals. The author claims that employees are ready to accept lower wages or even pay cuts to work under a grateful senior. Then finally, gratitude neutralizes anger and jealousy of a leader. Then, a calm and compose leader can easily touch the skies of success.

Table 4. Mediation analysis (results of hierarchical regression)

Hypothesized relationships	IV to DV	IV to Med	M to DV (with mediator)	IV to DV	Hypotheses support
G→ SA →L	0.549*	0.507*	0.629*	0.472*	*H8*
G →SS →L	549*	0.454*	0.736*	0.499*	*H9*
G →SIP →L	0.549*	0.376*	0.654*	0.521*	*H10*

Notes: NS: Not significant; IV – Independent variable; DV – Dependent variable; M – Mediator; G – Gratitude; SA – Social awareness; SS – Social skills; SIP – Social information processing; L – Leadership
*$p < 0.05$
Source: Primary data

Table 5. Mediation analysis (results of sobel test and bootstrapping)

	Sobel test			Bootstrapping estimates		
Hypothesized relationships	*z*-value	Sig.	Effect	SE	LLCI	ULCI
G→SA→L	6.386	0.000*	0.1421	0.0284	0.0882	0.1992
G→SS→L	6.932	0.000*	0.1653	0.0312	0.1037	0.2254
G→SIP→L	5.559 0.000*		0.1163	0.0257	0.0658	0.1677

Notes: NS: Not significant; IV – Independent variable; DV – Dependent variable; M – Mediator; G – Gratitude; SA – Social awareness, SS – Social skills, SIP – Social information processing. *$p < 0.05$
Source: Primary data

Table 6. Moderation analysis

	Standardized coefficients		
Hypothesized relationships	β_1	β_2	β_3
$L = \beta_0 + \beta_1 G + \beta_2 SA + \beta_3 (G * SA)$	0.309(S)	0.468 (S)	−0.133(NS)
$L = \beta_0 + \beta_1 G + \beta_2 SS + \beta_3 (G * SS)$	0.273(S)	0.613 (S)	0.005(NS)
$L = \beta_0 + \beta_1 G + \beta_2 SIP + \beta_3 (G * SIP)$	0.334(S)	0.528(S)	−0.063(NS)

Notes: *$p < 0.05$; as interaction is insignificant so moderation does not exist; S: Significant, NS: Not significant, IV – Independent variable, DV – Dependent variable, M – Mediator, G – Gratitude, SA – Social awareness, SS – Social skills, SIP – Social information processing
Source: Primary data

Considering such expert opinions, objects of a broaden-and-build theory of positive emotions, attachment perspective of a social cognitive theory of gratitude, Fredrickson's theoretical predictions, previous literature and advice of practitioners, this study hypothesized that gratefulness influences leadership. It is revealed that gratitude is positively associated with leadership. The present study also suggests a statistically robust and positive relationship between gratitude and SA, SS and social information processing. Also, the mediation and moderation effect of three dimensions of social intelligence (SA, SS and SIP) are explored. It reports that all these three dimensions of social intelligence partially mediate the relationship between gratitude and transformational leadership. The findings suggest that cultivating gratitude broadens one's social intelligence, which builds leadership skills. These findings are consistent with the broaden-and-build theory of positive emotions and also with a prediction of Fredrickson.

SA refers to one's ability to emphatically sense social cues about another person's thoughts, aspirations and emotions, which helps them to understand others appropriately. Then, here gratitude hits the bull's eye. A grateful leader is presumed to analyze others' contributions correctly and then acknowledge their efforts through grateful and appreciative behaviors at an appropriate public forum. Leaders generally resort to the oral expression of thankfulness, appreciation note, reward, promotion and challenging assignment to express their gratitude toward their subordinates.

SS are competencies that promote interaction and communication with others. Then, researches suggested that managers who express gratitude make their subordinates more comfortable for voicing their concerns. Grateful people appear more warmth and responsive to others, and hence, they command their trust and faith (Fredrickson, 2004), which makes them easy to approach. This easiness and comfort of subordinates allows free flow of interaction between two parties and thereby facilitates SS. As far as SIP is concerned, it is one's ability to regulate negative emotions such as anxiety, nervousness, etc. in social situations. This study reports that gratitude influence the SIP capacity of leaders. The findings could be understood with the help of the "undoing effect" of positive emotions. Several studies, including Sansone *et al.* (2010), Allen (2018) and Jans-Beken (2018) reported the undoing effect of gratitude through neutralizing stress, anxiety, nervousness and depression.

The findings of the present study complement the results of previous studies. Garg and Gera (2019) also reported a positive association between gratitude and leadership. Afterward on, Garg (2020) advocated that gratitude is both necessary and sufficient conditions for the development of seven essential constructs of educational leadership. Waters and Stokes (2015) also reported that an effective school leader attends both emotion-gratitude and action-gratitude. The authors also supported that gratuitous behaviors of leaders reciprocate, and it is equally effective in good and challenging time. Howells (2013) discussed the importance of gratitude for school leadership in great length. The results of the present study support her qualitatively researched propositions.

Conclusion

Theoretical contribution of the study

The study makes a few significant contributions to the existing literature. Firstly, the current study aspires to address the scarcity of gratitude studies in organizational contexts. Fredrickson (2004), in her famous book *The Psychology of Gratitude,* suggested an exploration of gratitude for workplace betterment. She promoted future researchers to investigate the association between gratitude and employee performance, leadership skills, employee satisfaction level and profitability of the firm. Unfortunately, her call has not been

attended enough in past 15 years. Then, this is one of the few studies in the direction enlightened by Fredrickson. Secondly, findings of the study provide an alternate explanation of broaden-and-build theory. It could be inferred that gratitude as positive emotions broadens one's SS, which build resources in terms of leadership skills. Further, the study stresses the significance of gratitude in the context of leadership. Probably, this is the first empirical study to investigate the relationship between gratitude and leadership. Previous researchers have discussed the role of leaders' emotions in the staff's performance and productivity (Humphrey, 2002; Garg, 2017b). Combining the results of these studies with the findings of this study reveals that gratitude can be doubly beneficial for leaders. Firstly, gratitude can sharpen leadership skills and secondly, as emotion, gratitude can influence subordinates' performance. Future researchers can undoubtedly investigate the impact of grateful leaders on follower's performance. Positive emotions are contagious (Hatfield *et al.*, 1993), and feelings of organizational leaders are especially contagious (George, 1995). Thus, the gratuitous behavior of leaders may produce a cascading effect through eliciting a similar response from subordinates.

Fourthly, this resolves the "black-box" of a path through which the effect of gratitude transmits to leadership. The results show that SA, SS and SIP partially mediate the relationship between gratitude and leadership. In summary, gratitude influences social intelligence, which horns one's leadership skills. Again, this study is one of the first explorations to explore the mediating and moderating effect of social intelligence amid the relationship between gratitude and transformational leadership. Then finally, this study is conducted in Indian settings, and most of the previous studies on gratitude were conducted in western contexts. As discussed, the findings of this study are complementing the broaden-and-build theory and social attachment perspective of a social cognitive theory of emotions. These theories were developed in non-eastern conditions. Given vast cultural differences, this study provides an appropriate avenue of application of these theories in Indian states.

Practical implications of the study

The findings of this study provide several practical implications for practitioners too. Firstly, leadership skill trainers can help participants in cultivating gratitude to promote leadership skills. Trainers may use different gratitude practices and interventions including "gratitude journals" (Froh *et al.*, 2008; Froh *et al.*, 2007), "gratitude letters" (Seligman, 2002; Sergeant and Mongrain, 2011), "gratitude letters visits" "cultivating mindfulness" (Langer, 2014; Langer *et al.*, 2010), "gratitude collages" (Reivich, 2009), "sharing positive events" (Gable *et al.*, 2004), "acts of kindness" (Sheldon *et al.*, 2013) and "meditation and progressive muscle relaxation techniques" (Shapiro, Schwartz and Santerre, 2002). Secondly, as discussed, leaders' expression of gratefulness is not only beneficial for leaders; it also influences subordinates' performance too. Thus, organizations are suggested to motivate their line managers to cultivate the virtue of gratitude. This could quickly be done by imbibing gratuitous behavior in organizational culture. Again as suggested by previous researchers that positive emotions are contagious, gratuitous behavior by top management will motivate line managers to follow suit. Thirdly, findings suggest that social intelligence is a significant predictor of leadership. Although more researches are required for generalization, a recruiter may start discussions and debates on the inclusion of social intelligence test for the selection of managers.

Also, Fehr *et al.* (2017) proposed a multilevel model of gratitude in organizations. According to this model, gratitude could be observed at three levels in the workplace, namely, episodic gratitude ("feeling grateful in response to a particular experience"),

persistent gratitude ("a stable tendency to feel grateful within a particular context") and collective gratitude ("persistent gratitude that is shared by the members of an organization"). The authors also argued that these different forms of gratitude complement and support each other. If an organization facilitates institutionalization of gratitude at organizational level, then it facilitates cultivation of both episodic and persistent gratitude at individual level. Thus, the present study recommends organizational level gratitude programs such as "Formal appreciation program," "Thanks-giving day or week," and "Personalized gratuitous letters from managers and leaders." In recent commendable initiatives, contemporary organizations have started paid leaves to employees to serve the needy people. Under the Corporate Social Responsibility initiative, these companies allow a group of employees to submit a socially-beneficial project. If such a project is approved, the organization not only funds the project, but employees are also given paid leaves. Such schemes foster an organizational culture based on gratitude and trust.

Limitations, scope for future studies and conclusion

The present study provides an empirical model of gratitude, social intelligence, and transformational leadership. However, these contributions come in light of a few limitations. Firstly, this study is based on a cross-sectional research design, and also data is collected using self-reported measures. Thus, there are chances of common method variance (CMV). Although CMV was not reported in this study, a longitudinal study is required to conclude any causal and hypothesized relationships. Future researchers may conduct longitudinal studies to confirm the findings of this study. Secondly, the present study is based on a sample of 252 employees only. A larger sample comprising of respondents from diverse industries would be appropriate for generalization of the results. Future research should explore a larger sample spread over different companies. Thirdly, this study examined the mediating effect of social intelligence only. Future researchers may test a range of variables such as contentment, satisfaction perceived organizational support for possible mediation effect. Also, gratitude is measured with the help of a one-dimensional scale (GQ-6). Future researchers may use multidimensional scales like GRAT for better results. The present has considered consolidated values of gratitude and leadership measures, and future researchers may opt for in-depth analysis by considering different constructs of gratitude and leadership. It would be interesting to examine

In conclusion, the study provides empirical evidence supporting a positive association between gratitude and transformational leadership. Moreover, all three dimensions of social intelligence proved to be partially mediators of the relationship between gratitude and leadership. Besides, SS were also reported as moderator of a hypothesized relationship. As this study is a pioneer study in the field of gratitude, social intelligence and leadership, the results bring significant contributions that enhance understanding and support further research in this specific area of study.

References

Agran, M., Hughes, C., Thoma, C.A. and Scott, L.A. (2016), "Employment social skills: what skills are really valued?", *Career Development and Transition for Exceptional Individuals*, Vol. 39 No. 2, pp. 111-120.

Algoe, S.B. and Stanton, A.L. (2012), "Gratitude when it is needed most: social functions of gratitude in women with metastatic breast cancer", *Emotion*, Vol. 12 No. 1, pp. 163-168.

Alkozei, A., Smith, R. and Killgore, W.D. (2018), "Gratitude and subjective wellbeing: a proposal of two causal frameworks", *Journal of Happiness Studies*, Vol. 19, pp. 1519-1542.

Baron, R.M. and Kenny, D.A. (1986), "The moderator-mediator variable distinction in social psychological research: conceptual, strategic and statistical considerations", *Journal of Personality and Social Psychology*, Vol. 51 No. 6, pp. 1173-1182.

Bartlett, M.Y. and DeSteno, D. (2006), "Gratitude and prosocial behavior helping when it costs you", *Psychological Science*, Vol. 17 No. 4, pp. 319-325.

Bass, B.M. (1990), "From transactional to transformational leadership: learning to share the vision", *Organizational Dynamics*, Vol. 18 No. 3, pp. 19-31.

Boyatzis, R.E. and Saatcioglu, A. (2008), "A 20-year view of trying to develop emotional, social and cognitive intelligence competencies in graduate management education", *Journal of Management Development*, Vol. 27 No. 1, pp. 92-108.

Bryman, A. and Bell, E. (2014), *Research Methodology: business and Management Contexts*, *Oxford University Press Southern Africa*, *Cape Town*.

Cronbach, L.J. and Shavelson, R.J. (2004), "My current thoughts on coefficient alpha and successor procedures", *Educational and Psychological Measurement*, Vol. 64 No. 3, pp. 391-418.

Davies, M., Cooper, G., Kettler, R.J. and Elliott, S.N. (2015), "Developing social skills of students with additional needs within the context of the Australian curriculum", *Australasian Journal of Special Education*, Vol. 39 No. 1, pp. 37-55.

Dik, B.J., Duffy, R.D., Allan, B.A., O'Donnell, M.B., Shim, Y. and Steger, M.F. (2014), "Purpose and meaning in career development applications", *The Counseling Psychologist*, Vol. 43 No. 4, pp. 1-28, doi: 10.1177/0011000014546872.

Edmondson, A. (2002), "The local and variegated nature of learning in organizations: a grouplevel perspective", *Organization Science*, Vol. 13 No. 2, pp. 128-146.

Emmons, R.A. and Crumpler, C.A. (2000), "Gratitude as human strength: appraising the evidence", *Journal of Social and Clinical Psychology*, Vol. 19 No. 1, pp. 56-69, doi: 10.1521/jscp.2000. 19.1.56.

Emmons, R.A. and McCullough, M.E. (2003), "Counting blessings versus burdens: an experimental investigation of gratitude and subjective well-being in daily life", *Journal of Personality and Social Psychology*, Vol. 84 No. 2, pp. 377-389.

Fehr, R., Fulmer, A., Awtrey, E. and Miller, J.A. (2017), "The grateful workplace: a multilevel model of gratitude in organizations", *Academy of Management Review*, Vol. 42 No. 2, pp. 361-381, doi: 10.5465/amr.2014.0374.

Fornell, C. and Larcker, D.F. (1981), "Evaluating structural equation models with unobservable variables and measurement error", *Journal of Marketing Research*, Vol. 18 No. 1, pp. 39-50.

Fredrickson, B.L. (1998), "What good are the positive emotions?", *Review of General Psychology*, Vol. 2 No. 3, pp. 300-319.

Fredrickson, B.L. (2004), "The broaden-and-build theory of positive emotions", *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, Vol. 359 No. 1449, pp. 1367-1378.

Fritz, M.S., Taylor, A.B. and MacKinnon, D.P. (2012), "Explanation of two anomalous results in statistical mediation analysis", *Multivariate Behavioral Research*, Vol. 47 No. 1, pp. 61-87.

Froh, J.J., Miller, D.N. and Snyder, S.F. (2007), "Gratitude in children and adolescents: development, assessment, and school-based intervention", *School Psychology Forum*, Vol. 2, pp. 1-13.

Froh, J.J., Sefick, W.J. and Emmons, R.A. (2008), "Counting blessings in early adolescents: an experimental study of gratitude and subjective well-being", *Journal of School Psychology*, Vol. 46 No. 2, pp. 213-233.

Froh, J.J., Emmons, R.A., Card, N.A., Bono, G. and Wilson, J.A. (2011), "Gratitude and the reduced costs of materialism in adolescents", *Journal of Happiness Studies*, Vol. 12 No. 2, pp. 289-302.

Gable, S.L., Reis, H.T., Impett, E.A. and Asher, E.R. (2004), "What do you do when things go right? The intrapersonal and interpersonal benefits of sharing positive events", *Journal of Personality and Social Psychology*, Vol. 87 No. 2, pp. 228-245.

Garg, N. (2017a), "Workplace spirituality and employee well-being: an empirical exploration", *Journal of Human Values*, Vol. 23 No. 2, pp. 1-19.

Garg, N. (2017b), "Workplace spirituality and organizational commitment: an empirical exploration", *Purushartha- A Journal of Management, Ethics, and Spirituality*, Vol. 10 No. 2, pp. 8-24.

Garg, N. (2018), "Promoting organizational performance in indian insurance industry: the roles of workplace spirituality and organizational citizenship behaviour", *Global Business Review*, Vol. 1, doi: 10.1177/0972150918778983.

Garg, N. (2020), "Exploring role of gratitude in developing teacher leadership in Indian universities", *International Journal of Educational Management*, Vol. 34 No. 5, pp. 881-901, available at: www.emerald.com/insight/publication/issn/0951-354X

Garg, N. and Gera, S. (2019), "Gratitude and leadership in higher education institutions: exploring the mediating role of social intelligence among teachers", *Journal of Applied Research in Higher Education*, Vol. ahead-of-print No. ahead-of-print.

Garg, N., Punia, B.K. and Jain, A. (2019), "Workplace spirituality and job satisfaction: exploring mediating effect of organizational citizenship behaviour", *Vision: The Journal of Business Perspective*, Vol. 23 No. 3, pp. 287-296.

George, J.M. (1995), "Leader positive mood and group performance: the case of customer service", *Journal of Applied Social Psychology*, Vol. 25, pp. 778-794.

George, D. and Mallery, P. (2003), *SPSS for Windows Step by Step: A Simple Guide and Reference*, 4the ed., Allyn, Bacon, Boston, 11.0 update.

Giles, D. and Yates, R. (2014), "Enabling educational leaders: qualitatively surveying an organization's culture", *International Journal of Organizational Analysis*, Vol. 22 No. 1, pp. 94-106.

Goleman, D. (2006), *Social Intelligence: The New Science of Human Relationships*, Bantam Books.

Grant, A.M. and Gino, F. (2010), "A little thanks goes a long way: explaining why gratitude expressions motivate prosocial behaviour", *Journal of Personality and Social Psychology*, Vol. 98 No. 6, pp. 946-955.

Grant, A.M., Wrzesniewski, A. (2010), "I won't let you down . . . or will I? Core Self-evaluations, other-orientation, anticipated guilt and gratitude, and job performance", *Journal of Applied Psychology*, Vol. 95 No. 1, pp. 108-121.

Greer, C.R. and Plunkett, W.R. (2000), *Supervision: Diversity and Teams in the Workplace*, 9th ed., Prentice Hall, Upper Saddle River, NJ.

Hair, J.F., Black, W.C., Babin, B.J. and Anderson, R.E. (2010), *Multivariate Data Analysis*, 7th ed., Prentice Hall, Upper Saddle River, NJ.

Hatfield, E. Cacioppo, J.T. and Rapson, R.L. (1993), "Emotional contagion", Vol. 2, No. 3, pp. 96-100, available at: https://doi.org/10.1111%2F1467-8721.ep10770953

Howells, K. (2013), "The role of gratitude in helping school leaders strengthen relationships", *Perspectives on Educational Leadership*, Vol. 6 No. 1, pp. 1-3.

Hu, X. and Kaplan, S. (2014), "Is "feeling good" good enough? Differentiating discrete positive emotions at work", *Journal of Organizational Behavior*, Vol. 36 No. 1, pp. 39-58.

Humphrey, R.H. (2002), "The many faces of emotional leadership", *Leadership Quarterly*, Vol. 13, No. 5, pp. 493-504.

Jans-Beken, L., Lataster, J., Peels, D., Lechner, L. and Jacobs, N. (2018), "Gratitude, psychopathology and subjective well-being: results from a 7.5-month prospective general population study", *Journal of Happiness Studies*, Vol. 19, pp. 1673-1689.

Kaplan, S., Bradley-Geist, J.C., Ahmad, A., Anderson, A., Hargrove, A.K. and Lindsey, A. (2014), "Test of two positive psychology interventions to increase employee well-being", *Journal of Business and Psychology*, Vol. 29 No. 3, pp. 367-380.

Kline, R.B. (2009), *Becoming a Behavioural Science Researcher: A Guide to Producing Research That Matters*, Guildford Press New York, NY, Chapter 7, Measurement.

Kong, F., Ding, K. and Zhao, J. (2015), "The relationships among gratitude, self-esteem, social support and life satisfaction among undergraduate students", *Journal of Happiness Studies*, Vol. 16 No. 2, pp. 477-489.

Langer, E.J. (2014), "Mindfulness forward and back", *Wiley Online Library*, available at: https://doi.org/10.1002/9781118294895.ch1

Langer, E., Djikic, M., Pirson, M., Madenci, A. and Donohue, R. (2010), "Believing is seeing using mindlessness (mindfully) to improve visual acuity", *Psychological Science*, Vol. 21, pp. 661-666.

Malhotra, N.K. and Dash, S. (2018), "Marketing research an applied orientation", Seventh Edition Revised.

Marlowe, H.A. (1986), "Social intelligence: evidence for multidimensionality and construct independence", *Journal of Educational Psychology*, Vol. 78 No. 1, pp. 52-58.

McCullough, M.E., Emmons, R.A. and Tsang, J.A. (2002), "The grateful disposition: a conceptual and empirical topography", *Journal of Personality and Social Psychology*, Vol. 82 No. 1, pp. 112-127.

McCullough, M.E., Emmons, R.A., Kilpatrick, S.D. and Larson, D.B. (2001), "Is gratitude a moral affect?", *Psychological Bulletin*, Vol. 127 No. 2, pp. 249-266.

Nieuwhof, C. (2017), "5 Crucial ways your gratitude impacts your attitude as leader", available at: https://careynieuwhof.com/5-ways-your-gratitude-impacts-your-attitude-as-a-leader/ (accessed 12 November 2019).

Northouse, G. (2007), *Leadership Theory and Practice*, 3rd ed., Sage Publications, Thousand Oaks, CA.

Peterson, B.E. and Stewart, A.J. (1996), "Antecedents and contexts of generativity motivation at mid-life", *Psychology and Aging*, Vol. 11, pp. 21-33.

Podsakoff, P.M., MacKenzie, S.B. and Podsakoff, N.P. (2012), "Sources of method bias in social science research and recommendations on how to control it", *Annual Review of Psychology*, Vol. 63 No. 1, pp. 539-569.

Reivich, K. (2009), "Fishful thinking: Cultivating gratitude in youth", *National Association of School Psychologists*, Vol. 38 No. 3, available at: www.nasponline.org/publications/cq/index.aspx

Rosenberg, E.L. (1998), "Levels of analysis and the organization of affect", *Review of General Psychology*, Vol. 2 No. 3, pp. 247-270.

Rowold, J. (2004), "MLQ-5X for sport coaches", German translation and adaptation of Bass & Avolio's Multifactor Leadership Questionnaire Redwood City: Mind Garden.

Sansone, R.A. and Sansone, L.A. (2010), "Gratitude and well-being: the benefits of appreciation", *Psychiatry*, Vol. 7 No. 11, pp. 18-22.

Seligman, M.E.P. (2002), *Authentic Happiness: Using the New Positive Psychology to Realize Your Potential for Lasting Fulfillment*, *Free Press*, *New York, NY*.

Sergeant, S. and Mongrain, M. (2011), "Are positive psychology exercises helpful for people with depressive personality styles?", *The Journal of Positive Psychology*, Vol. 6 No. 4, pp. 260-272, doi: 10.1080/17439760.2011.577089.

Shapiro, S.L., Schwartz, G.E.R. and Santerre, C. (2002), "Meditation and positive psychology", in Snyder, C.R. and Lopez, S. (Eds), *Handbook of Positive Psychology*, Oxford University Press, London, pp. 632-645.

Sheldon, K.M., Boehm, J.K. and Lyubomirsky, S. (2013), "Variety is the spice of happiness: the hedonic adaptation prevention (HAP) model", in Boniwell I. and David S. (Eds), *Oxford Handbook of Happiness*, Oxford University Press, Oxford, pp. 901-914.

Shrout, P.E. and Bolger, N. (2002), "Mediation in experimental and nonexperimental studies: new procedures and recommendations", *Psychological Methods*, Vol. 7 No. 4, pp. 422-445.

Silvera, D.H., Martinussen, M. and Dahl, T.I. (2001), "The tromsø social intelligence scale, a self-report measure of social intelligence", *Scandinavian Journal of Psychology*, Vol. 42 No. 4, pp. 313-319.

Spence, J.R., Brown, D.J. and Keeping, L.M., L., H. (2013), "Helpful today, but not tomorrow? Feeling grateful as a predictor of daily organizational citizenship behaviors", *Pers. Psychol*, Vol. 67, pp. 705-738.

Sun, P. and Kong, F. (2013), "Affective mediators of the influence of gratitude on life satisfaction in late adolescence", *Social Indicator Research*, Vol. 114 No. 3, pp. 1361-1369.

Uhder, J., Watkins, P.C. and Hammamoto, D. (2010), "Would the humble please stand: Can self-reported humility be valid?", Paper presented to the annual convention of the American Psychological Association, San Diego, CA.

Waters, L. (2012), "Predicting job satisfaction: contributions of individual gratitude and institutionalized gratitude", *Psychology*, Vol. 03 No. 12, pp. 1174-1176.

Waters, L. and Stokes, H. (2015), "Positive education for school leaders: exploring the effects of emotion-gratitude and action-gratitude", *The Educational and Developmental Psychologist*, Vol. 32 No. 1, pp. 1-22.

Watkins, P.C., Uhder, J. and Pichinevskiy, S. (2014), "Grateful recounting enchances subjective well-being: the importance of grateful processing", *The Journal of Positive Psychology*, Vol. 10 No. 2, pp. 91-98.

Watkins, P.C., Woodward, K., Stone, T. and Kolts, R.L. (2003), "Gratitude and happiness: development of a measure of gratitude, and relationships with subjective well-being", *Social Behavior and Personality: An International Journal*, Vol. 31 No. 5, pp. 431-451.

Wood, A.M., Froh, J.J. and Geraghty, A.W.A. (2010), "Gratitude and well-being: a review and theoretical integration", *Clinical Psychology Review*, Vol. 30 No. 7.

Yay, M. (2017), "The mediation analysis with the sobel test and the percentile bootstrap", *International Journal of Management and Applied Science (IJMAS)*, Vol. 3 No. 2, pp. 123-125.

Yüksel, A. and Oguz Duran, N. (2012), "Turkish adaptation of the gratitude questionnaire", *Eurasian Journal of Educational Research*, Vol. 46, pp. 199-216.

Zhao, X., Lynch, J.G. and Chen, Q. (2010), "Reconsidering Baron and Kenny: myths and truths about mediation analysis", *Journal of Consumer Research*, Vol. 37 No. 2, pp. 197-206.

Further reading

Al-Hashimi, M. and Al-Barri, Z. (2017), "Gratitude intervention at work: its impact on well-being, job satisfaction and organisational commitment", *International Journal Advances in Social Science and Humanities*, Vol. 5 No. 3, pp. 01-10.

Bartlett, M.Y., Condon, P., Cruz, J., Baumann, J. and Desteno, D. (2012), "Gratitude: prompting behaviors that build relationships", *Cognition and Emotion*, Vol. 26 No. 1, pp. 2-13.

Bono, G., Emmons, R.E. and McCullough, M.E. (2004), "Gratitude in practice and the practice of gratitude", *Positive Psychology in Practice*, Vol. 1, pp. 464-481.

Bothma, C.F.C. and Roodt, G. (2013), "The validation of the turnover intention scale", *SA Journal of Human Resource Management*, Vol. 11 No. 1.

Chen, X.P., Hui, C. and Sego, D.J. (1998), "The role of organizational citizenship behavior in turnover: Conceptualization and preliminary tests of key hypotheses", *Journal of Applied Psychology*, Vol. 83 No. 6, pp. 922-931.

Coyne, I. and Ong, T. (2007), "Organizational citizenship behavior and turnover intention: a cross-cultural study", *The International Journal of Human Resource Management*, Vol. 18 No. 6, pp. 1085-1097.

Du Plooy, J. and Roodt, G. (2010), "Work engagement, burnout, and related constructs as predictors of turnover intentions", *SA Journal of Industrial Psychology*, Vol. 36 No. 1.

Emmons, R.A. and Mishra, A. (2010), "Why gratitude enhances well-being: what we know", *What We Need to Know," Designing Positive Psychology: Taking Stock and Moving Forward*, *Oxford University Press*, *New York, NY*, pp. 248-262.

Emmons, R.A. and Stern, R. (2013), "Gratitude as a psychotherapeutic intervention", *Journal of Clinical Psychology*, Vol. 69 No. 8, pp. 846-855.

Erdil, O. and Müceldili, B. (2014), "The effects of envy on job engagement and turnover intention", *Procedia - Social and Behavioral Sciences*, Vol. 150, pp. 447-454.

Fabio, A.D., Palazzeschi, L. and Bucci, O. (2017), "Gratitude in organizations: a contribution for healthy organizational contexts", *Frontiers in Psychology*, Vol. 8, pp. 1-6.

Fehr, R., Fulmer, A., Awtrey, E. and Miller, J.A. (2014), "The grateful workplace: a multilevel model of gratitude in organizations", *Academy of Management Review*, Vol. 1

Fitzgerald, P. (1998), "Gratitude and justice", *Ethics*, Vol. 109 No. 1, pp. 119-153.

Foundation, J.T., Good, G., May, U.C.B. and Allen, S. (2018), "The science of gratitude", *The University of California at Berkeley Wellness Letter*, Vol. 32 No. 6.

Froman, L. (2009), *Positive Psychology in the Workplace*, Springer, New York, NY, pp. 59-69.

Garg, N. (2017c), "Workplace spirituality and organizational performance in indian context: mediating effect of organizational commitment, work motivation, and employee engagement", *South Asian Journal of Human Resources Management*, Vol. 4 No. 2, pp. 191-211.

Griffeth, R., Hom, P.W. and Gaertner, S. (2000), "A meta-analysis of antecedents and correlates of employee turnover: update, moderator tests, and research implications for the next millennium", *Journal of Management*, Vol. 26 No. 3, pp. 463-488.

Lambert, N.M., Fincham, F.D., Stillman, T.F. and Dean, L.R. (2009), "More gratitude, less materialism: the mediating role of life satisfaction", *The Journal of Positive Psychology*, Vol. 4 No. 1, pp. 32-42.

Layous, K., Sweeny, K., Armenta, C., Na, S., Choi, I. and Lyubomirsky, S. (2017), "The proximal experience of gratitude", *Plos One*, Vol. 12 No. 7, pp. 1-26.

Lin, C.P. (2008), "Clarifying the relationship between organizational citizenship behaviors, gender, and knowledge sharing in workplace organizations in Taiwan", *Journal of Business and Psychology*, Vol. 22 No. 3, pp. 241-250.

Lin, C.C. (2013), "A higher-Order gratitude uniquely predicts subjective well-being: incremental validity above the personality and a single gratitude", *Social Indicators Research*, Vol. 119 No. 2, pp. 909-924.

McCullough, M.E., Tsang, J.A. and Emmons, R.A. (2004), "Gratitude in intermediate affective terrain: links of grateful moods to individual differences and daily emotional experience", *Journal of Personality and Social Psychology*, Vol. 86 No. 2, pp. 295-309.

Podsakoff, P.M., MacKenzie, S.B., Paine, J.B. and Bachrach, D.G. (2000), "Organizational citizenship behaviors: a critical review of the theoretical and empirical literature and suggestions for future research", *Journal of Management*, Vol. 26 No. 3, pp. 513-563.

Podsakoff, N.P., Whiting, S.W., Podsakoff, P.M. and Blume, B.D. (2009), "Individual- and organizational-level consequences of organizational citizenship behaviors: a meta-analysis", *Journal of Applied Psychology*, Vol. 94 No. 1, pp. 122-141.

Podsakoff, N.P., Podsakoff, P.M., Mackenzie, S.B., Maynes, T.D. and Spoelma, T.M. (2014), "Consequences of unit-level organizational citizenship behaviors: a review and recommendations for future research", *Journal of Organizational Behavior*, Vol. 35, pp. 87-119.

Shanker, M. (2018), "Organizational citizenship behavior about employees' intention to stay in indian organizations", *Business Process Management Journal*, Vol. 24 No. 6.

Solom, R., Watkins, P.C., McCurrach, D. and Scheibe, D. (2016), "Thieves of thankfulness: traits that inhibit gratitude", *The Journal of Positive Psychology*, Vol. 12 No. 2, pp. 120-129.

Spence, J.R. (2013), "Helpful today but not tomorrow? Feeling grateful as a predictor of daily organizational citizenship behaviours", *Personnel Psychology*, No. 67, pp. 705-738.

Spence, J.R. (2014), "Helpful today but not tomorrow? Feeling grateful as a predictor of daily organizational citizenship behaviours", *Personnel Psychology*, No. 67, pp. 705-738.

Wood, A.M., Joseph, S. and Linley, P.A. (2007), "Coping style as a psychological resource of grateful people", *Journal of Social and Clinical Psychology*, Vol. 26 No. 9, pp. 1076-1093.

Appendix. Questionnaire

Gratitude scale

(1) I have so much in life to be thankful for.
(2) If I had to list everything that I felt grateful for, it would be a very long list.
(3) When I look at the world, I do not see much to be grateful for.*
(4) I am grateful to a wide variety of people.
(5) As I get older I find myself more able to appreciate the people, events, and situations that have been part of my life history.
(6) Long amounts of time can go by before I feel grateful to something or someone.

* *Items 3 and 6 are reverse-scored.

Social intelligence scale

(1) I can predict other people's behavior.
(2) I often feel that it is difficult to understand other's choices.
(3) I know how my actions will make others feel.
(4) I often feel uncertain around new people who I do not know.
(5) People often surprise me with the things they do.
(6) I understand other people's feelings.
(7) I fit in easily in social situations.
(8) Other people become angry with me without me being able to explain why.
(9) I understand other's wishes.
(10) I am good at entering new situations and meeting people for the first time.
(11) It seems as though people are often angry or irritated when I say what I think.
(12) I have a hard time getting along with other people.
(13) I find people unpredictable.
(14) I can often understand what others are trying to accomplish without the need for them to say anything.
(15) It takes a long time for me to get to know others well.
(16) I have often hurt others without realizing it.
(17) I can predict how others will react to my behavior.
(18) I am good at getting on good terms with new people.
(19) I can often understand what others really mean through their expression, body language etc.
(20) I frequently have problems finding good conversation topics.
(21) I am often surprised by others' reactions to what I do.

Leadership scale

(1) I make others feel good to be around me.
(2) I express with a few simple words what we could and should do.
(3) I enable others to think about old problems in new ways.
(4) I help others develop themselves.

(5) I tell others what to do if they want to be rewarded for their work.
(6) I am satisfied when others meet agreed-upon standards.
(7) I am content to let others continue working in the same way as always.
(8) Others have complete faith in me.
(9) I provide appealing images about what we can do.
(10) I provide others with new ways of looking at puzzling things.
(11) I let others know how I think they are doing.
(12) I provide recognition/rewards when others reach their goals.
(13) As long as things are working, I do not try to change anything.
(14) Whatever others want to do is O.K. with me.
(15) Others are proud to be associated with me.
(16) I help others find meaning in their work.
(17) I get others to rethink ideas that they had never questioned before.
(18) I give personal attention to others who seem rejected.
(19) I call attention to what others can get for what they accomplish.
(20) I tell others the standards they have to know to carry out their work.
(21) I ask no more of others than what is absolutely essential.

Corresponding author
Naval Garg can be contacted at: naval.garg@dtu.ac.in

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2999267, IEEE Journal of Emerging and Selected Topics in Power Electronics

Implementation Strategy for Resource Optimization of FPGA-Based Adaptive Finite Control Set-MPC using XSG for a VSI System

Vijay Kumar Singh, Ravi Nath Tripathi, and Tsuyoshi Hanamoto

***Abstract*— Finite control set-model predictive control (FCS-MPC) for power electronic systems has been implemented using field-programmable gate array (FPGA) considering an alternative solution to handle the computational burden of the control algorithm. However, FPGA resource utilization is a concern for optimum design and development of the FCS-MPC. This paper presents, an analytical approach based implementation strategies using the Xilinx system generator (XSG) considering the real-time operation of the FCS-MPC for a three-phase voltage source inverter (VSI). Moreover, an adaptive prediction model is proposed to adapt to the change in load current. The percentage THD, dynamic response, and steady-state error are evaluated for system performance analysis. The implementation strategies are compared based on FPGA resource utilization for the FCS-MPC in stationary *αβ* and rotating *dq* frames considering fixed, approximated and adaptive values of a coefficient used for current prediction. The look-up tables (LUTs), flip-flops and DSP slices have opted as utilization indices of FPGA resources for comparative analysis. The optimum design based controller model is used for FPGA-based experimental system implementation. The simulation and experimental results are used for the performance evaluation of the FCS-MPC in both *αβ* and *dq* frames. Moreover, change in switching state constraint is included in the cost function and controller performance is evaluated corresponding to the weighting factor for approximated and adaptive coefficient values.**

***Index Terms*— finite control set-model predictive control, field-programmable gate array, model-based design, voltage source inverter, Xilinx system generator.**

I. INTRODUCTION

MODEL predictive control (MPC) possesses appealing characteristics such as explicit use of a system model, flexibility of simultaneous handling of multiple constraints and easy inclusion of nonlinearities [1]–[4]. These distinctive features of MPC have created considerable attention towards researcher and MPC has been applied to a wide variety of power converter and drive applications. For instance, two-level voltage source inverter (VSI) [5], [6], three-level neutral-point clamped converter (NPC) [7], active front end rectifier [8], [9], cascaded H-Bridge inverter [10], three-phase direct matrix converters [11], [12], predictive control for UPS applications [13], predictive torque control (PTC) and field-oriented control (FOC) of an induction machine [2], [14], just to name a few.

The fundamental principle of the MPC is to have a plant or system model that is used for the prediction of the future behavior of control parameters and the control action is selected based on an optimization function or cost function in each sampling interval. Finite control set-model predictive control (FCS-MPC) is one of the classifications of MPC that uses the discrete-time model of a power converter. The cost function is computed based on the finite number of switching states of the power converter and the control action is directly applied to the converter. As compared to the generalized MPC method, the computational requirement is reduced in the FCS-MPC by using the discretized power converter model and a small prediction horizon (N_P = 1). The computations required during the implementation of predictive controls for power converters are accomplished by the microprocessor and digital signal processor (DSP) [3], [4], [14].

The prediction and cost function computations required for the algorithm in the practical implementation should strictly complete within a given sampling interval. However, the computational task is a concern especially for complex converter systems and creates a delay issue that deteriorates the quality of waveforms [5], [15]. This issue has been addressed and some methodologies for delay compensation have been presented to compute the optimum switching state within the specified sampling interval [16], [17]. However, delay compensation techniques have to deal with the increased computational burden and average switching frequency [16].

Field programmable gate array (FPGA) is a pragmatic choice for the implementation of MPC because of parallel processing capability [18]–[24]. In [18], a comparative analysis among various digital control platforms is presented for the implementation of model predictive control including FPGA. Moreover, the different approach is adopted for FPGA-based system implementation. In [19], an FPGA platform of NI-CRIO

The manuscript was submitted on November 18, 2019 for review.

V. K. Singh is with the Department of Life Science and Systems Engineering, Kyushu Institute of Technology, Wakamatsu-ku, Kitakyushu-shi, Fukuoka 808-0196 Japan (e-mail: vijay.ec1985@gmail.com).

R. N. Tripathi is with the Next Generation Power Electronics Research Center, Kyushu Institute of Technology, Wakamatsu-ku, Kitakyushu-shi, Fukuoka 808-0196 Japan (e-mail: ravi1989tripathi@gmail.com).

T. Hanamoto is with the Department of Life Science and Systems Engineering, Kyushu Institute of Technology, Wakamatsu-ku, Kitakyushu-shi, Fukuoka 808-0196 Japan (e-mail: hanamoto@life.kyutech.ac.jp).

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2020 at 05:20:42 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2999267, IEEE Journal of Emerging and Selected Topics in Power Electronics

reconfigurable system is used to implement MPC for a back-to-back converter. The FPGA implementation of MPC is performed using a model-based design (MBD) through MATLAB/Simulink in [20], [21] and FPGA code was obtained using hardware description language (HDL) coder functionality of Simulink. In [22], a high-speed FCS-MPC implementation is presented with a control algorithm coded in C++ using PROTOIP toolbox and HDL code was generated through Xilinx Vivado HLS. Xilinx system generator (XSG) as a digital simulator adopting the MBD platform was used to implement an FPGA-based predictive current controller in [23]–[26]. XSG is an integrated platform with MATLAB/Simulink and facilitates the HDL code generation for the developed system.

In addition to the implementation approach, resource utilization is another key aspect of FPGA-based system control implementation. The system control implementation based on FPGA resource utilization is compared in [22] and [25] for Xilinx FPGA. In [22], the prediction horizon is considered for the comparison of FPGA resource utilization. In [25], semi-parallel and fully serial implementation approaches for MPC are adopted and resource utilization is compared for the same. The look-up tables (LUTs), flip-flops and DSP slices are the main digital logic indices considered for resource utilization in the aforementioned papers.

There is a possibility to adopt different implementation strategies for an FPGA implementation to enhance resource utilization. In this way, the MBD approach can be advantageous by allowing different implementation strategies under the same environment and to compare the FPGA resource utilization. Moreover, the MBD approach provides better system visualization and easy debugging that is appealing for rapid controller prototyping.

This paper presents analytical implementation strategies using an MBD approach considering the reduction in the FPGA resource utilization for fully parallel implementation of the FCS-MPC algorithm. The predictive model coefficient computed using load parameters is one of the significant factors that influence system performance. Furthermore, an adaptive predictive model is proposed to compensate for the effect of change in load parameters by using simple mathematical relation of the three-phase inverter output voltage, dc voltage supply, and modulation index. The unity and non-unity modulation index is also included to investigate the proposed adaptive model design considering system performance.

A comparative analysis of FPGA resource utilization using analytical implementation strategies is considered corresponding to the predictive model coefficient of the FCS-MPC for a three-phase VSI system with RL load. Moreover, resource utilization for controller implementation using stationary $\alpha\beta$ and rotating dq reference frames is used for in-depth analysis. In general, the cost function is formulated based on stationary $\alpha\beta$ coordinates [27], [28]. Considering current control with the FCS-MPC, the sinusoidal future current references need be predicted accurately using an extrapolation method, however, the extrapolated reference causes unwanted oscillations that influence transient response [7], [9], [27]. Considering the above issue, the cost function can be designed using a rotating dq frame that does not require the extrapolation method due to continuous current references [27]. However, the comparative analysis based on design constraints and performance indices is required to be addressed for two frames. Furthermore, the computational burden, as well as the FPGA resource requirement corresponding to each coordinate system, is crucial for practical system implementation.

The FPGA-based implementation of the FCS-MPC is performed using the MBD approach provided by the XSG integrated to MATLAB/Simulink. The system performance is analyzed based on the implementation of the FCS-MPC using optimum FPGA resource utilization in $\alpha\beta$ and dq frames for the adaptive predictive model (unity and non-unity modulation index). The dynamic behavior and steady-state error are considered for analysis of system performance. Furthermore, a constraint of change in switching state [2], [29] is considered to analyze the performance of average switching frequency reduction and the current harmonics for the adaptive predictive model.

The paper is organized as follows: Section II describes the algorithm of the FCS-MPC for both stationary $\alpha\beta$ as well as rotating dq frames considering the discrete-time mathematical model of the three-phase VSI system. In Section III, the model-based design and development of the controller in the digital simulator are explained. Section IV covers the different implementation strategies considered for the development of the FCS-MPC algorithm and the comparative FPGA resource utilization. The simulation and experimental results are presented with detailed discussion in Section V. Finally, appropriate conclusions are drawn in Section VII.

II. Finite Control Set-MPC

The power circuit of a three-phase VSI with *RL* load is shown in Fig. 1 where R is the load resistance and L is the load inductance. The switching states of the converter are determined by the switching signals S_a, S_b, and S_c as shown in Table I. Considering all possible switching combinations of the switching signals, eight switching states $\boldsymbol{S}$ ($\boldsymbol{S}_0$–$\boldsymbol{S}_7$) and hence, eight voltage vectors $\boldsymbol{v}_i$ ($\boldsymbol{v}_0$–$\boldsymbol{v}_7$) are obtained as shown in Table II.

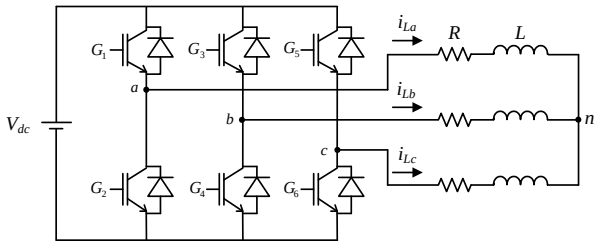


Fig. 1. Power circuit of three-phase VSI with RL load.

TABLE I
Gating Signals of Inverter

Leg 'a', S_a	Leg 'b', S_b	Leg 'c', S_c
G_1 ON, 1 G_2 OFF, 0	G_3 ON, 1 G_4 OFF, 0	G_5 ON, 1 G_6 OFF, 0
G_1 OFF, 0 G_2 ON, 1	G_3 OFF, 0 G_4 ON, 1	G_5 OFF, 0 G_6 ON, 1

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2020 at 05:20:42 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2999267, IEEE Journal of Emerging and Selected Topics in Power Electronics

TABLE II
VOLTAGE VECTORS AND SWITCHING STATES

Switching states	Voltage vectors
$\boldsymbol{S} = [S_a \ S_b \ S_c]$	$\boldsymbol{v_i} = [v_{i\alpha} \ v_{i\beta}]$
$\boldsymbol{S}_0 = [0 \ 0 \ 0]$	$\boldsymbol{v}_0 = [0, 0]$
$\boldsymbol{S}_1 = [1 \ 0 \ 0]$	$\boldsymbol{v}_1 = [2V_{dc}/3, 0]$
$\boldsymbol{S}_2 = [1 \ 1 \ 0]$	$\boldsymbol{v}_2 = [V_{dc}/3, \sqrt{3}V_{dc}/3]$
$\boldsymbol{S}_3 = [0 \ 1 \ 0]$	$\boldsymbol{v}_3 = [-V_{dc}/3, \sqrt{3}\ V_{dc}/3]$
$\boldsymbol{S}_4 = [0 \ 1 \ 1]$	$\boldsymbol{v}_4 = [-2V_{dc}/3, 0]$
$\boldsymbol{S}_5 = [0 \ 0 \ 1]$	$\boldsymbol{v}_5 = [-V_{dc}/3, -\sqrt{3}V_{dc}/3]$
$\boldsymbol{S}_6 = [1 \ 0 \ 1]$	$\boldsymbol{v}_6 = [V_{dc}/3, -\sqrt{3}V_{dc}/3]$
$\boldsymbol{S}_7 = [1 \ 1 \ 1]$	$\boldsymbol{v}_7 = [0, 0]$

The FCS-MPC based on the current control objective uses a discrete-time model of the load current for the formulation of the control algorithm and deals only with a limited number of possible switching states of the power converter. The discrete-time model can be obtained from the dynamics of the load current. The dynamics of the load current $\boldsymbol{i}_L$ (i_{La}, i_{Lb}, i_{Lc}) can be described by the vector differential equation as below:

$$\boldsymbol{v_i} = R\boldsymbol{i_L} + L\frac{d\boldsymbol{i_L}}{dt} \tag{1}$$

A. Discrete-time Predictive Model

The discrete-time model of the load current dynamics (1) for a sampling time T_S represents the predictive model. It will be used to predict the future value of load currents considering each voltage vector listed in Table II. The discrete-time model can be obtained by an approximation of load current derivative $d\boldsymbol{i}_L/dt$ using the forward Euler discretization method given as:

$$\frac{d\boldsymbol{i_L}}{dt} \approx \frac{\boldsymbol{i_L}(k+1) - \boldsymbol{i_L}(k)}{T_S} \tag{2}$$

After substituting (2) into (1), an expression of predicted future load current at sampling interval k+1 is obtained for each of the seven different voltage vectors $\boldsymbol{v_i}(k)$ in $\alpha\beta$ frame as:

$$\begin{aligned} i_{L\alpha}^{p}(k+1) &= k_1 i_{L\alpha}(k) + k_2 v_{i\alpha}(k) \\ i_{L\beta}^{p}(k+1) &= k_1 i_{L\beta}(k) + k_2 v_{i\beta}(k) \end{aligned} \tag{3}$$

$$\text{where } k_1 = \left(1 - \frac{RT_S}{L}\right),\ k_2 = \frac{T_S}{L}$$

$i_{L\alpha}(k)$, $i_{L\beta}(k)$ and $i^p{}_{L\alpha}(k+1)$, $i^p{}_{L\beta}(k+1)$ denote the measured load currents at instant k and the predicted future load currents at instant k+1, respectively. The coordinate transformations from abc to $\alpha\beta$ are computed using mathematical relation defined by Clarke transformation as

$$\begin{bmatrix} x_\alpha \\ x_\beta \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{3} & -1/\sqrt{3} \end{bmatrix} \begin{bmatrix} x_a \\ x_b \\ x_c \end{bmatrix} \tag{4}$$

where x can be any quantity voltage or current.

The voltages and currents in dq coordinates can be transformed from $\alpha\beta$ components using Park transformation relation given as:

$$\begin{bmatrix} x_d \\ x_q \end{bmatrix} = \begin{bmatrix} \cos\theta^* & \sin\theta^* \\ -\sin\theta^* & \cos\theta^* \end{bmatrix} \begin{bmatrix} x_\alpha \\ x_\beta \end{bmatrix} \tag{5}$$

where θ^* is the reference phase angle for the coordinate conversion. The predicted load currents in rotating dq-frame can also be obtained using forward Euler discretization considering feed-forward terms for the decoupling of the d and q components of the current as given below:

$$\begin{aligned} i_{Ld}^{p}(k+1) &= k_1 i_{Ld}(k) + k_2 \left\{ v_{id}(k) + k_3 i_{Lq}(k) \right\} \\ i_{Lq}^{p}(k+1) &= k_1 i_{Lq}(k) + k_2 \left\{ v_{iq}(k) - k_3 i_{Ld}(k) \right\} \end{aligned} \tag{6}$$

$$\text{where } k_1 = \left(1 - \frac{RT_S}{L}\right),\ k_2 = \frac{T_S}{L},\ k_3 = \omega^* L$$

ω^* is an angular frequency of the current reference and calculated as $\omega^* = d\theta^*/dt$.

B. Adaptive Predictive Model

The predictive model depends on the coefficients k_1, k_2, k_3, for current prediction and computed using the load parameters (R, L). Any change in load parameters disturbs the accuracy of the predictive model, which ultimately affects the performance of the FCS-MPC. The value of a coefficient can be updated to compensate for the effect of change in load parameters. In this paper, an approach to update the coefficient k_1 (adaptive k_1) corresponding to the change in load resistance R is considered by assuming unchanged load inductance L.

According to the phase voltage and line current relationship, the load impedance Z can be calculated as given below:

$$Z = \frac{v_{ph,rms}}{i_{L,rms}} \tag{7}$$

where $v_{ph,rms}$ is the rms values of voltage in any phase and $i_{l,rms}$ is the rms value of the load current. The $v_{ph,rms}$ can be simply obtained using the relation between the phase voltage and a dc supply to an inverter given as:

$$V_{dc} = \frac{2\sqrt{2}v_{ph,rms}}{m} \tag{8}$$

where m is a modulation index.

By substituting the value of $v_{ph,rms}$ from (8) into (7), the value of Z can be updated as follows

$$Z = \frac{mV_{dc}}{2\sqrt{2}i_{L,rms}} \tag{9}$$

The load impedance Z contains resistance (R) and inductive reactance (X_L) as given below

$$Z = \sqrt{R^2 + (X_L)^2},\ \text{where } X_L = \omega L \tag{10}$$

R can be computed by putting the value of Z from (9) to (10) and represented as:

$$R = \sqrt{\left(\frac{mV_{dc}}{2\sqrt{2}i_{L,rms}}\right)^2 - (X_L)^2} \tag{11}$$

R can be updated by (11) according to the change in load current. Further, the coefficient k_1 used for current prediction in (3) and (6) can be automatically updated by putting the value of

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2020 at 05:20:42 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2999267, IEEE Journal of Emerging and Selected Topics in Power Electronics

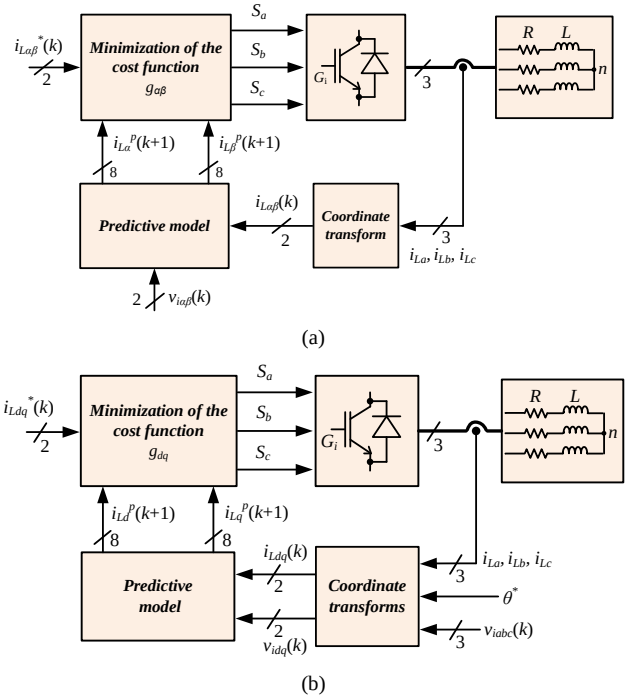


Fig. 2. Schematic diagram of load current control using the FCS-MPC in (a) $\alpha\beta$-frame and, (b) dq-frame.

the updated R from (11) and an expression for adaptive k_1 can be written as follows:

$$k_1 = \left(1 - \frac{T_S}{L}\sqrt{\left(\frac{mV_{dc}}{2\sqrt{2}i_{L,rms}}\right)^2 - \left(X_L\right)^2}\right) \tag{12}$$

Considering any system parameters, $\left(\frac{mV_{dc}}{2\sqrt{2}i_{L,rms}}\right)^2 \gg \left(X_L\right)^2$.

Thus, the expression for adaptive k_1 is approximated as:

$$k_1 \approx \left(1 - \frac{mV_{dc}T_S}{2\sqrt{2}Li_{L,rms}}\right) \tag{13}$$

The values of V_{dc}, T_S, and L are known and $i_{l,rms}$ can be obtained from the reference load current.

C. Cost Function

To select an optimum switching state according to the optimum voltage vector in each sampling and to minimize the current error for the next sampling interval, selection criteria is required to be defined considering control objectives.

The simple selection criteria i.e cost function can be expressed in orthogonal $\alpha\beta$ coordinates as follows:

$$g_{\alpha\beta} = \left|i^*_{L\alpha}\left(k+1\right) - i^p_{L\alpha}\left(k+1\right)\right| + \left|i^*_{L\beta}\left(k+1\right) - i^p_{L\beta}\left(k+1\right)\right| \tag{14}$$

where $i^*_{L\alpha}$ and $i^*_{L\beta}$ are the real and imaginary components of the reference current. The future reference current value required by (14) has to be predicted using the Lagrange extrapolation method discussed in [5]. However, for sufficient small T_S, a simple approximation $i^*_L(k+1) \approx i^*_L(k)$ can be used and no extrapolation is required. The same approximation is considered in this paper.

The cost function for the control of dq coordinate is formulated as:

$$g_{dq} = \left|i^*_{Ld}\left(k+1\right) - i^p_{Ld}\left(k+1\right)\right| + \left|i^*_{Lq}\left(k+1\right) - i^p_{Lq}\left(k+1\right)\right| \tag{15}$$

where i^*_{Ld} and i^*_{Lq} are the reference currents for d and q components respectively. The block diagram of the load current control using the FCS-MPC with stationary $\alpha\beta$ and rotating dq reference frames are depicted in Fig. 2(a) and Fig. 2(b) respectively.

The cost functions expressed in (14) and (15) are based on the primary control objective that is current control. Furthermore, the cost function with a constraint of change in switching state of the inverter is considered to compare the system performance for the FCS-MPC. The inclusion of this constraint reduces the number of commutations of power switches that leads to a reduction in switching frequency. Present switching state $\boldsymbol{S}(k)$ is compared with previously applied switching state $\boldsymbol{S}(k\text{-}1)$ to ensure fewer switching commutations for each phase. The cost function with the constraint is formulated as

$$g = g_{\alpha\beta/dq} + \lambda \cdot p \tag{16}$$

where λ is the weighting factor and p is the number of switching commutations defined as

$$p = \sum_x \left|S_x\left(k\right) - S_x\left(k-1\right)\right|, \quad x = a,b,c \tag{17}$$

III. Model-Based Design of the FCS-MPC

A low-level digital code (VHDL/Verilog) of the desired control system is required to program an FPGA device for the FPGA-based system implementation. The developments in the system implementation facilitate the generation of low-level code for FPGAs using the functionality of a model-based design approach [20], [23]. The XSG, an MBD platform, provides the Xilinx blockset for the MATLAB/Simulink to develop a virtual model of digital circuits specifically for Xilinx FPGAs. Xilinx blockset consists of the fundamental mathematical blocks, a block for the inclusion of low-level VHDL code (Blackbox) and a block for the inclusion of MATLAB code (MCode). During implementation, the Blackbox is used for the analog-to-digital converter (ADC) and digital-to-analog (DAC) interfaces.

The modeling of the FCS-MPC is performed based on the schematic diagram shown in Fig. 3. This is a fully parallel implementation approach and the control algorithm is divided

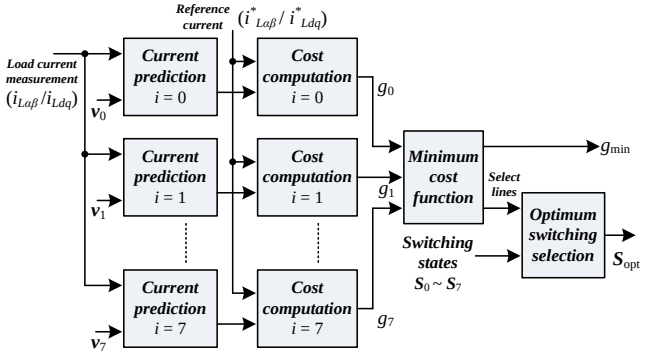


Fig. 3. Block diagram for parallel implementation of the FCS-MPC algorithm using model-based design approach.

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2020 at 05:20:42 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2999267, IEEE Journal of Emerging and Selected Topics in Power Electronics

into three steps: prediction of future load currents, computation of cost functions, and selection of optimum switching state. The cost function without constraint is opted to demonstrate the modeling steps and FPGA resource utilization based on implementation strategies considering equal resource requirements by the constraint during the FCS-MPC implementation with $\alpha\beta$ as well as dq-frames.

The prediction of future load currents is performed firstly using (3) and (6) corresponding to the FCS-MPC in $\alpha\beta$-frame and dq-frame respectively. Further, the cost functions are computed using (14) and (15) for both the frames described in section II. The future current predictions and cost functions are computed on the parallel basis for each inverter voltage vectors ($\boldsymbol{v}_0$–$\boldsymbol{v}_7$) defined for corresponding switching states (S_a, S_b, S_c) as given in Table II.

The modeling for the prediction and cost computation can be performed using only fundamental mathematical blocks based on (3), (14) and (6), (15) for $\alpha\beta$ and dq-frame respectively. However, there is a possibility to perform the same operation using other blocks of the Xilinx blockset. The following strategies are considered for the modeling of the FCS-MPC algorithm for both $\alpha\beta$ and dq frame:

- Fundamental mathematical blocks for the current prediction and cost function computation
- MCode block for the current prediction and cost function computation
- Approximation for the coefficient k_1
- Adaptive k_1

IV. FPGA Implementation Strategy and Resource Utilization

A. Implementation Strategy

The implementation strategy is crucial for the time required for computation and resource utilization. There is a possibility to reduce the required FPGA resources significantly by using an appropriate implementation strategy that may create the possibility of system implementation using smaller FPGAs. The following are the objectives of analytical implementation strategies mentioned in Section III:

- FPGA resource: minimization of resource utilization (LUTs, flip-flops, DSP slices) required for fully parallel implementation.
- Usability: the implementation strategy with minimum resource utilization can be used for semi-parallel/sequential implementation.
- Trade-off: maintaining appropriate control quality and computational speed with reduced resource utilization.

Besides, the optimum strategy can be beneficial to develop various controllers for the FPGA-based system implementation. The implementation strategies are described as follows:

1) Fundamental mathematical blocks

In this strategy, the future current predictions and cost function computations are performed using fundamental mathematical XSG blocks: AddSub, CMult, Mult, Absolute. AddSub block performs the addition and subtraction for two inputs, CMult block performs the function of gain multiplier for a single input, Mult block performs the function of multiplication for two inputs and Absolute block performs the function of absolute value computation for real and imaginary quantities in (14) and (15). The coefficients (k_1, k_2, k_3) used for future current predictions are provided by Xilinx Constant blocks with an appropriate number of bits and binary point settings for fixed-point number representation.

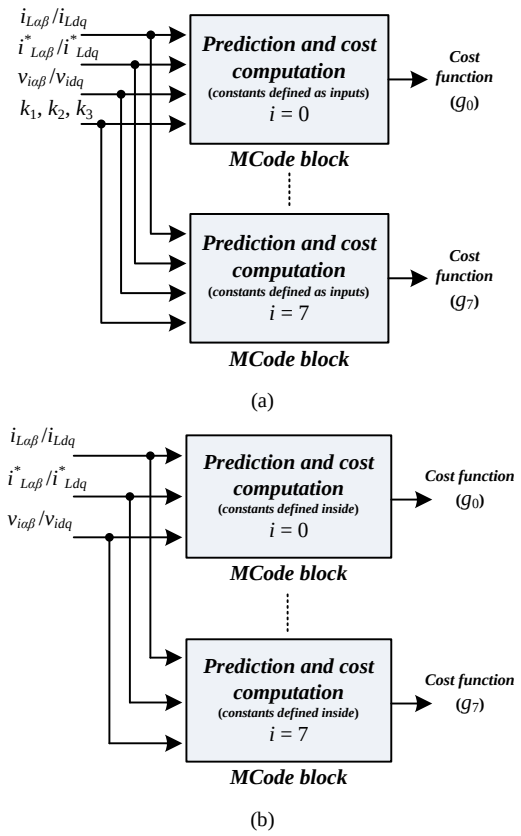


Fig. 4. Development methodologies for the FCS-MPC in XSG using MCode block (a) case-I, (b) case-II.

2) MCode blocks

In this strategy, the future current predictions and cost function computations are performed together through a single MCode block corresponding to each voltage vector using MATLAB code. This strategy is further subdivided into two cases based on the coefficients (k_1, k_2, k_3):

Case-I: coefficients defined as inputs to MCode blocks

Case-II: coefficients defined inside MCode blocks

The development methodologies are presented in Fig. 4 for case-I (Fig. 4(a)) and case-II (Fig. 4(b)). An additional case (case-III) is considered for the resource utilization in the current prediction using adaptive k_1 where other coefficients defined inside the MCode block like case-II.

The MATLAB code for the prediction (Ipred_real and Ipred_img) and cost computation using current errors (x and y) in $\alpha\beta$-frame for case-I and case-II is as follows:

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2020 at 05:20:42 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2999267, IEEE Journal of Emerging and Selected Topics in Power Electronics

MATLAB code: **Case-I**

```
function [cost] = mpc_case-I (Im_real, Im_img,
Iref_real, Iref_img, k1, k2, v_real, v_img)
Ipred_real = k1*Im_real + k2*v_real;
Ipred_img = k1*Im_img + k2*v_img;
x = Iref_real – Ipred_real;
y = Iref_img – Ipred_img;
    if (x<0)
        mod_x = – x;
    else
        mod_x = x;
    end
    if (y<0)
        mod_y = – y;
    else
        mod_y = y;
    end
cost = mod_x + mod_y;
end
```

MATLAB code: **Case-II**

```
function [cost] = mpc_case-II (Im_real, Im_img,
Iref_real, Iref_img, v_real, v_img)
k1 = xfix({xlSigned, 10, 8, xlRound, xlWrap}, 0.95);
k2 = xfix({xlSigned, 13, 13, xlRound, xlWrap},
0.005);
Ipred_real = k1*Im_real + k2*v_real;
Ipred_img = k1*Im_img + k2*v_img;
x = Iref_real – Ipred_real;
y = Iref_img – Ipred_img;
    if (x<0)
        mod_x = – x;
    else
        mod_x = x;
    end
    if (y<0)
        mod_y = – y;
    else
        mod_y = y;
    end
cost = mod_x + mod_y;
end
```

The values of k_1, k_2 are defined inside the MATLAB code in case-II using a command xfix that is used for the fixed-point number representation. Similarly, the MATLAB codes for the case-I and case-II of the FCS-MPC in *dq*-frame are designed using the coefficients k_1, k_2, and k_3 according to the codes designed for the FCS-MPC in $\alpha\beta$-frame mentioned above. An additional case (case-III) is considered for the resource utilization in the current prediction using adaptive k_1 where other coefficients defined inside the MCode block like case-II.

B. Resource Utilization

Every FPGA has a set of programmable logic, I/O, and memory resources. The FPGA resources consist of configurable logic blocks (CLBs) which are the main constituent to implement code on the FPGA. A CLB element comprises of a pair of slices and a slice contains LUTs and flip-flops [30], [31]. According to the family of the FPGA chip, the number of LUTs and flip-flops in a single slice are different. For instance, there are four LUTs and eight flip-flops in a slice considering Zynq-7000 series FPGAs [31]. In addition, there is a major component of the FPGA resources, called DSP slices that are mainly used to implement digital signal processing functions. The main constituents of a DSP slice are signed multipliers, adder/accumulator, arithmetic logic unit (ALU) logic functions and many more [32], [33].

The FPGA resource utilization considering implementation strategies mentioned in the previous section are listed in Table III and IV for the FCS-MPC in $\alpha\beta$-frame and *dq*-frame respectively. To select the optimum design strategy, only two logic indices LUTs and DSP slices need attention because the number of Flip-flops are the same considering $\alpha\beta$-frame as well as *dq*-frames.

1) Coefficient k_1 (k_1 = 0.95, k_1 = 1)

The coefficient k_1 is used in future current prediction using (3) and (6) as mentioned in Section II. The value of k_1 depends on the load parameters (R, L) and sampling time (T_S). The exact value of k_1 is 0.95 based on the parameters (R = 10 Ω, L = 10 mH and T_S = 50 μs) considered for simulation and experimental validations. In the case of k_1 = 0.95, the number of bits 10 and binary points 8 are specified and provides the coefficient value of 0.9492187 that utilizes significant FPGA resources. The approximation (k_1 = 1) can be opted considering the reduction in FPGA resource utilization due to the omission of the required bits and the multiplier used to multiply the exact value of k_1. Consequently, the required bits and the multipliers directly reduce the LUTs and DSP slices, respectively. However, the

TABLE III
FPGA RESOURCE UTILIZATION FOR THE DEVELOPMENT OF THE FCS-MPC IN $\alpha\beta$ - FRAME

Logic utilization indices	*Fundamental Mathematical Blocks*				*MCode Block*					
					Case-I		*Case-II*		*Case-III*	
	k_1 = 0.95	k_1 = 1	*Adaptive* k_1		k_1 = 0.95	k_1 = 1	k_1 = 0.95	k_1 = 1	*Adaptive* k_1	
			m = 1	m = 0.95, 0.9					m = 1	m = 0.95, 0.9
LUTs (53200)	5732 (11%)	4799 (9%)	6546 (12%)	7580 (14%)	4364 (8%)	3804 (7%)	4896 (9%)	3794 (7%)	5488 (10%)	6792 (14%)
Flip-flops (106400)	549 (1%)	549 (1%)	1551 (2%)	1820 (2%)	549 (1%)	549 (1%)	549 (1%)	549 (1%)	1551 (2%)	1820 (2%)
DSP slices (220)	21 (10%)	19 (9%)	26 (12%)	28 (13%)	25 (11%)	19 (9%)	19 (9%)	19 (9%)	27 (12%)	29 (13%)

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2020 at 05:20:42 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2999267, IEEE Journal of Emerging and Selected Topics in Power Electronics

TABLE IV
FPGA RESOURCE UTILIZATION FOR THE DEVELOPMENT OF THE FCS-MPC IN *dq* - FRAME

Logic utilization indices	*Fundamental Mathematical Blocks*				*MCode Block*					
					Case-I		*Case-II*		*Case-III*	
	$k_1 = 0.95$	$k_1 = 1$	*Adaptive* k_1		$k_1 = 0.95$	$k_1 = 1$	$k_1 = 0.95$	$k_1 = 1$	*Adaptive* k_1	
			$m = 1$	$m = 0.95, 0.9$					$m = 1$	$m = 0.95, 0.9$
LUTs (53200)	5017 (9%)	5004 (9%)	5489 (10%)	5784 (11%)	8534 (16%)	8304 (16%)	9967 (19%)	8072 (15%)	9133 (17%)	10646 (20%)
Flip-flops (106400)	804 (1%)	804 (1%)	1806 (2%)	2845 (3%)	804 (1%)	804 (1%)	804 (1%)	804 (1%)	1806 (2%)	2845 (3%)
DSP slices (220)	66 (30%)	62 (28%)	66 (30%)	70 (32%)	86 (39%)	74 (34%)	24 (11%)	24 (11%)	32 (15%)	36 (16%)

performance with $k_1 = 1$ is required to be analyzed through simulation as well as the experiment.

In case-II of MCode block, the equal number of DSP slices are required for both the subcases $k_1 = 0.95$ and 1, however, the required number of LUTs for $k_1 = 0.95$ is slightly higher as required for $k_1 = 1$. In the case of the FPGA implementation using controller development in $\alpha\beta$-frame, both cases (case-I and II) of MCode with $k_1 = 1$ are the optimum solutions, however, considering better controller performance based on current prediction using an exact value of k_1 with the cost of slightly increased LUTs, case-II of MCode block with $k_1 = 0.95$ can be the optimum choice.

On the other hand, for the controller development in *dq*-frame, case-II of MCode block is an optimum strategy due to the least percentage utilization of DSP slices among all the individual resource indices. Although the LUTs requirement has reduced significantly in the case of the development using fundamental mathematical blocks, the required DSP slices have increased approximately 3 times. Furthermore, the FPGA resource utilization for the development of the FCS-MPC in *dq*-frame is even higher as compared to $\alpha\beta$-frame due to the additional coordinate transformation ($\alpha\beta$ to *dq*) for voltage vector including measured current, reference phase angle θ^* generation using CORDIC SINCOS and an extra effort for computation of feed-forward term used for decoupling.

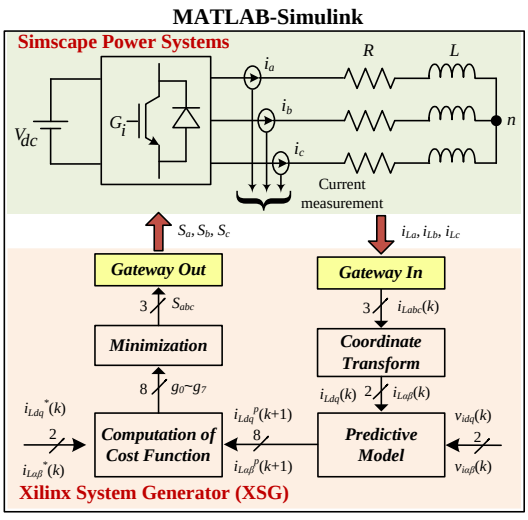


Fig. 5. Block diagram representation of methodology for simulation.

2) Adaptive k_1 (m = 1, 0.95, 0.9)

The adaptive k_1 is used for future current prediction using (3), (6) and (13) as mentioned in Section II. The adaptive k_1 depends on the modulation index (m) for a fixed dc supply. The unity value (m = 1) and non-unity value (m = 0.95, 0.9) are opted for system performance analysis as well as resource utilization.

In the case of the controller development using adaptive k_1, resource utilization is significantly higher than the optimum design cases (MCode block: case-II) in both $\alpha\beta$-frame as well as *dq*-frames. However, considering the updated predictive model corresponding to the load changes, the controller performance needs to be investigated for the controller development using adaptive k_1. Moreover, a trade-off between controller performance and the FPGA resource utilization can be a better choice to decide an appropriate implementation strategy corresponding to the system application.

An optimum implementation strategy corresponding to the minimum resource utilization is further utilized for the prediction and cost computation in both the frames. After the completion of current prediction and cost computation, the minimum cost function g_{min} and an optimum switching state $\boldsymbol{S}_{opt}$ are selected using a pipelining method for all three implementation strategies. The logic for the selection of g_{min} and $\boldsymbol{S}_{opt}$ are developed in XSG using a multiplexer (Mux block) and comparator (Relational block) based on the XSG diagram depicted in [26].

V. RESULTS AND DISCUSSION

Simulation methodology in Fig. 5 is represented using a block diagram to express the development approach of the controller. The power circuit of three-phase VSI is developed in MATLAB/Simulink using simscape toolbox, whereas the controller is developed in the digital environment of XSG using Xilinx toolbox. Due to the non-identical platforms, a signal from simscape to XSG is interfaced using Xilinx "Gateway In" block and a signal from XSG to simscape is interfaced using Xilinx "Gateway Out" block.

The FPGA architecture of the FCS-MPC with the schematic block diagram representation of the experimental setup in Fig. 6 is developed for a complete understanding of the FPGA-based implementation of the system. An adequate latency is provided to blocks used to predict load current and to compute cost

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2020 at 05:20:42 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2999267, IEEE Journal of Emerging and Selected Topics in Power Electronics

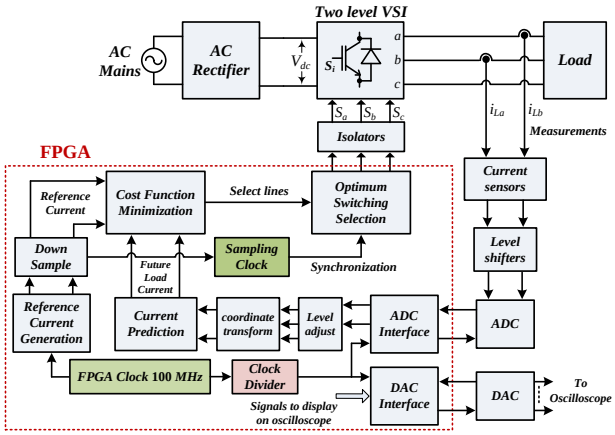


Fig. 6. Schematic diagram of FPGA architecture for experimental setup.

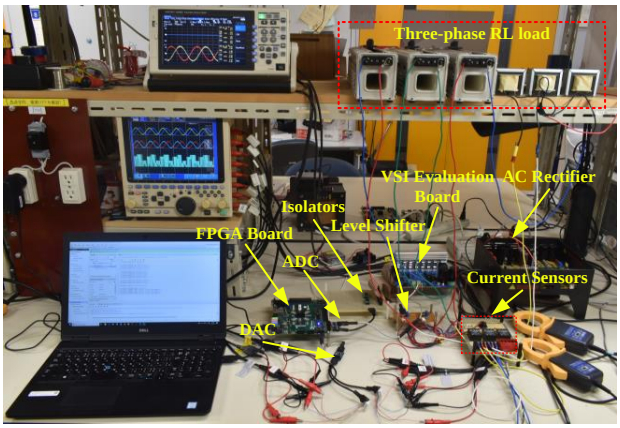


Fig. 7. Laboratory prototype of the experimental setup.

function for proper completion of the computational task by the corresponding block. To ensure the selection of the optimum switching state in the given sampling interval, the sampling clock is generated as a synchronizer. The sampling clock is used as an enabling signal and given to the multiplexer that is used for the selection of optimum switching state. Before sending switching signals directly to the VSI, the isolator circuit is used for electrical isolation of the FPGA board and inverter board considering safety precautions of the FPGA board (Zynq-XC7020). The FPGA code was generated through the modeled controller in XSG and programmed using dedicated software (Xilinx Vivado Design Suite) for the real-time implementation of the VSI system with the FCS-MPC. The laboratory prototype of the experimental setup is shown in Fig. 7. The experimental system components used for the development of laboratory prototype are listed in Table V.

TABLE V
DISCRIPTION OF COMPONENTS SPECIFICATIONS FOR THE EXPERIMENTAL SYSTEM

S.No.	Components	Specifications
1.	Three-phase VSI	STEVAL-IHM023V3, 1 kW
2.	DC supply	THDSHVMTRPFCKIT
3.	Current transducer	LA 25-NP
4.	Op-amp IC for level shifter	LM385N
5.	Isolator IC	ADuM3440
6.	ADC	Digilent Pmod AD1
7.	DAC	Digilent Pmod DA4
8.	FPGA board	Zedboard Zynq Evaluation and Development Kit

TABLE VI
PARAMETERS FOR SYSTEM IMPLEMENTATION

Parameter	Symbol	Value
Supply dc voltage	V_{dc}	145 V
Load resistance	R	10 Ω
Load inductance	L	10 mH
System frequency	f	50 Hz
Sampling frequency	f_S	20 kHz
FPGA clock frequency	f_{CLK}	100 MHz

The performance of the FCS-MPC in $\alpha\beta$ and dq frames was experimentally evaluated and the experimental data are obtained through HIOKI 8855 Memory Hicorder. Further, the data are plotted with the help of MATLAB plotting tool for demonstration and analysis of experimental results. The system parameters considered for the simulation and the experimental validation are listed in Table VI. The validation and analysis of FPGA-based FCS-MPC are performed using the implementation strategy having the optimum resource utilization.

A. *Effect of Predictive Model Coefficient*

The performance of the FCS-MPC in $\alpha\beta$ and dq frames is analyzed for the predictive model with exact (k_1 = 0.95), approximated (k_1 = 1) as well as the adaptive values of coefficient k_1 considering unity (m = 1) and non-unity (m = 0.95, 0.9) values of 'm'. Moreover, the simulation and experiment were carried out considering the cost function without constraint for reference

TABLE VII
PERFORMANCE OF THE FCS-MPC IN $\alpha\beta$ - FRAME

Indices		i_L = 4 A					i_L = 2.5 A				
		k_1 = 0.95	k_1 = 1	adaptive k_1			k_1 = 0.95	k_1 = 1	adaptive k_1		
				m=1	m=0.95	m=0.9			m=1	m=0.95	m=0.9
Simulation	THD (%)	3.58	3.64	3.46	3.66	3.68	5.27	5.37	5.12	5.22	5.36
	Average f_{sw} (Hz)	3637	3663	3677	3640	3713	3147	3057	3123	3023	3053
Experiment	THD (%)	4.9	5.2	4.6	4.7	4.9	8.5	9.1	7.5	7.7	8.1
	Average f_{sw} (Hz)	2556	2489	2445	2430	2533	2325	2150	2350	2378	2393

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2020 at 05:20:42 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2999267, IEEE Journal of Emerging and Selected Topics in Power Electronics

TABLE VIII
PERFORMANCE OF THE FCS-MPC IN *dq* - FRAME

Indices		i_L = 4 A					i_L = 2.5 A				
		k_1 = 0.95	k_1 = 1	adaptive k_1			k_1 = 0.95	k_1 = 1	adaptive k_1		
				m=1	m=0.95	m=0.9			m=1	m=0.95	m=0.9
Simulation	THD (%)	3.85	4.07	3.75	3.79	4.02	6.09	6.54	5.65	5.74	5.87
	Average f_{sw} (Hz)	3650	3607	3713	3693	3727	3027	2967	3127	3113	3027
Experiment	THD (%)	4.8	5.3	4.7	5.0	5.2	8.4	9.2	7.9	8.2	8.5
	Average f_{sw} (Hz)	2586	2524	2465	2438	2457	2365	2120	2410	2388	2347

currents with low load ($\boldsymbol{i}_L$ = 2.5A) and high load ($\boldsymbol{i}_L$ = 4A) conditions. To investigate the quality of the load current and the switching stress on the power devices, the harmonic content in the load current corresponding to the average switching frequency (f_{sw}) is also analyzed considering both the frames. The results obtained during the simulation as well as an experiment are listed in Table VII and VIII for the FCS-MPC in *αβ* and *dq* frames respectively. The harmonic content in the load current during the experiment was measured through HIOKI 3390 power analyzer and current clamp sensor.

1) Coefficient k_1 (k_1 = 0.95, k_1 = 1)

During the simulation, there is a slight difference in percentage THD for the FCS-MPC in *αβ* and *dq* frames corresponding to the coefficient k_1, however, the performance is almost similar during the experiment for both frames. The percentage THD is less in the case of high load with the cost of slightly higher switching stress as compared to low load during the simulation. However, a significantly high current THDs are observed during the experiment for the low load as compared to high load condition. The percentage THD of the load currents in the experiment is higher as compared to the simulation that is because of the simplified model of the VSI and the load used for the simulation study.

2) Adaptive k_1 (m = 1, 0.95, 0.9)

In this case, there is a slight difference in percentage THD for the simulation as well as the experiment in *αβ* and *dq* frames. This difference is more pronounced for low load current as well as non-unity values of '*m*'. In addition, percentage THD for adaptive k_1 with m = 1 is slightly better as compared to m = 0.95 and 0.9. Compared to the other cases of coefficient k_1 (k_1 = 1, 0.95), the controller performance is slightly better for current prediction with adaptive k_1 in both low load as well as high load conditions for unity value of '*m*', however, implementation with adaptive k_1 requires highest FPGA resources amongst the cases of k_1.

B. Dynamic Performance (k_1 = 0.95, 1; *adaptive* k_1: m = 1, 0.95, 0.9)

To investigate the dynamic performance of the FCS-MPC for the three-phase VSI system, the step change in the reference current is considered for current tracking during the transients. The comparative analysis for the system implementation using k_1 = 0.95, 1 and adaptive k_1 (m = 1, 0.95, 0.9) is performed considering simulation as well as experimental results. The tracking performance of the load current is demonstrated for the change in reference current from 2.5A to 4A at instant 0.062s and from 4A to 2.5A at instant 0.14s.

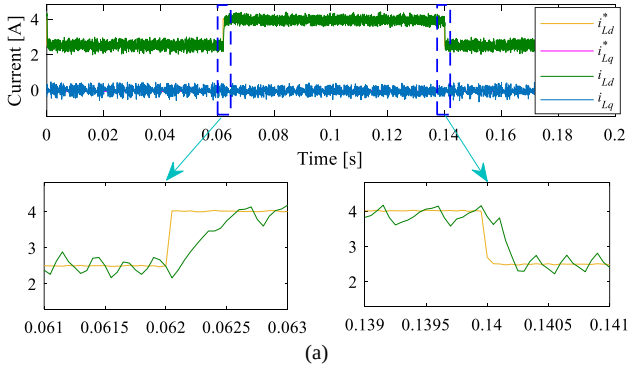
The dynamic performance of FCS-MPC in *αβ* and *dq* frames is represented in Table IX and X considering the settling time during transients for the cases of the coefficient k_1. The settling time for a step change at 0.062s ($\boldsymbol{i}_L$ = 2.5A to 4A) is significantly lower in *dq*-frame as compared to *αβ*-frame. Moreover, the proposed adaptive k_1 case demonstrates the fastest settling time amongst in simulation as well as experiment. The settling time for a step change at 0.14s ($\boldsymbol{i}_L$ = 4A to 2.5A) is slightly better for

TABLE IX
DYNAMIC RESPONSE: SETTLING TIME FOR STEP CHANGE AT 0.062S

Coefficient k_1	FCS-MPC in *αβ*-frame		FCS-MPC in *dq*-frame	
	Sim.	Exp.	Sim.	Exp.
k_1 = 0.95	375 µs	430 µs	350 µs	400 µs
k_1 = 1	400 µs	520 µs	270 µs	350 µs
k_1 adaptive: m = 1	300 µs	370 µs	250 µs	300 µs
k_1 adaptive: m = 0.95	350 µs	410 µs	270 µs	320 µs
k_1 adaptive: m = 0.9	350 µs	430 µs	280 µs	330 µs

TABLE X
DYNAMIC RESPONSE: SETTLING TIME FOR STEP CHANGE AT 0.14S

Coefficient k_1	FCS-MPC in *αβ*-frame		FCS-MPC in *dq*-frame	
	Sim.	Exp.	Sim.	Exp.
k_1 = 0.95	150 µs	250 µs	150 µs	230 µs
k_1 = 1	200 µs	310 µs	170 µs	250 µs
k_1 adaptive: m = 1	150 µs	240 µs	150 µs	220 µs
k_1 adaptive: m = 0.95	200 µs	300 µs	150 µs	240 µs
k_1 adaptive: m = 0.9	250 µs	320 µs	200 µs	300 µs



(a)

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2020 at 05:20:42 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2999267, IEEE Journal of Emerging and Selected Topics in Power Electronics

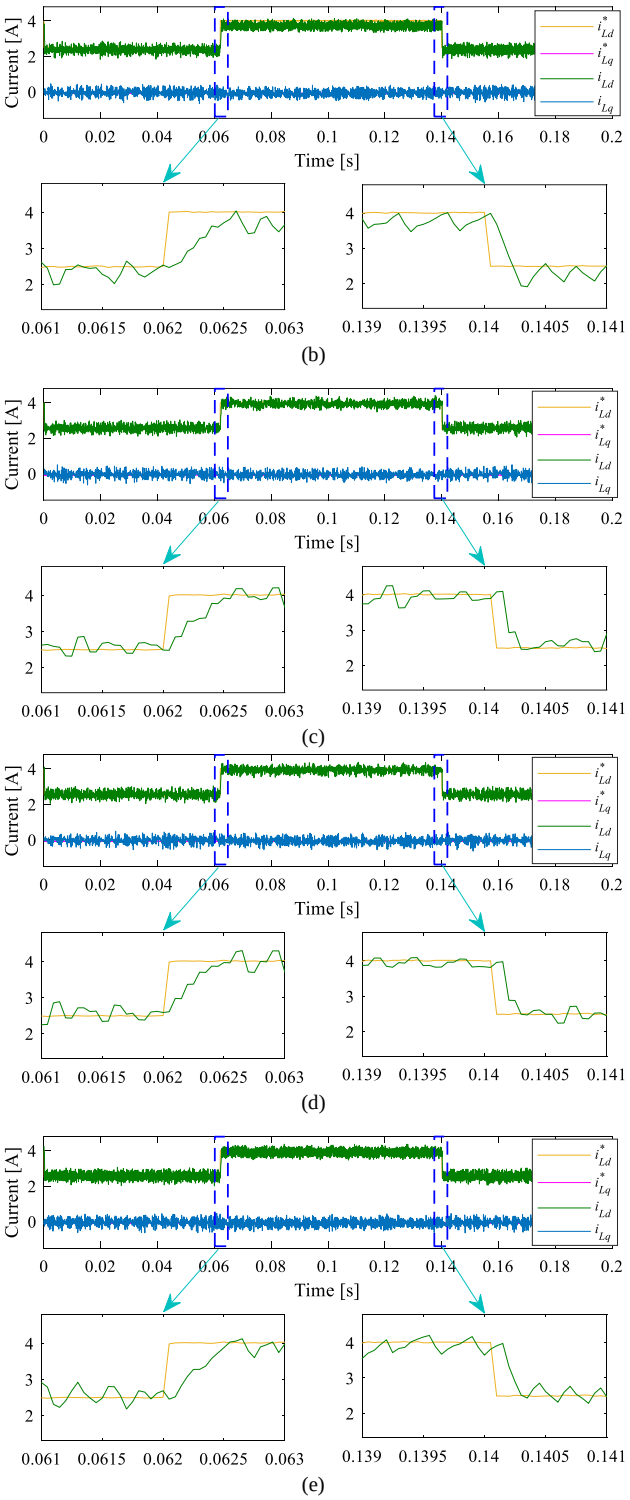


Fig. 8. Experimental results of load current dynamic performance for the FCS-MPC in *dq*-frame (a) coefficient k_1: k_1 = 0.95 (fixed), (b) coefficient k_1: k_1 = 1 (approximated), (c) adaptive k_1: m = 1, (d) adaptive k_1: m = 0.95, and (e) adaptive k_1: m = 0.9.

dq-frame as well as adaptive k_1 case as compared to *αβ*-frame and other cases of k_1.

The current tracking in Fig. 8 are depicted for the controller performance in *dq*-frame for coefficient k_1 = 0.95, 1 as well as adaptive k_1 with m = 1, 0.95, 0.9. In *dq*-frame, the step changes are applied in the *d*-axis component of the reference current, keeping *q*-axis reference current zero. The time axis is enlarged to demonstrate the current tracking during the transient caused by step changes in reference current. The transient response is almost the same for the current prediction using k_1 = 0.95 and 1. In addition, the transient response for adaptive k_1 with m = 1 is almost the same for m = 0.95 and 0.9 and slightly better as compared to k_1 = 0.95 and 1. However, the current tracking performance of adaptive k_1 with m = 1 is best amongst the aforementioned cases.

C. *Steady-state Error*

The steady-state error (SSE) is analyzed through the simulation as well as experiment for load current tracking considering adaptive k_1 (m = 1, 0.95, 0.9) and coefficient k_1 (k_1 = 1, 0.95). The comparative SSE analysis is performed for the $\boldsymbol{i}_L$ = 4A and $\boldsymbol{i}_L$ = 2.5A in Fig. 9 and Fig. 10 respectively using bar graphs for the FCS-MPC in both frames. The comparative analysis demonstrates a similar trend for the $\boldsymbol{i}_L$ = 4A and $\boldsymbol{i}_L$ = 2.5A in simulation as well as experimental results. The SSE

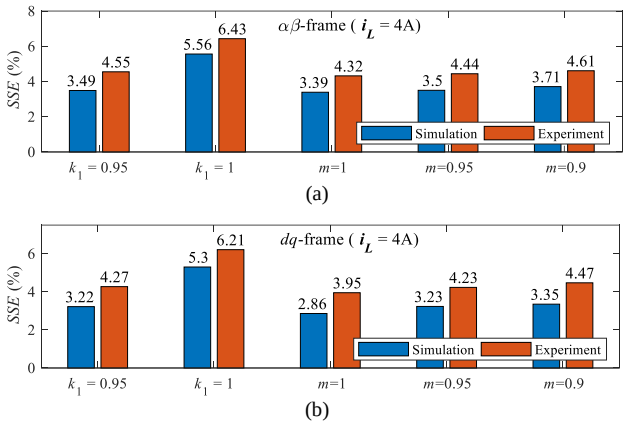


Fig. 9. Steady-state error performance for $\boldsymbol{i}_L$ = 4A with the current prediction using coefficient k_1: k_1 = 0.95 (fixed), k_1 = 1 (approximated) and adaptive k_1: m = 1, 0.95, 0.9 through the simulation as well as experiment for the FCS-MPC in (a) *αβ*-frame, and (b) *dq*-frame.

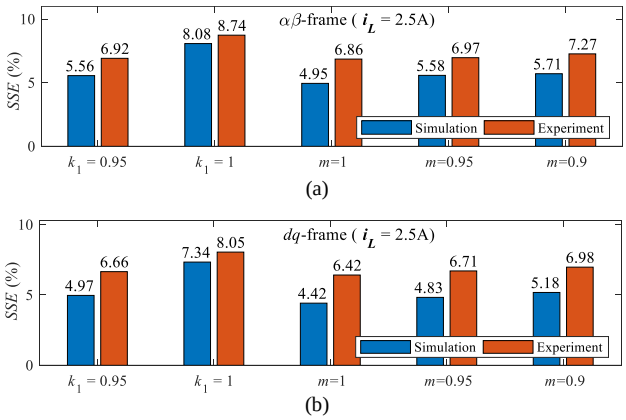


Fig. 10. Steady-state error performance for $\boldsymbol{i}_L$ = 2.5A with the current prediction using coefficient k_1: k_1 = 0.95 (fixed), k_1 = 1 (approximated) and adaptive k_1: m = 1, 0.95, 0.9 through the simulation as well as experiment for the FCS-MPC in (a) *αβ*-frame, and (b) *dq*-frame.

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2020 at 05:20:42 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2999267, IEEE Journal of Emerging and Selected Topics in Power Electronics

performance for adaptive k_1 with m = 1 is better as compared to other cases and it is more pronounced for the low load current $\boldsymbol{i}_L$ = 2.5A. Moreover, the SSE performance in *dq*-frame is better as compared to *αβ*-frame for the considered cases of coefficient k_1 as well as both load current conditions.

D. Switching State Constraint

The system performance is investigated for the cost function with the switching state constraint (16) as discussed in section II during the simulation and experiment. The constraint of change in switching state provides a reduction in the switching frequency corresponding to the variation in weighting factor λ. The system performance with varying λ is examined for average switching frequency f_{sw} as well as percentage THD in the load current considering the case of best system performance (adaptive k_1: m = 1) and poor performance (coefficient k_1 = 1).

1) For condition $\boldsymbol{i}_L$ = 4A (adaptive k_1: m = 1 and k_1 = 1)

The performance of the FCS-MPC in *αβ*, as well as *dq* frames for $\boldsymbol{i}_L$ = 4A, are depicted in Fig. 11 and Fig. 12 for the simulation and experiment results respectively. The average f_{sw} decreases corresponding to an increase in λ, consequently, the harmonic

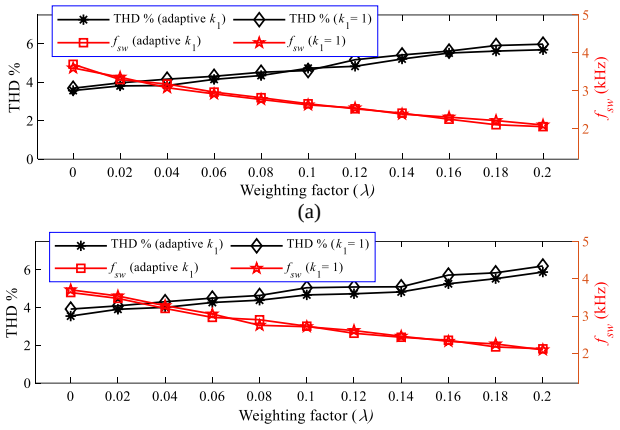


Fig. 11. Average switching frequency (f_{sw}) and percentage THD in load current for $\boldsymbol{i}_L$ = 4A during simulation with the FCS-MPC in (a) *αβ*-frame, and (b) *dq*-frame.

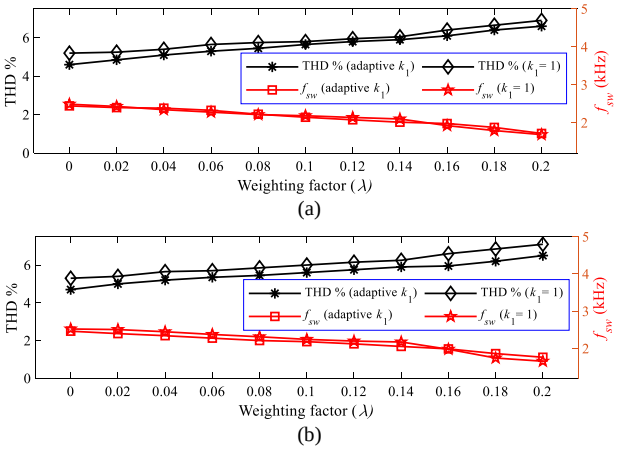


Fig. 12. Average switching frequency (f_{sw}) and percentage THD in load current for $\boldsymbol{i}_L$ = 4A during experiment with the FCS-MPC in (a) *αβ*-frame, and (b) *dq*-frame.

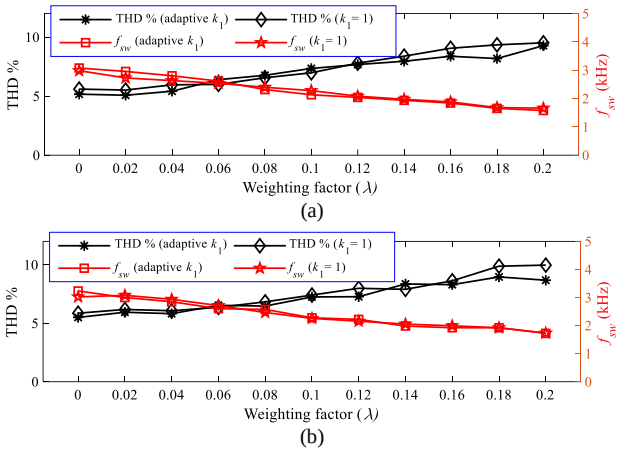


Fig. 13. Average switching frequency (f_{sw}) and percentage THD in load current for $\boldsymbol{i}_L$ = 2.5 A during simulation with the FCS-MPC in (a) *αβ*-frame, and (b) *dq*-frame.

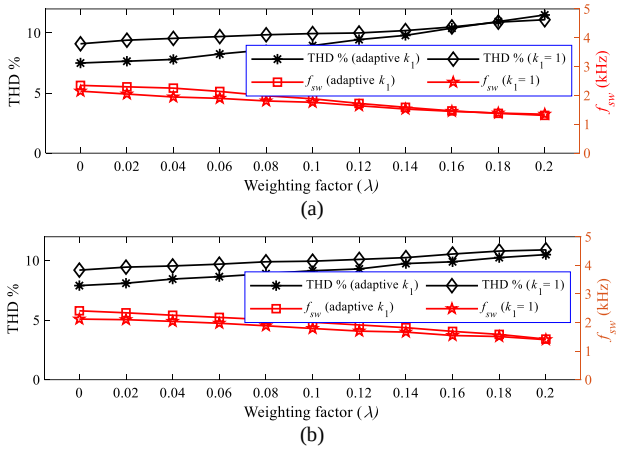


Fig. 14. Average switching frequency (f_{sw}) and percentage THD in load current for $\boldsymbol{i}_L$ = 2.5 A during experiment with the FCS-MPC in (a) *αβ*-frame, and (b) *dq*-frame.

content in the load current increases. A slightly higher percentage THD is observed corresponding to λ for k_1 = 1 as compared to the adaptive k_1 during the simulation as well as the experiment. However, THD is almost the same especially for the value of λ = 0.1 to 0.14 in the simulation as well as the experiment for both the frames.

2) For condition $\boldsymbol{i}_L$ = 2.5A: Adaptive k_1 (m = 1) and k_1 = 1:

Similar to the previous case, the performance in both the frames for $\boldsymbol{i}_L$ = 2.5A is presented in Fig. 13 and Fig. 14 for simulation and experiment respectively corresponding to the same range of λ.

In the case of the simulation, the gap between THD lines for adaptive k_1 and k_1 = 1 is negligible corresponding to variation in λ. However, the gap between THD lines decreases with the increase in λ for both the frames in the case of the experiment. And, similar to the case of $\boldsymbol{i}_L$ = 4A, the THD is almost the same for value λ = 0.1 to 0.18. This trend of THD is authenticated through the average f_{sw} corresponding to change in λ and THD increases with the reduction in average f_{sw}.

The controller performance is significantly changed

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2020 at 05:20:42 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2999267, IEEE Journal of Emerging and Selected Topics in Power Electronics

TABLE XI
FPGA RESOURCE UTILIZATION FOR THE DEVELOPMENT OF THE FCS-MPC WITH SWITCHING CONSTRAINT

Logic utilization indices	FCS-MPC in $\alpha\beta$-frame		FCS-MPC in dq-frame	
	adaptive k_1: $m = 1$	$k_1 = 1$	adaptive k_1: $m = 1$	$k_1 = 1$
LUTs (53200)	5763 (11%)	4074 (8%)	9405 (18%)	8352 (16%)
Flip-flops (106400)	2617 (2%)	1005 (1%)	2262 (2%)	1260 (1%)
DSP slices (220)	38 (17%)	26 (12%)	43 (20%)	31 (14%)

considering the percentage THD in load current for current prediction using fixed, adaptive and approximated value of coefficient k_1 as summarized in Table VII and Table VIII. In addition, the performance is dependent on constraint for switching frequency reduction that ultimately results in higher percentage THD. However, the switching constraint almost compensates for the effect of current prediction using the approximated value of coefficient k_1 for the higher value of λ.

3) FPGA resources

The FPGA resource utilization in the case of the FCS-MPC development with the change in switching state constraint using adaptive k_1 (MCode block: case-III) as well as $k_1 = 1$ (MCode block: case-II) is also listed in Table XI for both the frames. The FPGA resource utilization is minimum for the implementation strategy using MCode block: case-II, $k_1 = 1$. So, a trade-off between the FPGA resource utilization and controller performance can be opted considering the requirement of the system application using an appropriate implementation strategy.

VI. CONCLUSION

This paper presented analytical implementation strategies for the development of FCS-MPC using a fully parallel approach considering conventional current prediction and proposed adaptive current prediction for $\alpha\beta$ and dq coordinates. A significant reduction is achieved (around 65% in DSP Slices) in overall FPGA resource requirement in dq coordinate for system development through MCode block case-II (coefficients defined inside the MATLAB code) of Xilinx blockset. Especially, system development through an approximated value of a coefficient ($k_1 = 1$) provides a further reduction in FPGA resources. Furthermore, resources are optimized using case-II for adaptive current prediction (around 50% reduced DSP Slices) for unity and non-unity modulation index. However, a trade-off between the controller performance and optimum resource utilization is vital for the cost-performance ratio.

The FCS-MPC performance for percentage THD, dynamic behavior, and SSE is investigated for conventional current prediction and proposed adaptive current prediction in $\alpha\beta$ and dq frames using implementation strategy with optimum FPGA resources. The system performance for adaptive k_1 with $m = 1$ is significantly better for all performance indices under high as well as low current operation in both the frames. The performance for coefficient k_1 ($k_1 = 1$) is poor amongst all the case of current predictions and become quite poorer for the low current operation. Furthermore, change in switching state constraint considered for the in-depth analysis lead to an interesting conclusion that percentage THD performance for coefficient $k_1 = 1$ and adaptive k_1 with $m = 1$ become almost similar corresponding to the change in weighting factor under high as well as low current operation. However, for this case, the SSE performance is not investigated and considered as possible future work. This overall system implementation and performance analysis will be vital for achieving a trade-off between controller performance and FPGA resource utilization. The effect of inductive load parameter variation on controller performance is a possible future work considering constraint as well for FPGA-based FCS-MPC implementation using optimum resources.

REFERENCES

[1] S. Kouro, P. Cortes, R. Vargas, U. Ammann and J. Rodriguez, "Model Predictive Control – a Simple and Powerful Method to Control Power Converters," *IEEE Trans. Ind. Electron.*, vol. 56, no. 6, pp. 1826–1838, Jun. 2009.

[2] J. Rodríguez and P. Cortes, Predictive Control of Power Converters and Electrical Drives, 1st ed. New York: *Wiley-IEEE Press*, 2012.

[3] S. Kouro, M. A. Perez, J. Rodriguez, A. M. Llor and H. A. Young, "Model Predictive Control: MPC's Role in the Evolution of Power Electronics," *IEEE Ind. Electron. Mag*., vol. 8, no. 1, pp. 16–31, Mar. 2014.

[4] S. Vazquez, J. I. Leon, L. G. Franquelo, J. Rodriguez, H. A. Young, A. Marquez, and P. Zanchetta, "Model predictive control: A review of its applications in power electronics," *IEEE Ind. Electron. Mag*., vol. 9, no. 4, pp. 8–21, Dec. 2015.

[5] J. Rodríguez, J. Pontt, C. Silva, P. Correa, P. Lezana, P. Cortés, and U. Ammann, "Predictive current control of a voltage source inverter," *IEEE Trans. Ind. Electron.*, vol. 54, no. 1, pp. 495–503, Feb. 2007.

[6] P. Nikhil, K. Sonam, M. Monika, and S. Wagh, "Finite control set model predictive control for two level inverter with fixed switching frequency," in *2018 SICE International Symposium on Control Systems* (SICE ISCS), Mar 2018, pp. 74–81.

[7] F. Donoso, A. Mora, R. Cárdenas, A. Angulo, D. Sáez, and M. Rivera, "Finite-set model-predictive control strategies for a 3L-NPC inverter operating with fixed switching frequency," *IEEE Trans. Ind. Electron.*, vol. 65, pp. 3954–3965, May 2018.

[8] P. Cortes, J. Rodriguez, P. Antoniewicz, and M. Kazmierkowski, "Direct power control of an AFE using predictive control," *IEEE Trans. Power Electron.*, vol. 23, no. 5, pp. 2516–2523, 2008.

[9] J. Rodriguez, M. P. Kazmierkowski, J. R. Espinoza, P. Zanchetta, H. Abu-Rub, H. A. Young, and C. A. Rojas, "State of the art of finite control set model predictive control in power electronics," *IEEE Trans. Ind. Informat.*, vol. 9, no. 2, pp. 1003–1016, May 2013.

[10] C. Qi, X. Chen, P. Tu, and P. Wang, "Cell-by-cell-based finite-control-set model predictive control for a single-phase cascaded H-bridge rectifier," *IEEE Trans. Power Electron.*, vol. 33, no. 2, pp. 1654–1665, Feb. 2018.

[11] M. Rivera, C. A. Rojas, J. Rodriguez, P. Wheeler, B. Wu, and J. R. Espinoza, "Predictive current control with input filter resonance mitigation for a direct matrix converter," *IEEE Trans. Power Electron.*, vol. 26, no. 10, pp. 2794–2803, Oct. 2011.

[12] L. Wang, H.Dan, Y. Zhao, Q. Zhu, T. Peng, Y. Sun and P. Wheeler "A finite control set model predictive control method for matrix converter with zero common-mode voltage," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 6, no. 1, pp. 327–338, Mar. 2018.

[13] P. Cortes, G. Ortiz, J. Yuz, J. Rodríguez, S. Vazquez, and L. Franquelo, "Model predictive control of an inverter with output LC filter for UPS applications," *IEEE Trans. Ind. Electron.*, vol. 56, no. 6, pp. 1875–1883, 2009.

[14] S. Vazquez, J. Rodriguez, M. Rivera, L. G. Franquelo and M. Norambuena, "Model Predictive Control for Power Converters and Drives: Advances and Trends," *IEEE Trans. Ind. Electron.*, vol. 64, no. 2, pp. 935–947, Feb. 2017.

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2020 at 05:20:42 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2999267, IEEE Journal of Emerging and Selected Topics in Power Electronics

[15] P. Cortes, M. P. Kazmierkowski, R. M. Kennel, D. E. Quevedo, and J. Rodriguez, "Predictive control in power electronics and drives," *IEEE Trans. Ind. Electron.*, vol. 55, no. 12, pp. 4312–4324, Dec. 2008.
[16] P. Cortes, J. Rodriguez, C. Silva, and A. Flores, "Delay compensation in model predictive current control of a three-phase inverter," *IEEE Trans. Ind. Electron.*, vol. 59, pp. 1323-1325, 2012.
[17] Y. Yang, H. Wen, and D. Li, "A fast and fixed switching frequency model predictive control with delay compensation for three-phase inverters," *IEEE Access*, 5:17904–17913, 2017.
[18] O. Gulbudak, E. Santi, "FPGA-based model predictive controller for direct matrix converter," *IEEE Trans. Ind. Electron.*, vol. 63, no. 7, pp. 4560–4570, Jul. 2016.
[19] Z. Zhang, F. Wang, T. Sun, J. Rodríguez, guez, and R. Kennel, "FPGA-Based Experimental Investigation of a Quasi-Centralized Model Predictive Control for Back-to-Back Converters," *IEEE Trans. Power Electron.*, vol. 31, pp. 662-674, Jan. 2016.
[20] Z. Ma, S. Saeidi, and R. Kennel, "FPGA implementation of model predictive control with constant switching frequency for PMSM drives," *IEEE Trans. Ind. Informat.*, vol. 10, no. 4, pp. 2055–2063, Nov. 2014.
[21] S. Wendel, A. Dietz, and R. Kennel, "FPGA based finite-set model predictive current control for small PMSM drives with efficient resource streaming," in *2017 IEEE International Symposium on Predictive Control of Electrical Drives and Power Electronics* (PRECEDE), pp. 66–71, Sept. 2017.
[22] B. Stellato, T. Geyer, and P. J. Goulart, "High-Speed Finite Control Set Model Predictive Control for Power Electronics," *IEEE Trans. Power Electron.*, vol. 32, pp. 4007-4020, 2017.
[23] V. K. Singh, R. Tripathi and T. Hanamoto, "FPGA-Based Implementation of Finite Set-MPC for a VSI System using XSG-Based Modeling," *Energies*, vol. 13, no. 1, 260, January 2020.
[24] T. J. Vyncke, S. Thielemans, and J. A. Melkebeek, "Finite-set model based predictive control for flying-capacitor converters: Cost function design and efficient FPGA implementation," *IEEE Trans. Ind. Informat.*, vol. 9, no. 2, pp. 1113–1121, May 2013.
[25] P. M. Sanchez, O. Machado, E. J. B. Peña, F. J. Rodriguez, and F. J. Meca, "FPGA-based implementation of a predictive current controller for power converters," *IEEE Trans. Ind. Informat.*, vol. 9, no. 3, pp. 1312-1321, Aug. 2013.
[26] V. K. Singh, R. N. Tripathi and T. Hanamoto, "Xilinx System Generator Based Modeling of Finite State MPC", *9th International Power Electronics Conference* (IPEC 2018-ECCE Asia), May 20-24, 2018.
[27] H. A. Young, M. A. Perez, J. Rodriguez, and H. Abu-Rub, "Assessing finite-control-set model predictive control: A comparison with a linear current controller in two-level voltage source inverters," *IEEE Ind. Electron. Mag.*, vol. 8, no. 1, pp. 44–52, Mar. 2014.
[28] C. Xia, T. Liu, T. Shi and Z. Song, "A Simplified Finite-Control-Set Model-Predictive Control for Power Converters," *IEEE Trans. Ind. Informat.*, vol. 10, no. 2, pp. 991–1002, May. 2014.
[29] R. Vargas, P. Cortes, U. Ammann, J. Rodriguez, and J. Pontt, "Predictive control of a three-phase neutral point-clamped inverter," *IEEE Trans. Ind. Electron.*, vol. 54, no. 5, pp. 2697–2705, Oct. 2007.
[30] E. Monmasson, L. Idkhajine, M. W. Naouar, "FPGA-based controllers", *IEEE Ind. Electron. Mag.*, vol. 5, no. 1, pp. 14–26, April 2011.
[31] Zynq-7000 SoC: *Technical Reference Manual*, July 1, 2018.
[32] Zynq-7000 SoC: *Data Sheet Overview*, July 2, 2018.
[33] 7 Series DSP48E1 Slice: *User Guide*, March 27, 2018.



Vijay Kumar Singh received his B.Tech. degree from the Department of Electronics and Communication Engineering, United College of Engineering and Research, Allahabad (under UPTU Lucknow), Uttar Pradesh, India, in 2008 and M. Tech. degree (Power Electronics and ASIC Design) from the Department of Electrical Engineering, Motilal Nehru National Institute of Technology, Allahabad, Uttar Pradesh, India in 2012. He completed Ph.D. degree course from the Department of Life Science and Systems Engineering, Kyushu Institute of Technology, Kitakyushu, Fukuoka, Japan in 2020. His research interests include model predictive control, design, modeling, and development of control techniques for power electronic converters.



Ravi Nath Tripathi received his B.Tech (Electrical and Electronics Engineering) degree in 2011 from Graphic Era Institute of Technology (GEIT), India and M.Tech (Power Systems) degree in 2013 from Delhi Technological University (formerly Delhi College of Engineering), Delhi, India. He completed doctoral degree course in 2017 from Kyushu Institute of Technology. He has worked as Post-Doctoral Researcher on NEDO project from 2017-2020 and currently working as a researcher in Next Generation Power Electronics Research Center, Kyushu Institute of Technology, Japan. His research interest is design and development of Gate Drive Control for power semiconductor devices, design and development of controller for power electronic converters, Grid-integration of Solar PV system and motor drives.



Tsuyoshi Hanamoto was born in 1961. He received an M.S degree from Kyushu Institute of Technology, Japan in 1986 and then joined Kobe Works of Kobe Steel, Ltd. In 1990, he joined the Center for Cooperative Research of Kyushu Institute of Technology. From 1997 to 2000 he was with Department of Electrical Engineering and since April 2000, he has been with the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, where he is presently a Professor. His research interests include motor control, power electronics, and FPGA-based system development.

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2020 at 05:20:42 UTC from IEEE Xplore. Restrictions apply.



Taylor & Francis
Taylor & Francis Group

Transactions of the IMF
The International Journal of Surface Engineering and Coatings

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ytim20

Modelling and optimisation of magnetic abrasive finishing process based on a non-orthogonal array with ANN-GA approach

S. Ahmad , R. M. Singari & R.S. Mishra

To cite this article: S. Ahmad , R. M. Singari & R.S. Mishra (2020) Modelling and optimisation of magnetic abrasive finishing process based on a non-orthogonal array with ANN-GA approach, Transactions of the IMF, 98:4, 186-198, DOI: 10.1080/00202967.2020.1776966

To link to this article: https://doi.org/10.1080/00202967.2020.1776966

Published online: 17 Jul 2020.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ytim20

TRANSACTIONS OF THE IMF
2020, VOL. 98, NO. 4, 186–198
https://doi.org/10.1080/00202967.2020.1776966

Check for updates

Modelling and optimisation of magnetic abrasive finishing process based on a non-orthogonal array with ANN-GA approach

S. Ahmad, R. M. Singari and R.S. Mishra

Department of Mechanical, Production & Industrial, Automobile Engineering, Delhi Technological University, New Delhi, India

ABSTRACT

Magnetic abrasive finishing (MAF) is an advanced precise finishing method that achieves micro-level to nano-level surface roughness. In industries, MAF is highly recommended where zero or negligible post-process surface defects are an obligatory requirement. In the same context, process optimisation is essential for making it commercially viable. This study presents an artificial neural network and genetic algorithm (ANN-GA), a robust modelling and optimisation tool (applicable to any sort of data set orthogonal array design or non-orthogonal array design) that is applied to scrutinise and improve the performance of the magnetic abrasive finishing of stainless steel SS302. In addition, the results from ANN-GA modelling and optimisation have been compared with conclusions drawn from conventionally used Taguchi-ANOVA analysis. An L_{27} non-orthogonal array design has been opted for as per machining set-up restriction. Abrasive size, voltage, machining gap, and rotational speed were the design variables considered in the present research work. It was found that the parametric design used in this study provides a straightforward, methodical, and proficient method of modelling and optimisation of change of surface roughness or finishing behaviour during the MAF process. Modelling and optimisation done with ANN-GA show a maximum value of $(\Delta R_a)_{max}$ equal to 0.256 µm, which is 7% better than the result obtained from Taguchi-ANOVA analysis.

ARTICLE HISTORY
Received 13 February 2020
Accepted 22 April 2020

KEYWORDS
Magnetic abrasive finishing; non-orthogonal array; artificial neural network; genetic algorithm; ANOVA

1. Introduction

Expeditious advancements of latest materials with properties like high hot hardness, toughness, fragility, strength-weight ratio, etc. are becoming worthy of attention. These significant properties of such advanced materials have made them more demandable in various industrial and research applications. The processing of these materials by traditional manufacturing processes always has been a challenge. Also, the quality of surface finish always has a consequential effect on the functionality of any mechanical component.[1,2] For example, frictional power losses, wear, tear, and component fatigue life have a great dependency on surface quality and surface integrity.[3]

In recent times, a few non-traditional finishing and polishing processes have evolved. These processes are capable enough to process advanced material such as NiMoC alloys, silicon wafers, carbide, and CBN cutting tools, etc. as well, as they can obtain micro to a nano-level range of surface finish.[4–6] Abrasive flow machining (AFM), chemo-mechanical polishing (CMP), magnetic float polishing (MFP), magneto-rheological finishing (MRF), and magnetic abrasive finishing (MAF), etc., are some such non-traditional finishing and polishing processes.[3,7–10]

In the MAF process, application of gentle forces by using a flexible magnetic abrasives brush (FMAB) for removing surface unevenness has been a really innovative idea. During MAF, the magnetic forces allow the FMAB to shear off the material from the surface in the form of micro-chips.[11–15] Many authors have studied various control factors, and optimisation techniques to accomplish the best surface finish on different advanced materials using the MAF manufacturing processes. However, these optimisation techniques lack intelligent iterative optimisation methodology for better performance of the process.

Yamaguchi *et al.*[6] used diamond paste for the finishing of an uncoated carbide tool (UCT) surface using MAF to get better tool wear characteristics. Finishing time, pole tip feed, spindle speed, machining gap, steel grits mean-diameter, diamond abrasive, lubricant, were chosen as control factors. It was reported that roughness less than 25 nm could be achieved. Tool life of MAF processed UCT had been doubled compared to UCT not processed with MAF, when it was used in turning of titanium (Grade 5) rods. It is vital for coated tool finishing to use different sizes of the abrasive particle simultaneously adjusting magnetic force for desired finishing and less material removal.[16] Amineh *et al.*[1] utilised MAF to remove the recast layer formed by wire electro-discharge machining (WEDM). The MAF was able to shear off the recast layer generated during WEDM without any micro-cracking. It was found that increasing abrasive particles causes higher recast layer removal, and improved the surface roughness. In another study, Amnieh *et al.*[17] examined the effectiveness of finishing parameters such as tool rotational speed, abrasive mesh number, and finishing time. Jiang Guo *et al.*[2] found that the MAF was an effective finishing operation for a flat workpiece of polyamide 12 (PA12) 3-D printed parts under constant magnetic field. The process leaves no visible chemical change within the 10 µm subsurface level. Jiang Guo *et al.*[18] investigated the tribological performance of RSA-905 after finishing the surface by magnetic field-assisted finishing (MFAF). Magnetic abrasive (SiC, Al_2O_3, Carbonyl Iron Powder) was used for finishing and achieved the desired finish without any post-process subsurface defect. Tribological performance was improved, and

CONTACT R. M. Singari ranganath@dce.ac.in Department of Mechanical, Production & Industrial, Automobile Engineering, Delhi Technological University, Shahbad Daulatpur, Bawana Road, New Delhi 110042, India

© 2020 Institute of Materials Finishing Published by Taylor & Francis on behalf of the Institute

surface residual stresses were released. However, a small decrement in hardness was noticed. Yuewu Gao *et al.*[19] studied the influence of various process variables on the polish quality and compared the surface roughness of Cu-27400 and SS316 flat workpiece. An Al_2O_3 based atomised magnetic abrasive tool was selected for processing under the control of a permanent magnet (Nd-Fe-B). Shadab Ahmad *et al.*[20] examined the consequence of sintered magnetic abrasive (SMA) on the MAF process. The process parameters considered were voltage, machining gap, and rotational speed. There were three experimental arrays designed according to the L_9 orthogonal array. The Taguchi-ANOVA method was used to find the optimised surface roughness change and corresponding values of the parameter with their quantitative effect. The most influential parameter was voltage followed by machining gap and rotational speed, respectively and it was found that the abrasive size has a significant role in the MAF process.[20,21]

The application of different optimisation methods like Taguchi, Response Surface, Fuzzy Logic, Genetic Algorithm, Moth Flame Algorithm, and Particle Swarm Algorithm for the effective utilisation of the MAF process, is increasing.[22,23] The influence of voltage, machining gap, abrasive size, and rotational speed on surface roughness during the MAF process was studied by employing various methods by many researchers. Table 1 shows the finishing parameters and optimisation techniques employed in the research.

The above literature survey shows that many researchers have analysed the effects of process parameters and optimised the process with different techniques. In addition, very few researchers have employed an iterative optimisation method for more effective utilisation of the MAF process. Artificial neural networks (ANNs) have solved various parametric based problems in manufacturing processes, like pattern finding and learning in a set of data, developing the ability of pattern classification, forecasting, clustering, or modelling.[31–33] Genetic algorithms (GAs) can perform a global search, optimisation, and generalisation.[34,35] The neural network solves the problems where modelling by mapping input patterns to output results is important. However, to perform computational optimisation, tools like GAs are used and need a single stature of merit as a fitness function.

This current research study explains the development and application of an artificial neural network and genetic algorithm (ANN-GA) as a modelling and optimisation tool that is used to improve the performance of the magnetic abrasive finishing of SS302. In addition, the results from ANN-GA modelling and optimisation have been compared with conclusions drawn from conventionally used Taguchi-ANOVA analysis. The Taguchi-ANOVA method is used to analyse the data for comparison.

2. Magnetic abrasive finishing (MAF)

Magnetic Abrasive Finishing (MAF) is an advanced finishing operation to produce an efficient and excellent surface finish quality on a cylindrical surface, flat surface, and externally and internally in tubes.[36,37] This process has the flexibility to work without the restriction of workpiece material characteristics (physical and mechanical properties). The workpiece material could be electrically conductive or non-conductive, magnetic or non-magnetic.[5,16,36] Thus, there is no restriction if material is made of ferrous or non-ferrous. Conventionally

Table 1. Finishing parameters and optimisation techniques in previous studies.

S. No.	Authors	Workpiece/ Abrasive	Type of Magnetic Source	Voltage or current supply	Machining Gap /mm	Rotational Speed /rpm	Abrasive Size	Surface Roughness Range or effect	Remarks
1.	T. C. Kanish *et al.*[22]	SS316L/ SiC	–	15–20 V	1.5-2	270–540	400–1200 mesh	35–50%	Taguchi optimisation and fuzzy modelling used to predict SR
2.	Prateek Kala *et al.*[24]	Copper/Alumina	Permanent	-	1.5–2.5	200–400	800 mesh	–	Taguchi optimisation and analysis of finishing force and torque involved in DDMAF done
3.	K. B. Judal *et al.*[25]	AISI-304/ SiC	DC Supply	0.5-2.5 A	1–3	150–1200	10–180 µm	(0.28 µm, 0.32 µm) – 0.039 µm	RSM optimisation techniques were used. Rotational speed and electrolytic current lead to significant improvement in MRR and SR.
4.	Yi-Hsun Lee *et al.*[26]	SUS304/ SiC	Permanent	-	1–2	100–1000	1000–8000 mesh	0.13-0.03 µm	Taguchi optimisation done and 77% improvement found in SR
5.	Amer A Moosa[27]	Brass/Quartz	DC supply	1.5–3.5 A	1–2	175–525	150 µm	–	ANFIS used for mapping and optimisation of experimental results
6.	Prateek Kala *et al.*[28]	Copper Alloy C70600/ Al_2O_3	DC supply	70–100 V	2	112–224	600–1200 mesh	62% (MAP), 76% (UAMAP)	Taguchi method applied for process optimisation, and this study concludes that UAMAP is an improved technique over MAP.
7.	A. Cheng Wang *et al.*[29]	SKD11/ AL_2O_3, SiC	-	1–3 A	–	700–1300	4000–8000 mesh	0.029 µm	Claimed that MFGA is better than MAF, because of gel-based MA. Optimisation of the process done by the Taguchi method.
8.	Nitesh Sihag *et al.*[30]	Tungsten Alloy/ Al_2O_3	Permanent	–	2	150	1200 mesh	0.34–17 µm	86% improvement in SR with chemical oxidation RSM and ANOVA was used to design the experiments and analyse the results

SR (Surface Roughness); MRR (Material Removal Rate); DDMAF (Double Disk Magnetic Abrasive Finishing); RSM (Response Surface Methodology); ANFIS (Adaptive Neuro-Fuzzy Inference System); MAP (Magnetic Abrasive Polishing); UAMAP (Ultrasonic Assisted Magnetic Abrasive Polishing); MFGA (Magnetic Finishing with Gel Abrasives); MA (Magnetic Abrasives); MAF (Magnetic Abrasive Finishing); ANOVA (Analysis Of Variance).

in MAF, magnetic abrasives particles (MAPs) composed of ferromagnetic material with abrasive particles are employed as cutting tools to shear-off the material from the target surface. But this material removal is greatly affected by finishing pressure, which works on the influence of a magnetic field across the machining gap.[38] The schematics of the MAF process are illustrated in Figure 1 representing the equipotential lines across which magnetic abrasive is arranged on the work surface to remove material.[20,38] The rotational motion causes the cutting force parallel to the targeted surface, and abrasive penetration inside the workpiece is due to the normal force generated by the magnetic field intensity.[39] MAPs become arranged with each other magnetically along a line of magnetic force.[9] A Fishbone diagram representing all the factors that affect the MAF process is shown in Figure 2.

3. Materials and method

3.1 Material and tools

Stainless steel grade 202 was the workpiece material for the study of this particular finishing operation. Sintered magnetic abrasive (SMA) was working as many tiny cutting tools with random cutting edges in simultaneous action. These abrasives performed shearing operation to remove the material from the targeted surface in the form of microchips.[40] The constituent of SMA was alumina powder, and cast iron powder sintered together in 40/60 by weight ratio, its preparation techniques have been discussed in the literature.[20,21] An electromagnet (32 mm pole diameter) mounted on a lathe tool post connected to a DC Supply produces a static magnetic field, as shown in Figure 3. Various independent control factors influence the performance of the MAF process, including abrasive size, voltage, machining gap, and rotational speed.

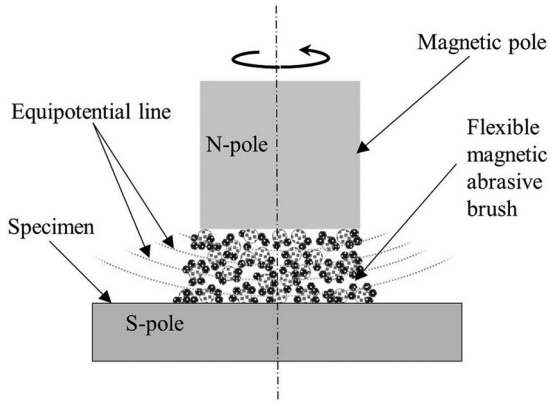


Figure 1. Schematic of the MAF process.

3.2 Experimental detail

The Taguchi technique is a prevailing tool for designing, arrangement, and collection of experimental data in a controlled way. It also tells us the effects of process variables over some particular variables for the design of efficient systems, and simultaneously, it reveals the effects of uncontrolled factors.[41,42]

The L_9 orthogonal array for three levels and three factors was used by Ahmad *et al.*[20]; each row of each array (three different arrays set for three different abrasives sizes, for investigation of a good range of abrasive size) represents the trial situation with a combination of factors. Columns were corresponding to each control parameter or process variable in the array, and there were nine combinations, as it is in the L_9 orthogonal array. The response parameter was Change in Surface Roughness (ΔR_a). This technique has been effectively utilised by numerous researchers to understand the finishing impacts during the MAF process.[4,20,21] Nevertheless, the satisfactory decision of the researched parameters (control factors) is fundamental.

The observations in this examination were led according to a non-orthogonal array[43,44] as per machining set-up restrictions. The determination of the non-orthogonal array depends on relying on the prerequisite that did not fit by the orthogonal array. Table 2 shows the control factors, their units, definition, and levels. It is notable that the control factors level is non-uniform for abrasive size because of sieve size restriction. The experimental data[20] presented in Table 3 have been used to develop the ANN model. The following observations were set in a fashion using an L_{27} non-orthogonal array with four inputs and one output. The input parameters

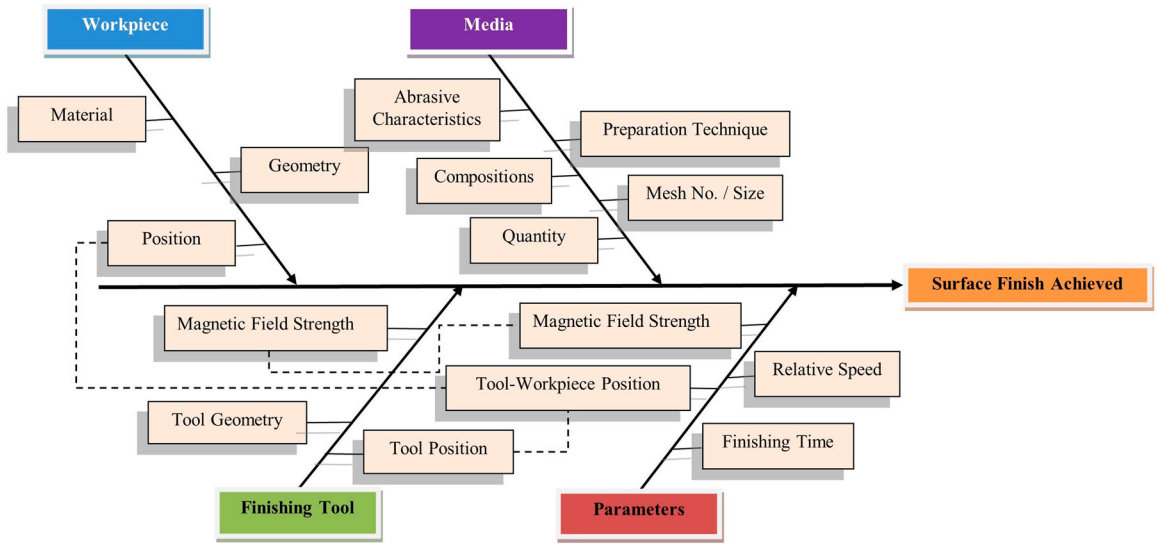


Figure 2. Fishbone diagram for MAF process.

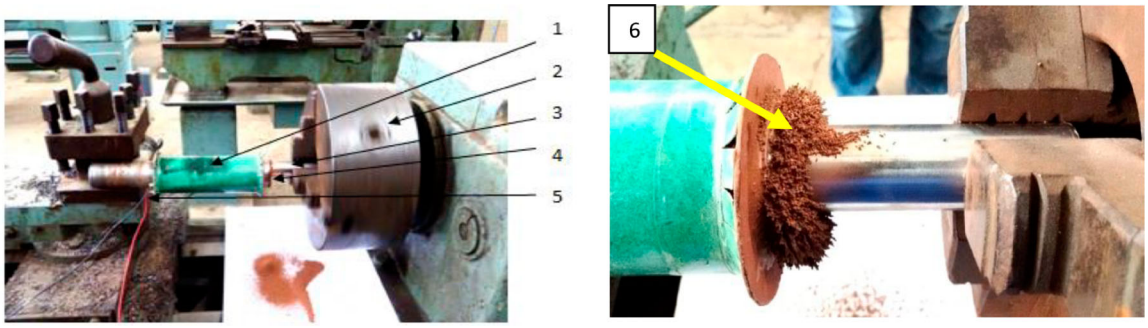


Figure 3. Images of the MAF setup: (1) Electromagnet (2) Chuck, (3) Workpiece, (4) Flexible Magnetic Abrasive Brush (FMAB), and (5) D.C. Power supply (6) FMAB in a closer view.

Table 2. Control factors, units, definition, and levels.

Control factor	Unit	Definition	Level 1	Level 2	Level 3
Abrasive Size (*d*)	µm	The average size of abrasive particles	90	150	300
Voltage (*V*)	Volts	Voltage applied to electromagnet to alter magnetic intensity	6	12	18
Machining Gap (*x*)	mm	The gap between the electromagnetic tool and workpiece	1	1.5	2
Rotational speed (*S*)	rpm	The relative rotational motion between tool and workpiece	60	90	120

Time 20 min for each run.

were abrasive size, voltage, machining gap, and rotational speed. The Change in Surface Roughness (ΔR_a) was an output/response parameter, measured in microns (µm).

In the Taguchi method, the signal-to-noise (S/N) ratio response has been formed corresponding to each observation, and 'larger-the-better' quality characteristic was targeted in this assessment, as more difference in surface roughness required. The control factors and the process response were taken as inputs to calculate the S/N ratio. Moreover, ANOVA analysis was carried out to scrutinise which controlling factor was statistically significant, so the best combination of the control factors, i.e. optimal combination, can be forecasted.[42] The S/N ratio (larger the better quality) for the change of surface roughness given by the Taguchi method is given below:

$$S/N = -10\log\frac{1}{n}\left(\frac{1}{y_1^2}+\frac{1}{y_2^2}+\cdots+\frac{1}{y_n^2}\right)$$

where $y_1, y_2, \ldots, y_n$ are the responses of the change of surface roughness and n is the number of observations. For each level, the average of selected characteristics of the factor has been shown in Table 3 (see *S/N* ratio). The response, *S/N* ratio, merges the reiterations, and the impact of noise levels into one whole coherent information. ANOVA for the *S/N* ratio was carried out for the classification of the significant factor.[45] The mean-response graphs were plotted, whereas the percentage contribution of factors was resolved by ANOVA.

4. Results and discussion

The prime motive of this study was to calculate the maximum Change in Surface Roughness $(\Delta R_a)_{max}$ and optimal values of factors correspondingly that have the most significant influence. Experimental data, which connect the influence of abrasive size, voltage, machining gap, and rotational speed, are shown in Table 3. These factors have a vital impact on the MAF process.

4.1 Analysis of variance (ANOVA)

For the computation of the process quality and its characteristics, the experimental data has been converted into S/N ratio. Based on S/N ratio, a ranking of control factors have been shown in Table 4. It is clear from Table 4 that the prevailing factor that affects the ΔR_a is voltage, and then machining gap, followed by abrasive size and rotational speed, respectively. Also, the variations in S/N ratio are shown in the main effects plot (Figure 4) where the most optimum result by Taguchi analysis is observed as abrasive size, voltage, machining gap, and rotational speed at 150 µm, 18 V, 1 mm, and 90 rpm respectively.

The analysis of the experimental data was carried out using ANOVA. The ANOVA analyses of the ΔR_a for all the analysed factors are shown in Table 5. The factors, having *P*-value less than 0.05 show a statistically considerable effect on the process performance. Also, Table 5 shows the percentage contribution of all factors in the overall result.

By scrutinising the results from Table 5, it can be noticed that voltage has the biggest influence on ΔR_a (% contribution = 41.60%), so much so, that this factor should be the centre of attention. Besides voltage, the change in surface roughness is also affected by machining gap (% contribution = 27.98%), abrasive size (% contribution = 10.95%), and rotational speed (% contribution = 9.39%).

The calculated S/N ratio was satisfactory because of the points on the normal probability plot in Figure 5 showing the approximate linear pattern with normal distribution having small disintegration.[19,45] Also, frequency versus residual plot Figure 5 confirms better the data variation as there were no outliers, and the negatively skewed histogram shows the initial observed S/N ratio has more impacts on results. The plot of residuals versus fitted value shows a random pattern (illustrated in Figure 5), which means residuals have approximately constant variance, and the residuals versus order plot shows that the data collected can be used to find the non-random error.

4.2 Interaction plot and the effect of factors on the change in surface roughness

The effect of the various process factors (Data means) on ΔR_a is shown in Figure 6. Factors corresponding to near-

Table 3. Experimental design.

Exp. Run	Abrasive Size/ µm	Voltage/ V	Machining Gap/ mm	Rotational Speed/ rpm	Experimental ΔR_a (A)	S/N ratio	ANN Predicted ΔR_a (B)	% Error ($\|A-B\|/A*100$)
1	90	6	1	60	0.13	−17.7211	0.1300	0
2	90	6	1.5	90	0.11	−19.1721	0.1099	0.0909
3	90	6	2	120	0.11	−19.1721	0.1101	0.0909
4	90	12	1	90	0.21	−13.5556	0.2103	0.1428
5	90	12	1.5	120	0.14	−17.0774	0.1396	0.2857
6	90	12	2	60	0.11	−19.1721	0.1097	0.2727
7	90	18	1	120	0.23	−12.7654	0.2300	0
8	90	18	1.5	90	0.17	−15.3910	0.1702	0.1176
9	90	18	2	60	0.15	−16.4782	0.1508	0.5333
10	150	6	1	60	0.15	−16.4782	0.1498	0.1333
11	150	6	1.5	90	0.14	−17.0774	0.1401	0.0714
12	150	6	2	120	0.13	−17.7211	0.1294	0.4615
13	150	12	1	90	0.24	−12.3958	0.2397	0.1250
14	150	12	1.5	120	0.15	−16.4782	0.1494	0.4000
15	150	12	2	60	0.13	−17.7211	0.1300	0
16	150	18	1	120	0.25	−12.0412	0.2497	0.1200
17	150	18	1.5	90	0.19	−14.4249	0.1900	0
18	150	18	2	60	0.18	−14.8945	0.1791	0.5000
19	300	6	1	60	0.11	−19.1721	0.1073	2.4540
20	300	6	1.5	90	0.10	−20.0000	0.1002	0.2000
21	300	6	2	120	0.12	−18.4164	0.1198	0.1666
22	300	12	1	90	0.17	−15.3910	0.1707	0.4117
23	300	12	1.5	120	0.16	−15.9176	0.1654	3.3750
24	300	12	2	60	0.10	−20.0000	0.1015	1.5000
25	300	18	1	120	0.21	−13.5556	0.2097	0.1428
26	300	18	1.5	90	0.16	−15.9176	0.1597	0.1875
27	300	18	2	60	0.14	−17.0774	0.1402	0.1428

Table 4. Response table for S/N Ratios – Larger is better.

Level	Abrasive size	Voltage	Machining gap	Rotational speed
1	−16.72	−18.33	−14.79	−17.63
2	−15.47	−16.41	−16.83	−15.93
3	−17.27	−14.73	−17.85	−15.91
Delta	1.80	3.60	3.06	1.73
Rank	3	1	2	4

horizontal line show insignificant effect, whereas factors corresponding to the highest inclined lines have the most significant effect. The most effective factor on ΔR_a was Voltage as its increment was highest, while the other factors show a lesser impact. ΔR_a increases with the increase of Voltage and Rotational Speed. Simultaneously, ΔR_a decreases with the increase of Machining Gap and Abrasive Size. The highest ΔR_a was found for the highest Voltage and the highest Rotational Speed.

4.3 Linear regression model

The linear dependence between the value of controlling factors (Abrasive Size, Voltage, Machining Gap, and Rotational Speed) and ΔR_a was obtained. Using MINITAB 17 software, the linear regression equation was derived from set values of abrasive size, voltage, machining gap, and rotational speed and ANOVA. However, a linear regression model was obtained as following to predict ΔR_a. A linear dependence between parameters and the unknown quantity is provided by the

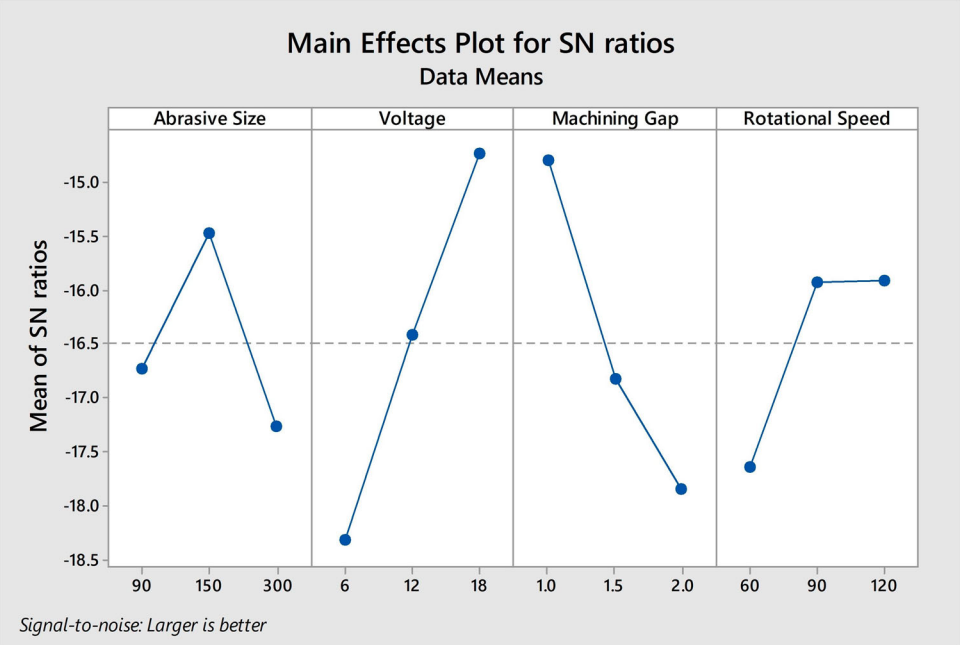


Figure 4. Main effects plot for S/N Ratio for ΔR_a.

Table 5. Analysis of variance for S/N ratio for change in surface roughness

Source	DF	Adj SS	Adj MS	*F*-value	*P*-value	% contribution
Abrasive size	2	15.350	7.6748	14.44	0.000	10.95
Voltage	2	58.344	29.1718	54.88	0.000	41.60
Machining gap	2	39.225	19.6126	36.89	0.000	27.98
Rotational speed	2	13.165	6.5826	12.38	0.000	9.39
Error	18	9.569	0.5316			6.80
Total	26	140.238				

developed model.

$$\text{Experimental } \Delta R_a(A) = 0.1511 - 0.000080d + 0.005370X_1 - 0.0549X_2 + 0.000403X_3 \quad (1)$$

$\Delta R_a = 0.23938$ (at abrasive size $d = 150$ µm, voltage $X_1 = 18$ V, machining gap $X_2 = 1$ mm, rotational speed $X_3 = 90$ rpm)

The developed linear regression equation with the help of MINITAB 15 for the ΔR_a (particularly for 150-micron size abrasive particle) is as below.[20]

$$\Delta R_a = 0.1850 + 0.00556X_1 - 0.0646X_2 + 0.00021X_3 \quad (2)$$

$\Delta R_a = 0.21713$ (at $d = 150$ µm, $X_1 = 18$ V, $X_2 = 1$ mm, $X_3 = 90$ rpm)

Based on the above Equations (1) and (2), a difference has been shown between non-orthogonal array and an orthogonal array, and it applied for the chosen controlling factors and their values. However, from Equations (1) and (2), coefficients showed that each factor has individual impacts, on ΔR_a which were dissimilar. The ΔR_a increases with increasing

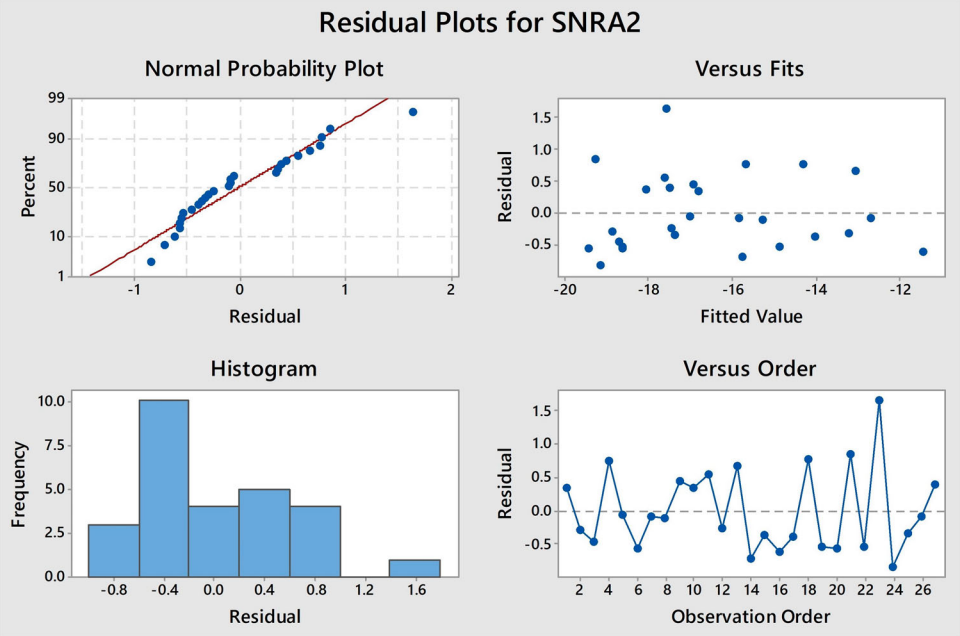


Figure 5. Residual plot for S/N Ratio for the Change in Surface Roughness (ΔR_a).

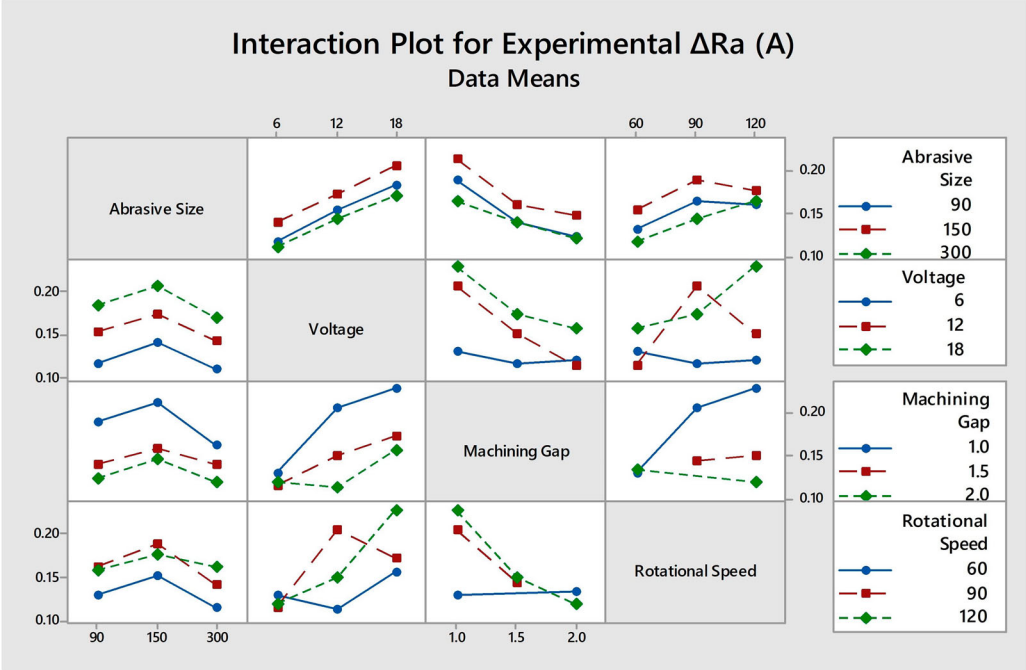


Figure 6. Interaction plot for ΔR_a.

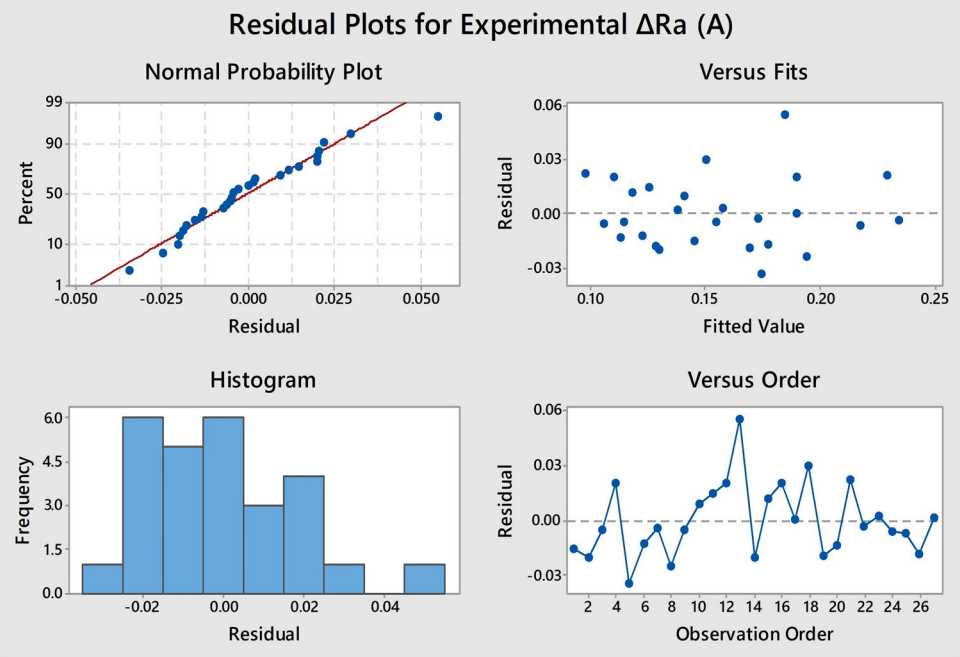


Figure 7. Residual plot for the change of surface roughness (ΔR_a).

Voltage and Rotational Speed, and decreases with increasing Machining Gap and Abrasive Size. The model formulated for prediction of ΔR_a, represented by Equation (1) was satisfactory because of the points on the normal probability plot in Figure 7 showing the approximate linear pattern with normal distribution having small disintegration.[44,45] In Figure 7 the plot of residuals versus fitted value shows a random pattern, which suggests that residuals have approximately constant variance, and the residuals versus order plot shows that the data collected can be used to find the non-random error. Also, the frequency versus residual plot in Figure 7 confirms the finer data variation as there were no outliers; however, the histogram is negatively skewed showing that the initial observations have more weight on the overall results.

4.4 Artificial Neural Network (ANN)

Artificial Neural Networks (ANN), normally referred to as Neural Networks, are analogous to biological neural networks. These networks have been used to approximate a function (in mathematical model form) that depends on a huge number of inputs. Due to the highly complex nature of calculations involved during the development of the network, software is normally used.[46] For data training and testing, MATLAB R2015a software has been used in this work.

Neural networks contain an input layer which is composed of neurons corresponding to the various process parameters. The experimental or realtime data (x_i) related to these parameters are subsequently fed into the neurons of the input layer. Input data are transferred to the hidden layer and output layer by multiplying weights (w_{ij}) between two neurons and adding using a summation function as shown in Equation (3). The feed-forward ANN model developed for modelling of ΔR_a (response or output) contains 4 different input layers of neurons equal to four input parameters (in this case, abrasive size, voltage, machining gap, and rotational speed). The developed ANN architecture was 4-10-8-1, i.e. four inputs, two hidden layers (10 and 8 neurons, respectively), and one output, as illustrated in Figure 8. This ANN architecture (4-10-8-1) has been found more appropriate for the modelling of data for which error is nearly zero as shown in Figure 9(c). For the present data, the ANN training algorithm was '*trainlm*' (trainlm is often the fastest backpropagation algorithm in the toolbox). The activation transfer functions (mathematical function that converts and processes the input data of each

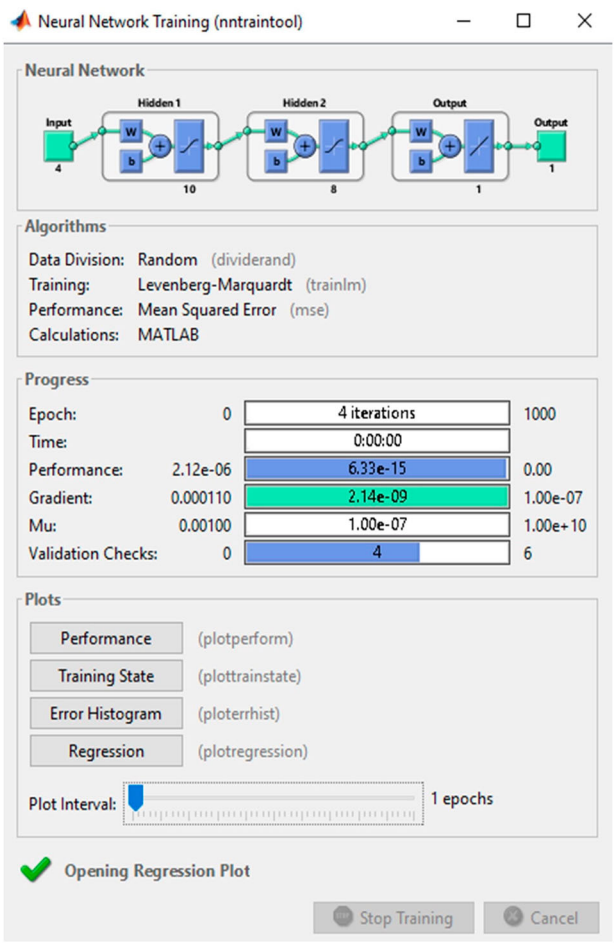


Figure 8. Neural network training.

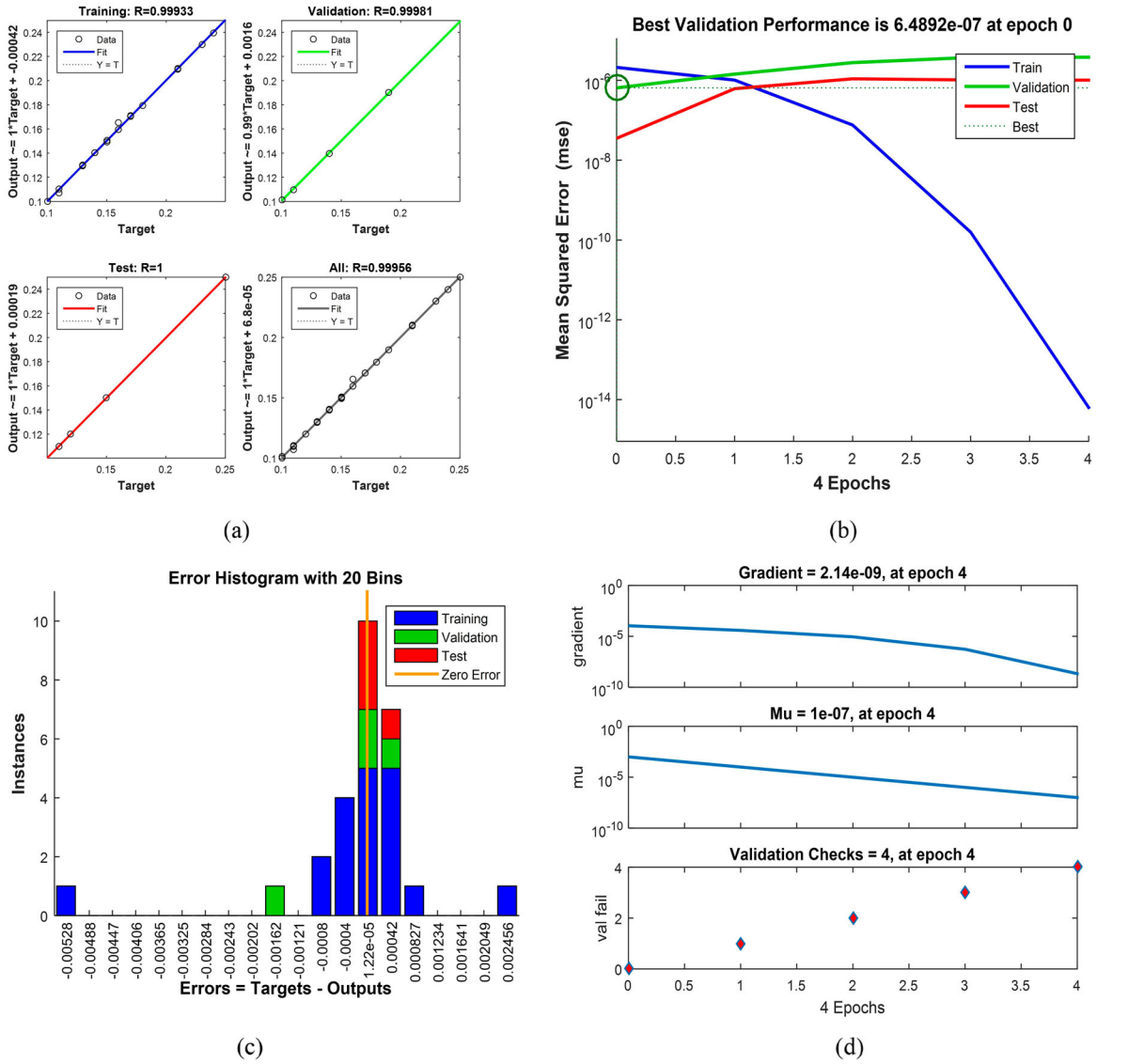


Figure 9 Training plots of neural network for ΔR_a: (a) regression Performance for trained ANN (b) variation of error with epochs (c) Error histogram plot (d) training state of ANN.

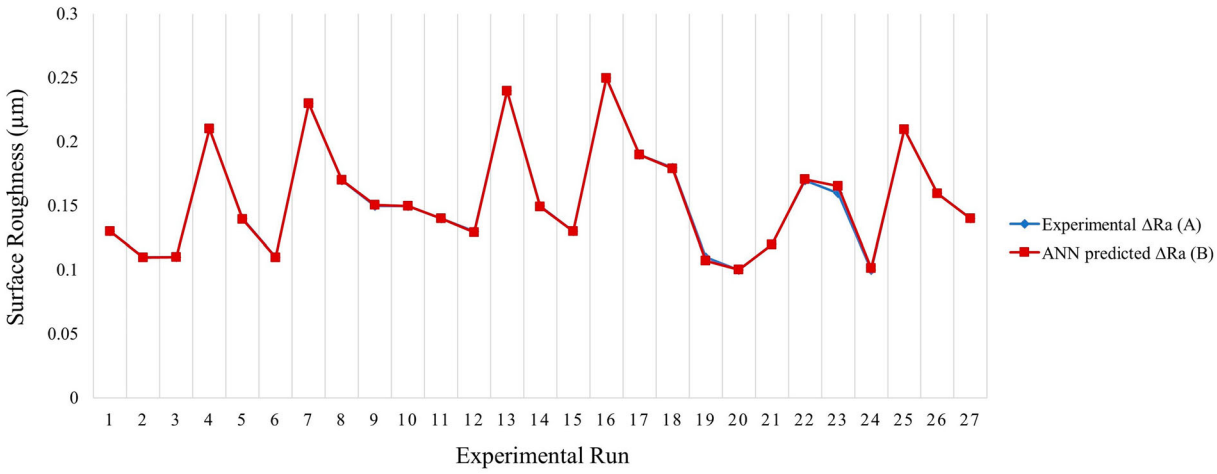


Figure 10 Comparison for surface roughness.

layer to outputs) '*logarithmic sigmoid logsig*' and the linear transfer function '*purelin*' were used as they offer better predictions.[31,32] This training algorithm updates weights and biases in pursuance of '*Levenberg–Marquardt*' optimisation (used to solve non-linear least squares problems). For the training, testing and validation of ANN, 70%, 15%, and 15% data are used, respectively. Regression coefficients can be acquired for the training, validation, and testing after ANN

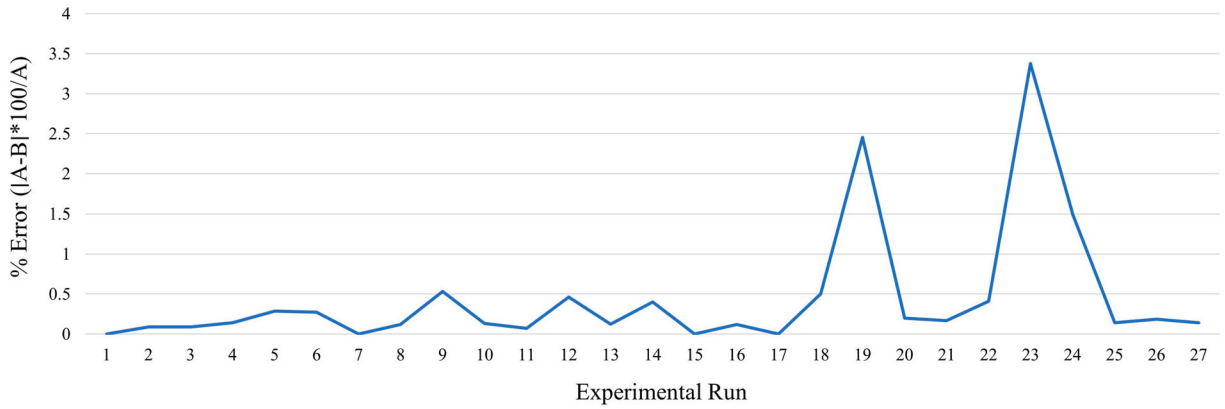


Figure 11 Percentage error plot between experimental and ANN predicted result.

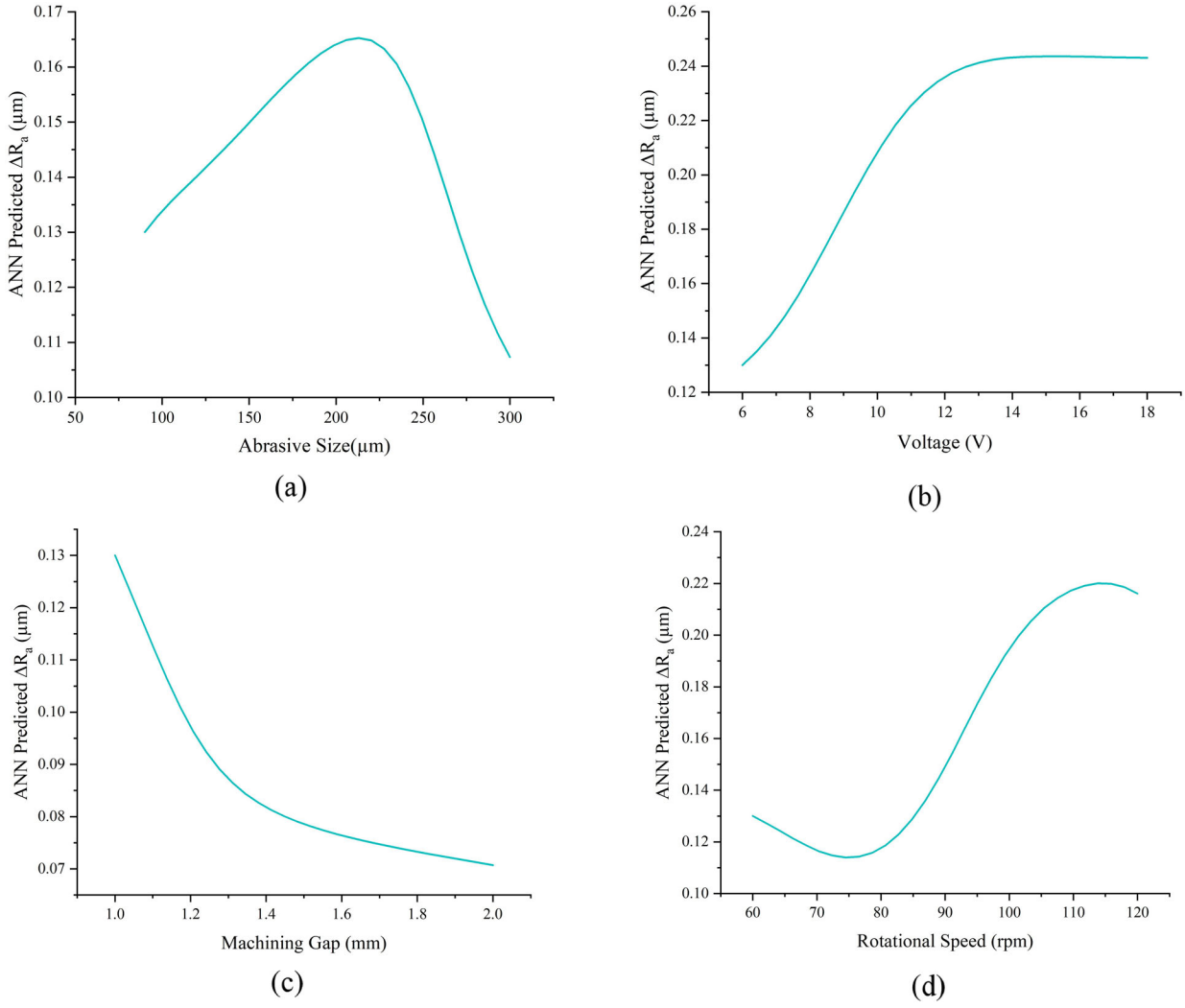


Figure 12. ANN predicted the individual effect of process variables (a) effect of abrasive size on ΔR_a (b) effect of voltage on ΔR_a (c) effect of machining gap on ΔR_a (d) effect of rotational speed on ΔR_a.

training.[32,33,46]

$$\text{Output} = f\left(\sum_{i=1}^{n} b + x_i w_i\right) \tag{3}$$

where, b is bias, experimental or realtime data (x_i), corresponding weights (w_{ij}), and f is activation function.

The trained ANN performance is shown in Figure 9(a), which points to the rationally good performance of trained ANN because the validation and test curves are perfectly on line; the error histogram also confirms this fact in Figure 9(c). It cannot be ignored that the best validation performance has attained zero epoch (which shows the number of passes the ANN training has performed over the entire training dataset) with the value of $6.4892*10^{-7}$, where the training persists until the 4th epoch, Figure 9 (b), showing the best validation performance. Figure 9(d) shows the training state of ANN, where the gradient

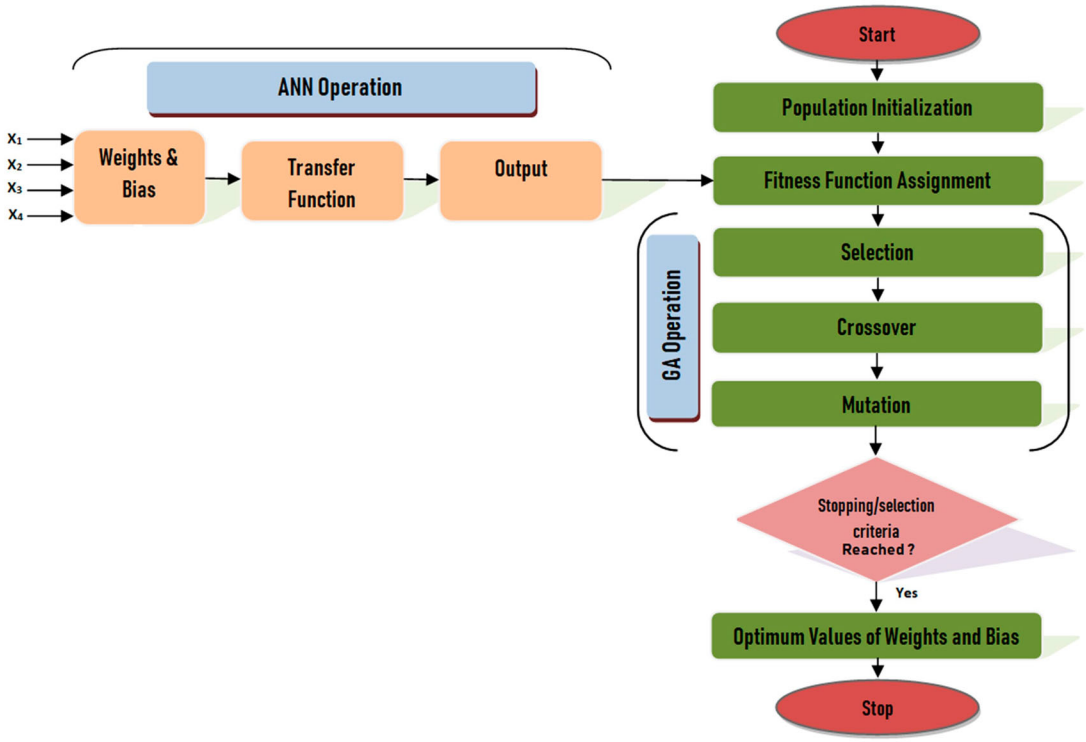


Figure 13. Flow chart of ANN-GA model.

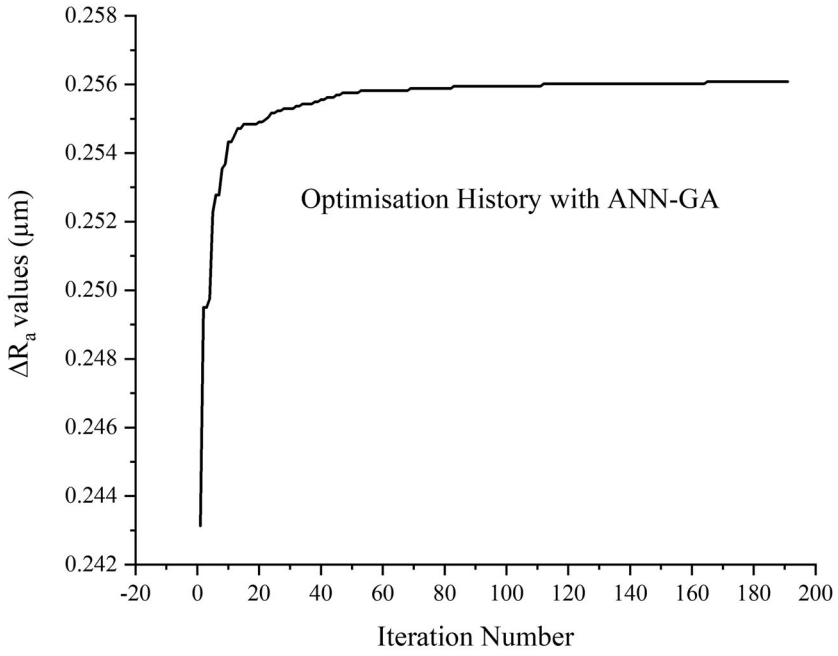


Figure 14. Optimisation history plot.

coefficient's final value is $2.14*10^{-9}$, approximately zero at the 4th epoch. Also, a diminishing value of gradient can be seen with the increase of epochs number. From Figure 10, it is clear that the ANN predicted and experimental values of ΔR_a between initial, and final states conform with each other to a very high degree, with the maximum error equal to 3.43%, and the other errors are negligible. Figure 11 shows the percentage error plot. Hence, it is certain that the developed ANN model has effectively learned the relationship between the input values and ΔR_a values, and thus can be used for maximising the change in surface roughness.

4.5 The individual effect of process parameters predicted by ANN Training

ANN training has predicted the relationship between the ΔR_a with controlling factors by varying the value of one factor, keeping constant value of other factors.

A relationship between the ΔR_a with abrasive size is shown in Figure 12(a). Initially, the trend is increasing until a certain limit, then it decreases. Due to the weaker magnetic bond formation on the edge of an irregularity,[39] the abrasive particles which are tinier than the peaks of irregularities get detached and cannot perform the shearing action satisfactorily. The

Table 6. Parameters of selection

Population size	Population type	Crossover fraction	Mutation rate	Max. generations	Initial population	Initial scores
50	Double vector	0.8	0.01	191	50	50

increasing size of the abrasive particles improves the shearing action and starts removing irregularities. The particles of size bigger than the irregularities shear-off the irregularities of the surface with a ploughing action. These ploughing actions further cause higher surface roughness by removing the material from the root.[47]

Since voltage increase causes an increase in the strength and contact area of the magnetic brush with the target surface, this escalates the normal magnetic force (F_n); consequently, the cutting force (F_c) increases.[4] This increased F_c helps the magnetic abrasive particles to remove irregularities of the surface more efficiently. Hence, the increase in voltage causes an increase in ΔR_a and the same trend has been predicted by ANN as illustrated in Figure 12(b).

When the face of the electromagnet comes closer to the targeted surface, then dense electromagnetic flux causes a higher magnetic force which controls the magnetic abrasive particles to perform more intensely, and vice-versa when further away. Thus, ΔR_a decreases when the machining gap increases, the relationship illustrated in Figure 12(c).

Figure 12(d) shows an increase in ΔR_a when rotational speed in most of the ranges increases; initially it decreases slightly because the normal magnetic force marginally decreases with increase in speed.[4,48]

4.6 Optimisation

The optimal selection of parameters has been carried out to increase the utility and commercial viability of manufacturing processes. Genetic Algorithms (GAs) solve optimisation problems that depend on several variables. Using an input-output pattern of data from the MAF process, by combining the capability of ANN and GA, can solve both the modelling and the optimisation problems.[46]

The basic optimisation problem for this ANN-GA approach is as in the following Equation (4):[49]

$$\text{Optimise } y = f(x, w) \tag{4}$$

where y represents the response or outcome of GA iteration, such as a change in surface roughness; x is a variable (instantaneous process factor during iteration) in the neural network, and w represents weights that evolve in the network training process. $f(.)$ is a fitness function that depends on the nature of input–output relation.

For the accomplishment of the target, a two-phase process has been applied. These two phases are modelling phase and optimisation phase. Conventionally, the standard optimisation problem is to be defined as an objective function. In this surface finishing operation, the difference between the initial and final roughness has to be maximum to achieve minimum surface roughness.

Thus, the present problem is the maximisation of ΔR_a and the objective function can be expressed as:

$$\left.\begin{array}{l}\text{Find values of:(V, x, S, d)for which, } \Delta R_a\text{is maximum}\\ \text{Subjected to range of magnetic abrasive finishing parameters:}\\ 6\text{ Volts} \le \text{V} \le 18\text{ Volts}\\ 1\text{ mm} \le \times \le 2\text{ mm}\\ 90\text{ rpm} \le \text{S} \le 120\text{ rpm}\\ 90\,\mu\text{m} \le \text{d} \le 300\,\mu\text{m}\end{array}\right\} \tag{5}$$

where ΔR_a is change in the surface roughness, 'V' is voltage, 'x' is maching gap, 'S' is rotational speed and 'd' is abrasive size. ΔR_a is a function of the affecting parameters, trained by neurals network.

An effective global optimisation method genetic algorithm (GA) can be used to solve the optimisation problem expressed by Equation (4). GA is a search algorithm, analogous to the process of natural selection that belongs to the class of evolutionary algorithms (EA). Genetic algorithms optimise the problems using methods motivated by the phenomenon of reproduction, mutation, crossover (recombination), and selection.[50] Manufacturing problems that contain data sets as information and defining a continuous and smooth mathematical formula is not possible for 'optimisation objective function,' these can be modelled by ANN and optimised by GA. Figure 13. shows a methodology for the ANN-GA operation.

The decisive parameters of GA optimisation operations are the population size, crossover fraction mutation rate, iterations number (i.e. generations). The opted initial population size was 50, the crossover rate of 0.8, the mutation rate of 0.1, and a number of iterations of 191 were utilised; the optimisation was terminated, and the optimisation history plot up to 191 iterations is shown in Figure 14. Parameters of selection for GA operation are illustrated in Table 6.

The optimisation problem in Equation (4) has been solved with the constraint (lower and upper limits of controlling factors) to search the optimum values of finishing parameters. The contribution of the optimisation process in maximising the ΔR_a is evident when compared with the Taguchi-ANOVA optimised result. From Table 7, ANN-GA shows $(\Delta R_a)_{max}$ of 0.256 µm, whereas Taguchi and ANOVA analysis find $(\Delta R_a)_{max}$ of 0.239 µm.

Table 7. Comparison of results obtained from the Taguchi method and ANN-GA Optimisation

Control factor	Taguchi predicted	ANN-GA Predicted
Abrasive Size/µm	150	105.158
Voltage/V	18	17.999
Machining Gap/mm	1	1.001
Rotational Speed/rpm	90	91.890
Change in Surface Roughness/µm	0.239	0.256

5. Conclusions

Based upon the above results and discussion the following conclusions can be made-

1. ANN-GA is a robust modelling and optimisation tool for any sort of data set (orthogonal array design or non-orthogonal array design) appropriate to scrutinising the finishing of SS302, and it was noticed that the parametric design used in this study provides a straightforward, methodical, and proficient method of modelling and

optimisation of change of surface roughness or finishing behaviour during the MAF process.
2. Modelling and optimisation done with ANN-GA show a maximum value of $(\Delta R_a)_{max}$ equal to 0.25610 µm, which is 7% better than the result obtained from Taguchi-ANOVA analysis.
3. From the regression analysis, ΔR_a increases with the increase of Voltage and Rotational Speed and decreases with increasing Machining Gap and Abrasive Size, whereas the ANN modelling predicts the same results except for the effect of abrasive size, i.e. ΔR_a increases then decreases over the range, and this argument is justified by the optimum value of abrasive size (between the lower and upper limits) obtained from both the optimisation techniques.
4. Voltage has the highest influence on ΔR_a, i.e. 41.60%, followed by Machining Gap (27.98%), while Abrasive Size (10.95%) and Rotational Speed (9.39%) show almost four times lower effect.
5. An ANN model for predicting ΔR_a was developed, and a linear regression equation was developed, where ΔR_a depends on abrasive size, voltage, machining gap, and rotational speed at a confidence level of 99.5%. By comparing the results obtained from ANN methods, and ANOVA, it can be concluded that the ANN model is capable of predicting with very high accuracy and is more efficient.

Acknowledgment and Funding

Authors offer the most sincere gratitude to the Council of Scientific and Industrial Research (CSIR), Government of India, for financial support in the form of Senior Research Fellowships.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

1. S. Khalaj Amineh, A. Fadaei Tehrani and A. Mohammadi: Improving the surface quality in wire electrical discharge machined specimens by removing the recast layer using magnetic abrasive finishing method. *Int. J. Adv. Manuf. Technol.*, 2012, **66**, 1793–1803. doi:10.1007/s00170-012-4459-7.
2. J. Guo, J. Bai, K. Liu and J. Wei: Surface quality improvement of selective laser sintered polyamide 12 by precision grinding and magnetic field-assisted finishing. *Mater. Des.*, 2018, **138**, 39–45. doi:10.1016/j.matdes.2017.10.048.
3. A. K. Singh, S. Jha and P. M. Pandey: Parametric analysis of an improved ball end magnetorheological finishing process. *Proc. Inst. Mech. Eng. Part B J. Eng. Manuf.*, 2012, **226**, 1550–1563. doi:10.1177/0954405412453805.
4. D. K. Singh, V. K. Jain and V. Raghuram: Parametric study of magnetic abrasive finishing process. *J. Mater. Process. Technol.*, 2004, **149**, 22–29. doi:10.1016/j.jmatprotec.2003.10.030.
5. M. Mosavat and A. Rahimi: Numerical-experimental study on polishing of silicon wafer using magnetic abrasive finishing process. *Wear*, 2019, **424–425**, 143–150. doi:10.1016/j.wear.2019.02.007.
6. H. Yamaguchi, A. K. Srivastava, M. A. Tan, R. E. Riveros and F. Hashimoto: Magnetic abrasive finishing of cutting tools for machining of titanium alloys. *CIRP Ann.*, 2012, **61**, 311–314. doi:10.1016/j.cirp.2012.03.066.
7. V. K. Jain: Magnetic field assisted abrasive based micro-/nano-finishing. *J. Mater. Process. Technol.*, 2009, **209**, 6022–6038. doi:10.1016/j.jmatprotec.2009.08.015.
8. V. K. Jain: Abrasive-based nano-finishing techniques: an overview. *Mach. Sci. Technol.*, 2008, **12**, 257–294. doi:10.1080/10910340802278133.
9. N. Sihag, P. Kala and P. M. Pandey: Chemo assisted magnetic abrasive finishing: experimental investigations. *Procedia CIRP*, 2015, **26**, 539–543. doi:10.1016/j.procir.2014.07.067.
10. S. R. Bhagavatula and R. Komanduri: On chemomechanical polishing of Si_3N_4 with Cr_2O_3. *Philos. Mag. A*, 1996, **74**, 1003–1017. doi:10.1080/01418619608242173.
11. A. Misra, P. M. Pandey and U. S. Dixit: Modeling of material removal in ultrasonic assisted magnetic abrasive finishing process. *Int. J. Mech. Sci.*, 2017, **131–132**, 853–867. doi:10.1016/j.ijmecsci.2017.07.023.
12. S. A. Sirwal and A. K. Singh: Analysis of the surface roughness for novel magnetorheological finishing of a typical blind hole workpiece. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.*, 2019, **233**, 1541–1561. doi:10.1177/0954406218776036.
13. V. K. Jain: 'Advanced machining processes', 2013, New Delhi: Allied Publishers.
14. S. Yang and W. Li: 'Surface finishing theory and new technology', 2018, Berlin, Heidelberg: Springer Berlin Heidelberg.
15. D. K. Singh, V. K. Jain, V. Raghuram and R. Komanduri: Analysis of surface texture generated by a flexible magnetic abrasive brush. *Wear*, 2005, **259**, 1254–1261. doi:10.1016/j.wear.2005.02.030.
16. H. Yamaguchi, A. K. Srivastava, M. Tan and F. Hashimoto: Magnetic abrasive finishing of cutting tools for high-speed machining of titanium alloys. *CIRP J. Manuf. Sci. Technol.*, 2014, **7**, 299–304. doi:10.1016/j.cirpj.2014.08.002.
17. S. K. Amnieh, P. Mosaddegh and A. F. Tehrani: Study on magnetic abrasive finishing of spiral grooves inside of aluminum cylinders. *Int. J. Adv. Manuf. Technol.*, 2017, **91**, 2885–2894. doi:10.1007/s00170-016-9970-9.
18. J. Guo, H. Wang, M. Goh and K. Liu: Investigation on Surface Integrity of rapidly solidified aluminum RSA 905 by magnetic field-assisted finishing. *Micromachines (Basel)*, 2018, **9**, 146. doi:10.3390/mi9040146.
19. Y. Gao, Y. Zhao, G. Zhang, G. Zhang and F. Yin: Polishing of paramagnetic materials using atomized magnetic abrasive powder. *Mater. Manuf. Process*, 2019, **34**, 604–611. doi:10.1080/10426914.2018.1532087.
20. S. Ahmad, S. Gangwar, P. C. Yadav and D. K. Singh: Optimisation of process parameters affecting surface roughness in magnetic abrasive finishing process. *Mater. Manuf. Process*, 2017, **32**, 1723–1729. doi:10.1080/10426914.2017.1279307.
21. C.-T. Lin, L.-D. Yang and H.-M. Chow: Study of magnetic abrasive finishing in free-form surface operations using the Taguchi method. *Int. J. Adv. Manuf. Technol.*, 2007, **34**, 122–130. doi:10.1007/s00170-006-0573-8.
22. T. C. Kanish, P. Kuppan, S. Narayanan and S. D. Ashok: A fuzzy logic based model to predict the improvement in surface roughness in magnetic field assisted abrasive finishing. *Procedia Eng.*, 2014, **97**, 1948–1956. doi:10.1016/j.proeng.2014.12.349.
23. R. K. Singh, S. Gangwar, D. K. Singh and V. K. Pathak: A novel hybridization of artificial neural network and moth-flame optimisation (ANN–MFO) for multi-objective optimisation in magnetic abrasive finishing of aluminium 6060. *J. Braz. Soc. Mech. Sci. Eng.*, 2019, **41**, 270. doi:10.1007/s40430-019-1778-8.
24. P. Kala and P. M. Pandey: Experimental study on finishing forces in double disk magnetic abrasive finishing process while finishing paramagnetic workpiece. *Procedia Mater. Sci.*, 2014, **5**, 1677–1684. doi:10.1016/j.mspro.2014.07.356.
25. K. B. Judal, V. Yadava and D. Pathak: Experimental investigation of vibration assisted cylindrical–magnetic abrasive finishing of aluminum workpiece. *Mater. Manuf. Process.*, 2013, **28**, 1196–1202. doi:10.1080/10426914.2013.811725.
26. Y.-H. Lee, K.-L. Wu, J.-H. Jhou, Y.-H. Tsai and B.-H. Yan: Two-dimensional vibration-assisted magnetic abrasive finishing of stainless steel SUS304. *Int. J. Adv. Manuf. Technol.*, 2013, **69**, 2723–2733. doi:10.1007/s00170-013-5242-0.
27. A. A. Moosa: Utilizing a magnetic abrasive finishing technique (MAF) via adaptive nero fuzzy (ANFIS). *Am. J. Mater. Eng. Technol.*, 2013, **1**, 49–53. doi:10.12691/materials-1-3-5.
28. P. Kala, S. Kumar and P. M. Pandey: Polishing of copper alloy using double disk ultrasonic assisted magnetic abrasive polishing. *Mater. Manuf. Process*, 2013, **28**, 200–206. doi:10.1080/10426914.2012.746704.
29. A. C. Wang, L. Tsai, C. H. Liu, K. Z. Liang and S. J. Lee: Elucidating the optimal parameters in magnetic finishing with gel abrasive. *Mater. Manuf. Process*, 2011, **26**, 786–791. doi:10.1080/10426914.2010.505620.
30. N. Sihag, P. Kala and P. M. Pandey: Analysis of surface finish improvement during ultrasonic assisted magnetic abrasive finishing on

chemically treated tungsten substrate. *Procedia Manuf.*, 2017, **10**, 136–146. doi:10.1016/j.promfg.2017.07.040.
31. E. Deniz, O. Aydogmus and Z. Aydogmus: Implementation of ANN-based selective harmonic elimination PWM using hybrid genetic algorithm-based optimisation. *Measurement (Mahwah N J)*, 2016, **85**, 32–42. doi:10.1016/j.measurement.2016.02.012.
32. H. Oktem, T. Erzurumlu and F. Erzincanli: Prediction of minimum surface roughness in end milling mold parts using neural network and genetic algorithm. *Mater. Des.*, 2006, **27**, 735–744. doi:10.1016/j.matdes.2005.01.010.
33. B. Stojanović, A. Vencl, I. Bobić, S. Miladinović and J. Skerlić: Experimental optimisation of the tribological behaviour of Al/SiC/Gr hybrid composites based on Taguchi's method and artificial neural network. *J. Braz. Soc. Mech. Sci. Eng.*, 2018, **40**, 311. doi:10.1007/s40430-018-1237-y.
34. G. Candan and H. R. Yazgan: Genetic algorithm parameter optimisation using Taguchi method for a flexible manufacturing system scheduling problem. *Int. J. Prod. Res.*, 2015, **53**, 897–915. doi:10.1080/00207543.2014.939244.
35. S. Datta, C. Garai and C. Das: Efficient genetic algorithm on linear programming problem for fittest chromosomes. *J. Glob. Res. Comput. Sci.*, 2012, **3**, 8.
36. K. Hanada, H. Yamaguchi and H. Zhou: New spherical magnetic abrasives with carried diamond particles for internal finishing of capillary tubes. *Diam. Relat. Mater.*, 2008, **17**, 1434–1437. doi:10.1016/j.diamond.2008.01.100.
37. H. Kumar, S. Singh and A. Srivastava: Parametric investigations into internal surface modification of brass tubes with alternating magnetic field. *Procedia Manuf.*, 2016, **5**, 1234–1248. doi:10.1016/j.promfg.2016.08.097.
38. D. K. Singh, V. K. Jain and V. Raghuram: Experimental investigations into forces acting during a magnetic abrasive finishing process. *Int. J. Adv. Manuf. Technol.*, 2006, **30**, 652–662. doi:10.1007/s00170-005-0118-6.
39. T. Mori, K. Hirota and Y. Kawashima: Clarification of magnetic abrasive finishing mechanism. *J. Mater. Process. Technol.*, 2003, **143–144**, 682–686. doi:10.1016/S0924-0136(03)00410-2.
40. G. C. Verma, P. Kala and P. M. Pandey: Experimental investigations into internal magnetic abrasive finishing of pipes. *Int. J. Adv. Manuf. Technol.*, 2017, **88**, 1657–1668. doi:10.1007/s00170-016-8881-0.
41. H. J. C. de Souza, M. B. Silva, C. B. Moyses, F. L. Alberto, F. J. Pontes, U. R. Ferreira, R.N. Duarte, C. E. Sanches da Silva: 'Robust design and Taguchi method application' in 'Design of experiments – Applications', (ed. M. Borges Silva), 1–20, 2013, InTech.
42. R. K. Roy: 'A primer on the Taguchi method', 2nd ed, 2010, Dearborn, MI, Society of Manufacturing Engineers.
43. C.-X. Ma, K.-T. Fang and E. Liski: A new approach in constructing orthogonal and nearly orthogonal arrays. *Metrika*, 2000, **50**, 255–268. doi:10.1007/s001840050049.
44. R. Mead: The Non-Orthogonal Design of Experiments. *J. R. Stat. Soc. Ser. A Stat. Soc.*, 1990, **153**, 151–201. doi:10.2307/2982800.
45. C. M. Judd, G. H. McClelland and C. S. Ryan: 'Data analysis: a model comparison approach to regression, ANOVA, and beyond', 3rd ed., 2017, New York, Routledge, Taylor & Francis Group.
46. K. Wang, H. L. Gelgele, Y. Wang, Q. Yuan and M. Fang: A hybrid intelligent method for modelling the EDM process. *Int. J. Mach. Tools Manuf.*, 2003, **43**, 995–999. doi:10.1016/S0890-6955(03)00102-0.
47. V. C. Shukla, P. M. Pandey, U. S. Dixit, A. Roy and V. Silberschmidt: Modeling of normal force and finishing torque considering shearing and ploughing effects in ultrasonic assisted magnetic abrasive finishing process with sintered magnetic abrasive powder. *Wear*, 2017, **390–391**, 11–22. doi:10.1016/j.wear.2017.06.017.
48. V. K. Jain, P. Kumar, P. K. Behera and S. C. Jayswal: Effect of working gap and circumferential speed on the performance of magnetic abrasive finishing process. *Wear*, 2001, **250**, 384–390. doi:10.1016/S0043-1648(01)00642-1.
49. K. Wang, G. L. Kovacs, M. Wozny and M. Fang, Eds.: 'Knowledge enterprise: intelligent strategies in product design, manufacturing, and management', Proceedings of PROLAMAT 2006, IFIP TC5 International Conference, June 15–17, 2006, Shanghai, China, vol. 207. Boston, MA, Springer US, 2006.
50. A. Ghosh and S. Tsutsui: 'Advances in evolutionary computing theory and applications', 2013, Berlin, Springer Berlin.

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Modelling Logistic Growth Model for COVID-19 Pandemic in India

Minni Jain
Delhi Technological University
Department of CSE
New Delhi, India
minnijain@dtu.ac.in

Prajwal Kumar Bhati
Delhi Technological University
Department of CSE
New Delhi, India
prajwalroxman@gmail.com

Pranav Kataria
Delhi Technological University
Department of CSE
New Delhi, India
pranavkataria_bt2k16@dtu.ac.in

Rohit Kumar
Delhi Technological University
Department of CSE
New Delhi, India
rohitmamc2016@gmail.com

Abstract-- **An early analysis of growth dynamics for infectious diseases, like COVID-19, is needed to dissect the crucial driving factors that result in rapid disease transmission, refine the measures taken to control the pandemic and improve disease forecast. The phenomenological models are used to identify the initial climbing growth period of COVID-19 outbreak in India and have modelled 3 major epidemic growth models: Generalized logistic growth, Logistic growth and Generalized growth, to predict the growth in the total number of positive cases, daily increase in the number of positive tested cases and the daily growth rate in confirmed positive cases, dated from Apr 10, 2020, to Apr 20, 2020. The bootstrap resampling method is applied for data prediction to process the sample data, dated from Jan 31, 2020, to Apr 10, 2020, and to calculate the 3 major growth parameters: r (Rate of growth at an early stage), K (Final epidemic size) and C(Number of aggregate cases at time t), which are used to calculate confidence intervals which predict the future direction of the curve and increase in the number of confirmed cases with 95% accuracy for the interval Apr 10, 2020, to Apr 20, 2020. Our models predict exponential and sub-exponential spread rate in the number of positive cases in India from Apr 10, 2020, to Apr 20, 2020. Our findings reveal that significant measures are needed to control the transmission rate of the virus in the community, as the models predict sub-exponential growth in India.**

Keywords – COVID-19; Phenomenological models; Epidemic Growth Models; Logistic Growth; Bootstrap Resampling Method; Growth parameters; Confidence Intervals; Data Prediction.

I. INTRODUCTION

In the early stages of an epidemic, the growth rate of the infection spread follows an exponential path. Extrapolation of available data is the most crucial method used to depict the growth dynamics of the epidemic. The predicted data help the policymakers to take preventive measures in the right direction to curb the spread of disease in the community.[11] In case of an epidemic, epidemic growth models are constructed to extrapolate the rate at which the disease might spread in the community. These epidemic growth models are constructed using the parameters obtained by exploring available data, using bootstrap resampling methods. In the bootstrap resampling method, multiple simulations run on the available dataset. In consecutive simulations, parameters are adjusted to extrapolate the most precise estimates that are further used to plot the respective epidemic growth models.[12] COVID-19 disease is now proclaimed as a pandemic around the globe. India comes among one of the few countries in the world which have implemented complete lockdown as a preventive measure to decrease the growth in positive cases of COVID-19 disease at an early stage of an outbreak. It has now spread across the whole world, badly affecting nations like the USA, Italy and Spain. This paper focuses on extrapolation of available data for India to construct early-stage epidemic growth models to dissect the crucial driving factors that are resulting in the rapid spread of disease, to refine the measures taken by authorities to control the pandemic and improve disease forecast.[4,5].

Logistic growth model construction is favourable and can unravel the growth and fall in the number of cases around the inflection point. If the logistic curve converges at an early stage, it means the pandemic is under control, suggesting that the steps taken to control the pandemic are working. Whereas, if the logistic curve does not converge and shows an exponential growth even after the occurrence of the inflection point, then it indicates that the measures taken in order to manage the pandemic are not working, and the epidemic is still in its early growth stage. Logistic models also involve calculation of important feature parameters like rate of growth at early stage(r), final epidemic size (K) and the number of aggregate cases at time t (C), growth deceleration (p), disease turning point(t)

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:55:03 UTC from IEEE Xplore. Restrictions apply.

which helps in accurate modelling of curve. Hence, the proposed method of logistic modelling is meritorious over conventional methods.

I. LITERATURE SURVEY

Novel coronavirus (COVID-19) belongs to a family of coronaviruses, which can cause various illness like SARS and MERS.[9] According to the World Health Organization, COVID-19 first appeared in the city of Wuhan, Hebei province, in China in December 2019. SARS-CoV-2 is a new mutation Coronavirus, which is identified as the cause of coronavirus disease of 2019.[8]This pandemic has now affected almost every country on the planet, and researchers around the world are racing to find a cure (or vaccine) for this virus.[9] USA, Italy and Spain are the worst hit nations.The situation in USA is very severe, because more than 50,000 people have lost their lives to this virus.[8] Active research in the field of Computers and Mathematics has helped researchers and doctors, to track the rate of growth of this virus in different countries.[1] Research referenced in [1] has shown that the rate of growth of virus in every country follows the same pattern, that is exponential in nature.

II. METHODS AND THEORY

A. DATASET SOURCES

This report is focused on daily data for the number of positive recorded cases from India. The dataset is sourced from the COVID-19 Patient Database[4] which consists of total confirmed cases state wise consisting of patient number, patient gender, age-bracket, travel history,detected city and current status for all cities in India. The data is updated daily. In this report the data from Jan 31 is used when the first case in India was reported to Apr 10 and used the models to forecast outbreak for next 5 to 10 days.

B. EPIDEMIC GROWTH MODELS

India is currently at the initial stage of the spread, that is, the number of positive cases has not reached a saturation peak yet, and the daily increase is still widening.

In the present scenario, the general epidemic growth models like Richards model and complete interval epidemic curves which constitute an extra variable 'α' signifying the imbalance between rise and sink for the growth frequency curve, cannot be constructed as the amount of data is very limited.

Early epidemic growth models usually result in an 'S' shape when they reach saturation stage. This happens when the epidemic is at an early stage and the data available for processing the model is limited.

Hence, this research focuses on 3 major models, which are constructed using few parameters:

1. Logistic Growth Model

$$\frac{dC(t)}{dt} = r(1 - \frac{C(t))}{K})C(t) \quad (1)$$

2. Generalized Logistic Growth Model(GLM):

$$\frac{dC(t)}{dt} = r(1 - \frac{C(t))}{K})C^p(t) \quad (2)$$

3. Generalized Growth Model (GGM):

$$\frac{dC(t)}{dt} = rC^p(t) \quad (3)$$

Here, the following main parameters are calculated to construct accurate growth models:

1. R : Rate of growth at early stage.
2. K : Final epidemic size.
3. C(t) : Number of aggregate cases at time t.
4. p : Signify profile of distinct growth.

Parameter p lies from 0 to 1, where if:

a. p = 0 : Implies constant growth rate.
b. 0 < p < 1 : Implies sub-exponential growth rate.
c. p = 1 : Implies exponential growth rate.

The 2 basic models, which are derived from Richard's model keeping the parameter α equal to 1 (α=1) and p=1, result in logistic development and generalized logistic development models. The development rate for these models deteriorates proportionally to the number of days. These curves tend to saturate the development in the figures for positive cases at an early juncture of epidemic resulting in optimistic results.

The (3) generalized growth model (GGM) allows for a steady increase in outbreak in the early phase (p<1), but it is unable to explain the decomposition of the frequency rate at advanced stages. Therefore, it acts as an upper barrier, or the worst case scenario that is obtained on the presumption that the upsurge will continue to develop following an expanding pattern as before.Logistic development (1) and Generalized Logistic development (2) usually result in an 'S' shaped curve, when the growth in the number of confirmed cases saturates. They predict very optimistic results, which can be seen as the best case scenario in the early stage analysis of the pandemic.

The generalized logistic growth model (2) results in sub-exponential development rate at a preliminary phase, hence it is preferred to study the imbalance between decay and growth dynamics. These three models are constructed using the available data for the total number of actual recorded cases in India.

C. PARAMETER ESTIMATION

For the parameter estimation stage, the Damped least square method, also known as Marquardt-Levenberg algorithm has been used to resolve the problem of least square nonlinear regression for curve fitting. In order to fit the normal function of normal growth in generalized case of Richard's model (α =1, p = 1), commence the starting spike and permit the algorithm to

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:55:03 UTC from IEEE Xplore. Restrictions apply.

calculate one of the 3 parameters optimally. Since, the early phase of growth models does not observe the logistic rate of development, hence to calculate the variability in our model variables, bootstrap resampling method along with a Poisson function is utilized for error estimation $f(\lambda)$, where λ is the expected value of the distribution as in prior inspections [2].

D. *CONFIDENCE INTERVAL*

The confidence interval mean is calculated for parameter estimation developed for Models (1), (2) and (3) by iterating 200 sets of realizations using parametric bootstrap resampling technique as used in prior studies [2].Parameters were then estimated from the realizations to construct 95% accurate Confidence Intervals. A Poisson error structure was used in the parametric bootstrap method for confidence interval estimation. Problems with the appropriate procedure occur when the model cannot be specified separately from the data resulting in unstable variations of the estimated parameters.The models parameters were evaluated for accurate identification using the simulations.

III. RESULTS

The Generalized linear model parameter p, which signifies the distinct profile growth of infected population is estimated to be 0.99 (95% Confidence Interval: [0.98, 1]). This high value of parameter p signifies that the growth rate is heading towards an exponential trend, and major preventive measures need to be enforced at the earliest.

On the contrary, the logistic model depicts a very optimistic best-case scenario , forecasting the total number infected population at 9,547 (95% Confidence Interval: [9349, 9746]) after 5 days (April 15), and 10,723 (95% Confidence Interval: [10514, 10933]) after 10 days.It might be a very optimistic result if the actual infected cases follow the logistic growth model. India announced their confirmed case for the first time on Jan 30. Since 10 April 2020, according to Indian administration, there were 206 deaths in India and 7,062 cases of infection, indicating that the mortality rate stands at 2.91% among the infected population.

While the growth rate of the infected population is slowing as predicted in Figure 5. , future conditions stay undetermined as the total number of infected individuals is close to the tip of inflection. Also, there is a possibility of new major outbursts to occur in the near future.

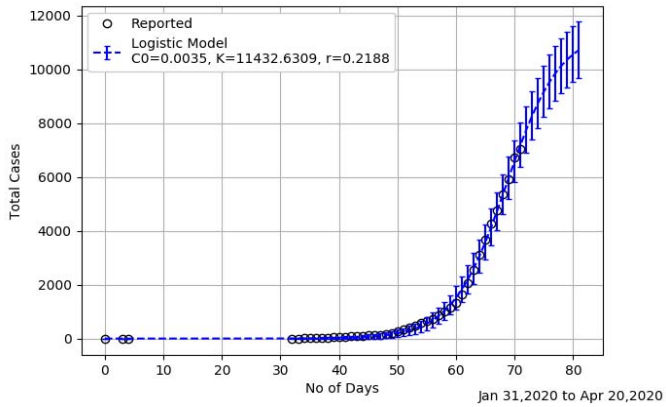


Figure 1. Estimation of Logistic Model to total cumulative confirmed cases in India since Jan31. The error bar is 95% CI extracted from 200 simulations using a poisson error structure.

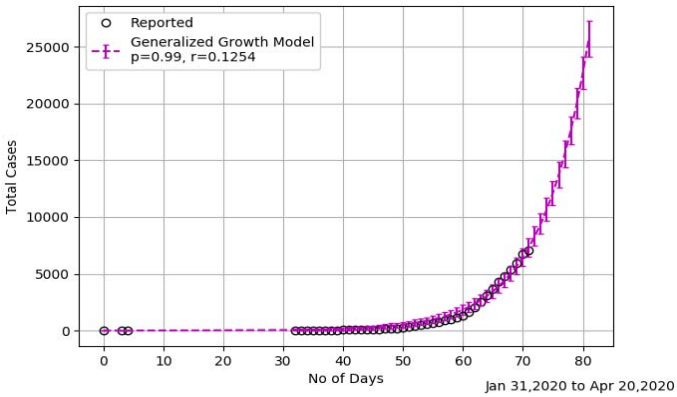


Figure 2. Estimation of Generalized Growth Model to total cumulative confirmed cases in India since Jan31. The error bar is 95% CI extracted from 200 simulations using a poisson error structure

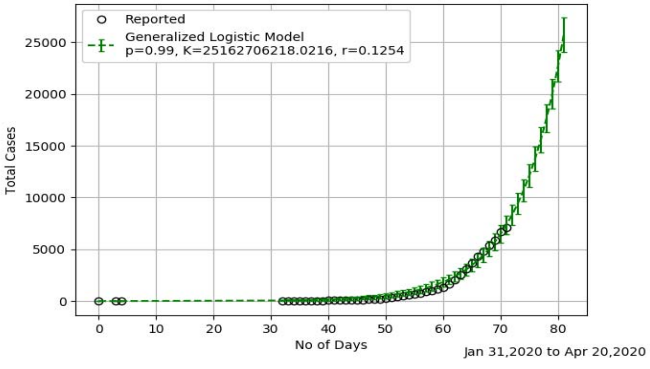


Figure 3. Estimation of Generalized Logistic Model in green to total cumulative confirmed cases in India since Jan31. The error bar is 95% CI extracted from 200 simulations using a poisson error structure.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:55:03 UTC from IEEE Xplore. Restrictions apply.

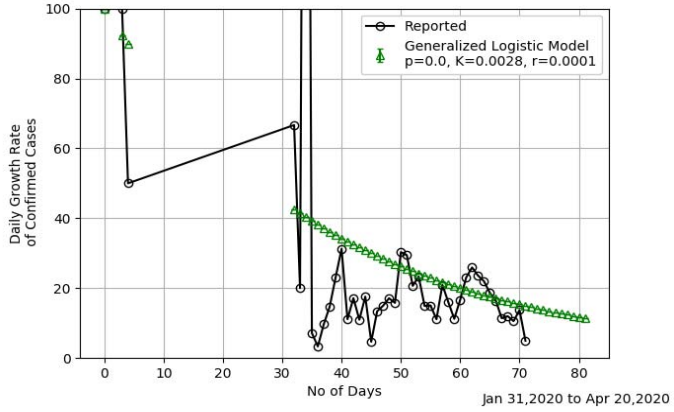


Figure 4. Estimation of models[3](*Generalized Logistic Development Model[1] (in green)*) to rate of daily development of infected population in India since Jan 31. The error bar is 95% CI extracted from 200 simulations using Poisson Error structure[1]

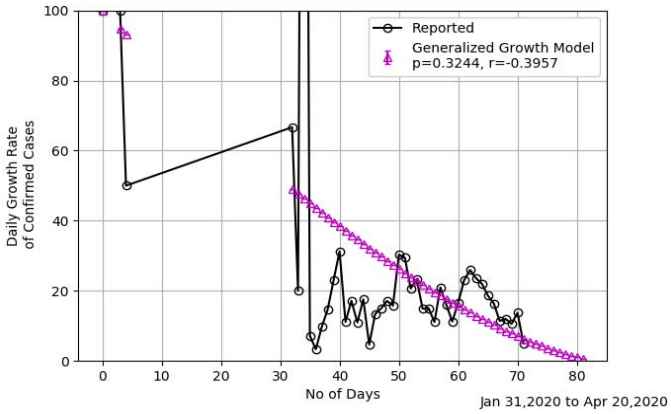


Figure 6. Estimation of models[3](*Logistic Development Model[1] (in blue*),Generalized Growth Model (*in pink*) to rate of daily development of infected population in India since Jan 31. The error bar is 95% CI extracted from 200 simulations using Poisson Error structure[1]

The probability is that the growth rate can keep on fluctuating at current levels for sometime, where the generalized growth model tends to be more favorable for drawing inferences. In this case, the infected population will increase to 13735 (95% CI: [13499, 13970]) in 5 days (April 15th), also to 25706 (95% CI: [25377,26034] within next 10 days (April 20th).

Figure 7. predicts that the point of inflection was passed a few days back, and also the daily number of confirmed infections has started decreasing. This seems to be a positive indication suppose the disease to break out following the logistic depreciation in cases in the following days, presuming the absence of other major public transmission, as was witnessed during phase III of most hotspots in China's research [1].

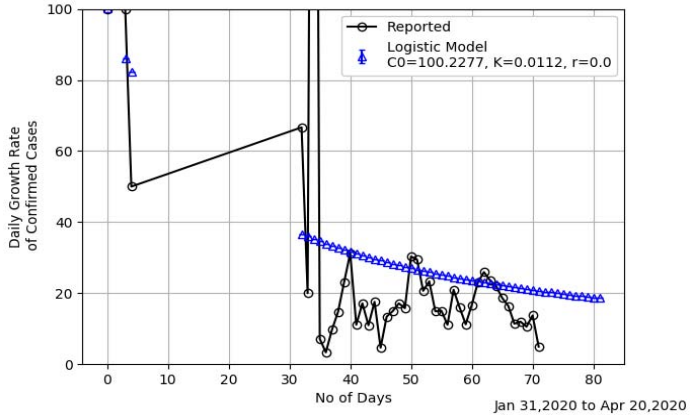


Figure 5. Estimation of models[3](*Logistic Development Model[1] (in blue*),Generalized Growth Model (*in pink*)to rate of daily development of infected population in India since Jan 31. The error bar is 95% CI extracted from 200 simulations using Poisson Error structure[1]

The daily development in the infected population is likely to deteriorate from this junction with an estimated daily surge of 603 (95% Confidence Interval:[550,655]) in 10 days(Apr 20) as predicted by the Logistics Model.

A pessimistic result is obtained from the Generalized Logistic model and Generalized Growth model. Both models estimate the daily increase of infected cases to be : 1355 (95% Confidence Interval: [1279, 1431]) from GLM and 1780 (95% Confidence Interval: [1697,1862]) from GGM, both of which could be predicted to be attained within 10 days

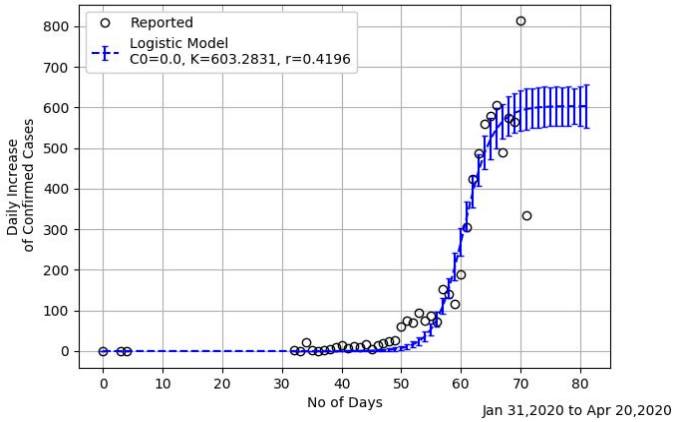


Figure 7. Estimation of models[3] (*Logistic Development Model[1] (in blue)*) to daily increase of infected population in India since Jan 31.The error bar is 95% CI extracted from 200 simulations using Poisson Error structure[1]

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:55:03 UTC from IEEE Xplore. Restrictions apply.

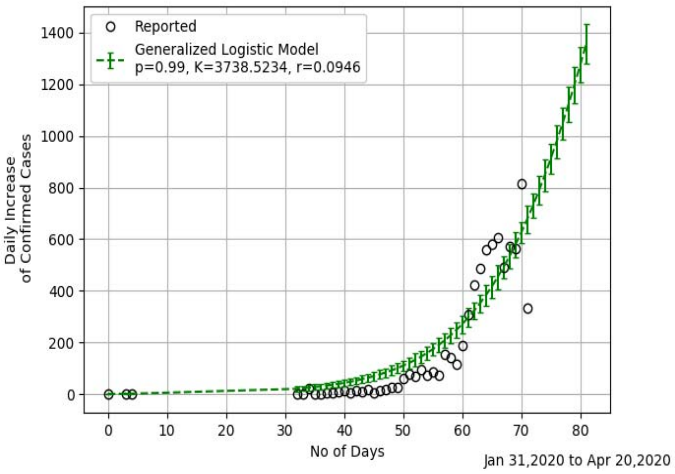


Figure 8. Estimation of models[3](*Generalized Logistic Development Model*[1] *(in green)*) to daily increase of infected population in India since Jan 31. The error bar is 95% CI extracted from 200 simulations using Poisson Error structure[1]

As shown in Figure 4. and Figure 6. the daily growth rate of the confirmed positive cases is definitely not growing exponentially, because the growth rate reduced notably from 40% in February to 1.5% by April 10.The initial period of the Indian outbreak occurred on Jan 30, where most cases are admitted to have a clear travel history to infected places.Although, it is observed from the results that the surge in figures is approaching a constant trend, the uncertainty in future cases might occur.

IV. CONCLUSION AND FUTURE WORK

In this research paper, uniquely introduced variations of logistic models to data of covid-19 pandemic in India and have employed a special 3 parameter Generalized Logistic Model and 2-parameter Generalized Growth Model to model the spread of this outbreak in India. Levenberg–Marquardt algorithm is used for parameter estimation to carry-out least square nonlinear minimization for curve fitting in Python and have calculated confidence intervals with 95% average accuracy, using bootstrap resampling method , where 200 simulations are performed using Poisson error structure. For some days, the number of total cases is predicted with more than 97% of accuracy, which shows the reliability of the models constructed.

With this research paper, tried to model the following growth models: (1)logistic growth development model, the (2)generalized growth development model,(3)generalized logistic growth model to the actual infected population of Covid-19 pandemic from Jan. 30 to Apr 10 of India. Our analysis tries to discover the growth of the covid-19 pandemic in India and the influence of National Lockdown at the average level.

On the grounds of experience in China[1], a high level risk in India is predicted with approximate total actual cases on April 14 to be 12116 (i.e. 95 % CI: [11895, 12336]), 25706 (i.e. 95 % CI: [25337, 26034]) by April 20.The GLM parameter p is around 0.99 (95% Confidence Interval: [0.98, 1]). This indicates that domestic transmission sources in India are out of control as growth is nearing an exponential one.On the contrary, the best case scenario would be the one described in the logistic model, giving the total infected population as 9547 (95% CI: [9349, 9746]) after 5 days (April 15), and 10723 (95% CI: [10514, 10933]) after 10 days. It might be a very good result if the actual infected cases follow the logistic model.Extrapolation of data for India has helped to determine the course of future measures that need to be taken in order to curb the development in COVID-19 disease in India. National lockdown has definitely helped to curb the surge in positive cases in India, but the infection might start spreading rapidly if the lockdown is lifted without prior precautions.

Accuracy of results depends proportionally on the size of the sample data collected. In our research, modelling logistic growth models for COVID-19 pandemic in its most nascent early stages, inferring that the size of sample data available for data extrapolation is very small. Hence, as the sample data size is very limited, saw differences between actual and predicted forecasts using the proposed method.Some external factors also might have resulted in mis-reporting of actual cases.

Table I: Short Term Forecast for daily increase of confirmed cases as predicted by GLM along with actual data and accuracy(%)

Day	Date	Actual	Predicted	Accuracy (%)
72	11 April 2020	854	677	79.27
73	12 April 2020	758	730	96.30
74	13 April 2020	1243	787	63.31
75	14 April 2020	1031	847	82.15
76	15 April 2020	886	911	97.25
77	16 April 2020	1061	978	92.17
78	17 April 2020	922	1188	77.6
79	18 April 2020	1371	1268	92.48
80	19 April 2020	1580	1274	80.63

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:55:03 UTC from IEEE Xplore. Restrictions apply.

Under-testing due to unavailability of testing kits in India might have also resulted in mis-reporting in actual recorded cases, which further resulted in deviation from actual figures in India.

Also, new hotspots might have originated in most affected states like Maharashtra, which might have led to sudden disruptions in the number of reported cases in April 2020.

Regression convergence depends significantly on preliminary values of parameters K, r and p, it might fail to converge accurately for a limited sample dataset.

Therefore, logistic models are highly data-driven in nature and the accuracy in predicted data depends on how well can the data collected mimic the actual dynamics of the pandemic.

The logistic models constructed mimic both the spread of epidemic and also the control in epidemic (S-shaped curve). Curve starts to converge after inflection point, predicting the control in the epidemic. Non-convergence of logistic models implies that the pandemic is not in control and further measures need to be undertaken.

REFERENCES

[1] Ke Wu1,2, Didier Darcet3, Qian Wang4 and Didier Sornette. Generalized logistic growth modeling of the COVID-19 outbreak in 29 provinces in China and in the rest of the world. Preprint: https://www.medrxiv.org/content/10.1101/2020.03.11.20034363v1

[2] Chowell et al., 2007; Chowell et al., 2006a. Transmission dynamics of the great influenza pandemic of 1918 in Geneva, Switzerland: Assessing the effects of hypothetical interventions. Preprint at: https://www.sciencedirect.com/science/article/abs/pii/S0022519305005096

[3] Junling Ma · Jonathan Dushoff · Benjamin M. Bolker · David J.D. Earn. Estimating Initial Epidemic Growth Rates. Preprint at: https://link.springer.com/content/pdf/10.1007/s11538-013-9918-2.pdf

[4] COVID-19 Patient Database for India available at https://docs.google.com/spreadsheets/d/e/2PACX-1vSz8Qs1gE_IYpzlkFkCXGcL_BqR8hZieWVi-rphN1gfrO3H4lDtVZs4kd0C3P8Y9lhsT1rhoB-Q_cP4/pubhtml#

[5] A Literature Review of 2019 Novel Coronavirus During the Early Outbreak Period: Epidemiology, Causes, Clinical Manifestation and Diagnosis, Prevention and Control.

[6] Tian-Mu Chen, Jia Rui, Qiu-Peng Wang,Ze-Yu Zhao, Jing-An Cui, Ling Yin;A mathematical model for simulating the phase-based transmissibility of a novel coronavirus. Preprint at: **https://doi.org/10.1186/s40249-020-00640-3**

[7] Xinguang Chen, Bin Yu; First two months of the 2019 Coronavirus Disease (COVID-19) epidemic in China: real-time surveillance and evaluation with a second derivative model. https://doi.org/10.1186/s41256-020-00137-4

[8] Poudel Adhikari, S.; Meng, S.; Wu, Y.; Mao, Y.; Ye, R.; Wang, Q.; Sun, C.; Sylvia, S.; Rozelle, S.; Raat, H.; Zhou, H. A Literature review of 2019 Novel Coronavirus During the Early Outbreak Period: Epidemiology, Causes, Clinical Manifestation and Diagnosis, Prevention and Control. *Preprints* **2020**, 2020020060 https://www.preprints.org/manuscript/202002.0060/v2

[9] World Health Organization article(WHO); Naming the coronavirus disease (COVID-19) and the virus that causes it https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it

[10] Cécile Viboud, Lone Simonsen, Gerardo Chowell; A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks https://doi.org/10.1016/j.epidem.2016.01.002

[11] Epidemiology article on wikipedia: https://en.wikipedia.org/wiki/Epidemiology

[12] Bootstrap resampling method; Medium Article https://towardsdatascience.com/an-introduction-to-the-bootstrap-method-58bcb51b4d60

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:55:03 UTC from IEEE Xplore. Restrictions apply.

This is an open access article published under an ACS AuthorChoice License, which permits copying and redistribution of the article or any adaptations for non-commercial purposes.

ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Morphological and Swelling Potential Evaluation of *Moringa oleifera* Gum/Poly(vinyl alcohol) Hydrogels as a Superabsorbent

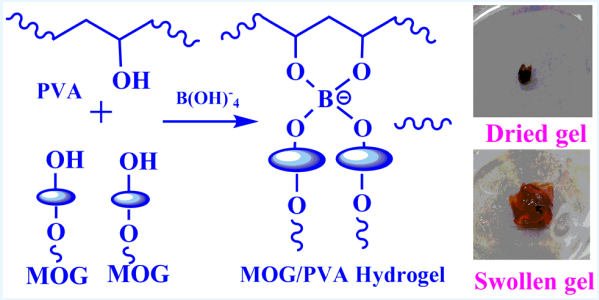
Suhail Ahmad, Kaiser Manzoor, Roli Purwar, and Saiqa Ikram*

Cite This: https://dx.doi.org/10.1021/acsomega.0c01023

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Synergistic use of superabsorbent hydrogels has vital contribution to the daily life. This work gives an account of a facile approach to synthesize superabsorbent hydrogels based on *Moringa oleifera* gum and polyvinyl alcohol cross-linked with borax. Fourier transform infrared, X-ray diffraction, and scanning electron microscopy were employed to characterize the structure, crystallinity, and surface morphologies of the samples. The cross-link density, swelling ratio, reswelling, water retention properties, and salt sensitivity of hydrogels were investigated. Reaction parameters for the hydrogel synthesis were optimized on the basis of water absorbency, and the reaction condition of greater water absorbency after 12 h at room temperature was taken as an ideal condition. Optimum conditions were obtained as [poly(vinyl alcohol) PVA] = 10% (w/v) and [borax] = 1.05×10^{-3} mol/L. Under the optimized conditions, the maximum swelling ratio of MOG/PVA hydrogel reached 1163 g/g in deionized water and 290 g/g in 0.9 wt % NaCl solution. Furthermore, all hydrogels exhibited salt sensitivity and excellent water retention capacity under the high temperature state and displayed smart swelling behaviors in physiological saline solutions. The water absorbency, reusability, and salt sensitivity of the hydrogels give these smart polymer wide promising applications.



1. INTRODUCTION

Superabsorbent hydrogels (SHs) are three-dimensional networks of linear (or branched) hydrophilic polymers interconnected by chemical and/or physical cross-linking. They can absorb and retain large quantities [swelling ratio (SR) > 100] of liquids,[1] blood solutions,[2] and bodily fluids.[3] SHs are broadly used in agriculture,[4] biomedical applications,[5] personal care products,[6] and even in sewage treatments.[7] Moreover, on the basis of their functional groups including −COOH, −OH, $-CONH_2$, $-SO_3H$, and their cross-linking nature,[8] SHs may perform by various stimuli factors such as osmotic pressure, pH, and temperature, which results in a gradual or sudden variation in the SR.[9]

Commonly, water-absorbing gels experience swelling in various phases. Initially, the gel outside comes in contact with water molecules and become wetted through hydrophilic groups present on the gel surface.[10] Second, the water molecules enter in polymer matrices through micropores, which leads in the expansion of gel molecular volume.[11] Lastly, the osmotic pressure of water and the binding force of the molecular matrices reached to equilibrium, as a result, the most extreme water absorbency is achieved.[12] However, the polymer hydrophilic groups tie salt constituent particles effectively by a mechanism similar to complexation.[13] Therefore, the gel structures got blocked, obstructing the entrance of water molecules in the initial phases and volume expansion of the gel in the last phase. Therefore, the swelling capacity decreases abruptly in aqueous fluids, such as blood,[14] urine,[15] and soil solutions.[16] To fulfil these gaps, a simple tactic applied to synthesize superabsorbent materials depends upon natural polysaccharide such as natural gums, cellulose, chitosan, and starch, which might rise their biocompatibility, biodegradability, and swelling capacity.[12]

Moringa oleifera (family: Moringaceae) is a wild plant that grows in many countries of tropics and subtropics. The stem of the *M. oleifera* tree exudes a gum commonly known as *M. oleifera* gum (MOG), which is initially white in color but changes to brownish black and reddish brown on exposure. It gives a highly tacky solution when dipped in water.[17] It is composed of D-galactose, D-mannose, D-glucoronic acid, D-xylose, L-arabinose, and L-rhamnose.[18] All the parts of *M. oleifera* are nontoxic and beneficial for human health and contain minerals, nutrients, and various antioxidants.[19,20] In India, MOG is usually known as sajna gum.

Received: March 6, 2020
Accepted: July 2, 2020

Downloaded via DELHI TECHNOLOGICAL UNIV on July 24, 2020 at 09:26:34 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

© XXXX American Chemical Society

Polymer like poly(vinyl alcohol) (PVA) has fair reputation as a biocompatible and nontoxic linear biopolymer.[21] We selected PVA as the other component because of its fair adhesive properties and water solubility.[22]

The aim of this work was to develop natural gum-based SH with better swelling, physicochemical, and reswelling properties that can vie with synthetic counterparts concerning to the environmental problems. The hydrogel was synthesized using biodegradable and cheap materials: MOG and PVA. The synthesis procedure is aqueous-based, catalyst-free, and environmentally friendly. Here, we described its advantages, procedure, and swelling properties to evaluate its promising application.

2. RESULTS AND DISCUSSION

2.1. Fourier Transform Infrared Spectra. Fourier transform infrared (FTIR) analysis (Figure 1) of MOG/PVA hydrogel displayed a strong absorption band in the region 3401 cm^{-1}, which can be assigned to stretch vibrations of O−H. The band in the region at 2906 cm^{-1} is because of C−H symmetric stretching vibrations. Further a wavenumber at about 1663 cm^{-1} represents hydroxyl bending mode. The characteristic absorption bands at 1163 and 1457 cm^{-1} correspond to C−O−C stretching and CH_2 bending vibrations, respectively. In case of cross-linked hydrogel, the strong absorption band of MOG[17] at 3403 cm^{-1} and PVA[23] at 3400 cm^{-1} was less pronounced and slightly moves to longer wavelength. The appearance of bands around 1429 and 1343 cm^{-1} is because of asymmetric stretching of B−O−C, and also, a wavenumber around 1661 cm^{-1} corresponding to H−OH bending has weakened, and a wavenumber around 837 cm^{-1} corresponding to C−H bending has almost disappeared, which provides evidence that the cross-linking between hydroxyl groups of MOG and PVA by borax.[24]

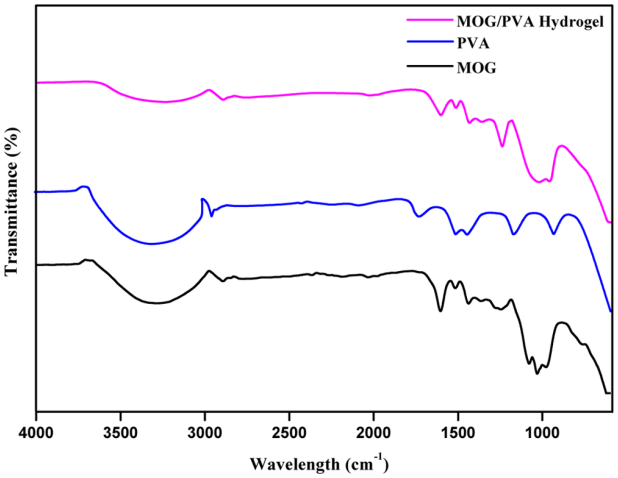


Figure 1. FTIR spectra of MOG, PVA, and MOG/PVA hydrogel.

2.2. X-ray Diffraction Studies. The X-ray diffraction (XRD) pattern (Figure 2) of MOG exhibits a broad peak around 2θ = 10−28°, which corresponds to the amorphous nature of pure MOG.[17] The PVA showed two typical peaks[25] at 2θ of 19.4 and 41°. In case of MOG/PVA hydrogel, the intensity of diffraction peaks at 2θ = 21.9° decreases with slight change in position, suggesting decomposition of semi-crystalline and amorphous nature of PVA and MOG, respectively, which indicates that MOG and PVA were chemically cross-linked.

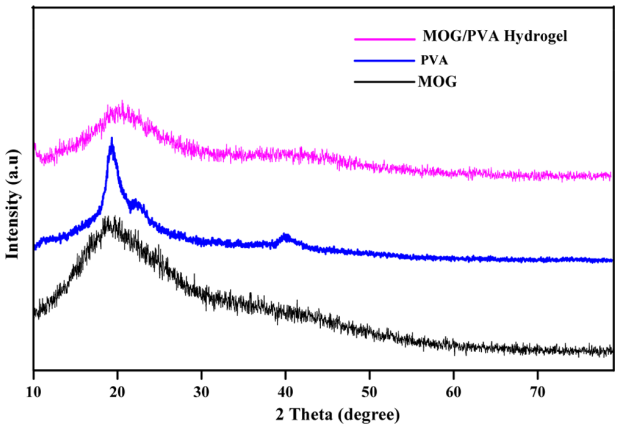


Figure 2. XRD of MOG, PVA, and MOG/PVA hydrogel.

2.3. Scanning Electron Microscopy. Scanning electron microscopy (SEM) images of MOG, PVA, and borax cross-linked MOG/PVA hydrogel are presented in Figure 3a−c, respectively. The MOG exhibited a smooth and compact surface, as shown in Figure 3a. PVA has a smoother and sheet-like morphology, as shown in Figure 3b, whereas the cross-linked hydrogel exhibited a fractured and scaly surface with significant pore spaces, as shown in Figure 3c. These pores are the regions for water permeation and interaction sites of external stimuli with the hydrophilic groups, which leads to increase in SR.

2.4. Swelling in Deionized Water. Figure 4 shows the SR of MOG/PVA hydrogels at different concentrations of PVA and borax at room temperature in deionized water. The water absorbency of hydrogels increases sharply during the first 120 min and afterward turned out to be slower until reaching a plateau. The swelling ability of hydrogels decreases with increasing cross-linker borax concentration, owing to the reason that the increase of cross-linking density makes polymer chains self-entangled, resulting in more compact structure. However, when the PVA content increases, the swelling of hydrogels first increased and then decreased, and the initial increase in swelling may be because of the increase in hydrophilicity and the formation of network of optimum pore size. Thereafter, further increase in PVA concentration has led to decrease in swelling of hydrogels, this may be because of the blocking of active sites on polymer chains, leading to decrease in the hydrophilic groups. The best result was obtained with SR 1163 (g/g), which was in agreement that the MOG/PVA hydrogels were superabsorbent.

2.5. Swelling in Various Saline Solutions. Figure 5 shows the equilibrium swelling capacity of MOG/PVA hydrogels in various saline solutions (D-glucose, urea, physiological saline water, and synthetic urine). The order of equilibrium swelling capacity for SHs in above solutions in the decreasing order was ranked as urea solution > D-glucose solution > physiological saline water > synthetic urine. The equilibrium swelling capacity of SHs in D-glucose solution was found to be slightly lower compared to the values measured in deionized water, whereas the equilibrium swelling capacity subsequently decreases in synthetic urine and in physical saline water. The shrinking behavior of hydrogels in synthetic urine and physiological saline water is because of the charge-

https://dx.doi.org/10.1021/acsomega.0c01023
ACS Omega XXXX, XXX, XXX−XXX

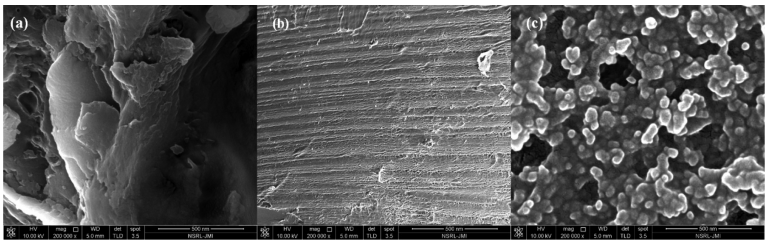


Figure 3. SEM images of (a) MOG, (b) PVA, and (c) MOG/PVA hydrogel.

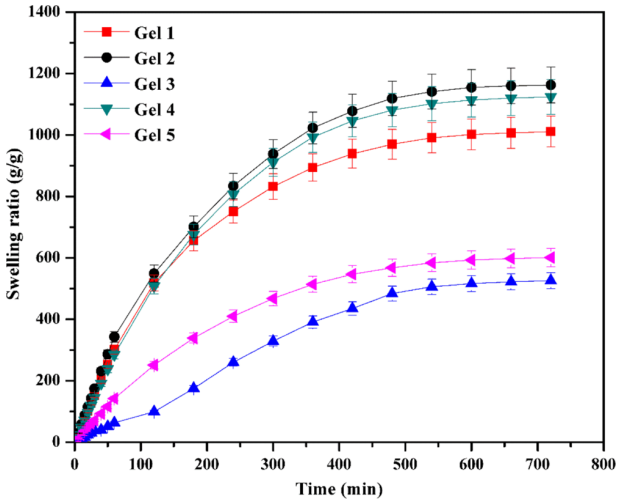


Figure 4. SR of MOG/PVA hydrogels in deionized water at room temperature.

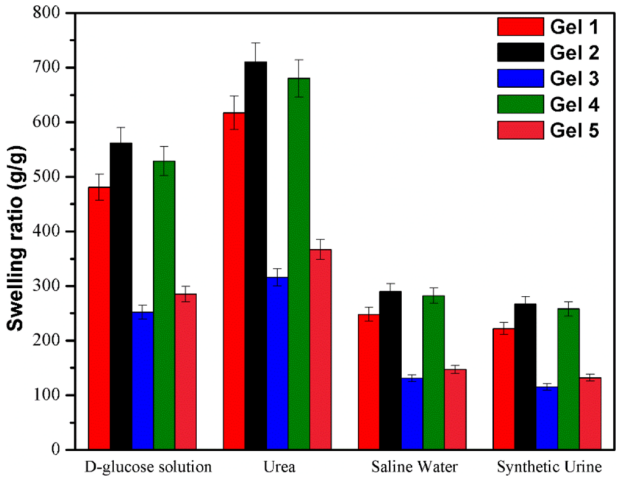


Figure 5. SR of MOG/PVA hydrogels in various saline solutions.

screening effect of the counterions (Na^+, K^+, Ca^{2+}, and Mg^{2+}) resulting in the decline of anion−anion repulsions, which reduces the osmotic pressure between the internal hydrogel network and the external solution.[26]

2.6. Water Retention Capacity. Figure 6 depicts the water retention capacity of the equilibrium swollen MOG/PVA hydrogels. As can be seen from the curves, the water retention capacity decreases with increasing time. After dehydration, gel samples became plastic-like; moreover, water retention decreases with increasing borax amount. GEL 3 losses water rapidly compare to GEL 2, and became

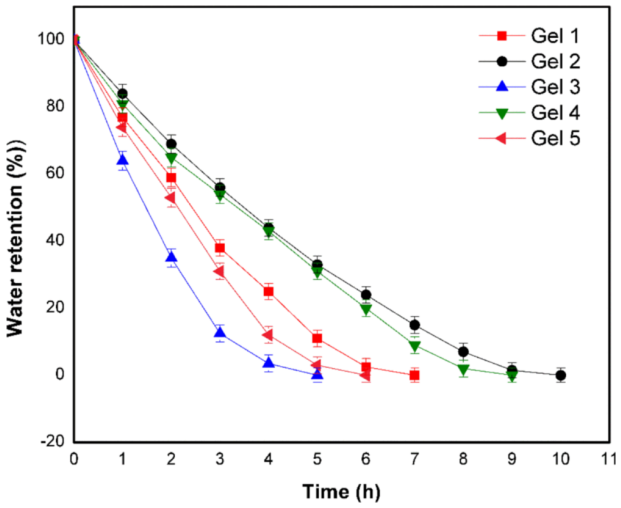


Figure 6. Water retention capacity of the MOG/PVA hydrogels at constant temperature (60 °C).

highly dehydrated after 5 h, while water retention capacity of GEL 2 was reached to 31% at this time interval. This may be because of the H-bonding interactions and van der Waals forces between water molecules and hydrogel matrices, which directly affect the water retention capacity of hydrogel.[27] Considering the above, MOG/PVA hydrogels have outstanding water retention capacity, which allows for a further application prospects in agriculture.

2.7. Reswelling Capability Studies. The reswelling capability of hydrogels in deionized water following five consecutive swelling/drying cycles is depicted in Figure 7. The enormous weight and larger pressure inside the hydrogel system are caused by absorbed water molecules, resulting in the breaking of polymer networks.[28] Thus, the voids in the hydrogel porous structure collapses slowly, resulting in the reduced swelling capability. The equilibrium swelling capacity of bone-dried GEL 2 was 1035 g/g for cycle 1 and remained at 756 g/g for cycle 5, exhibiting swelling capability loss to 35%. The swelling capability of GEL 4, GEL 1, GEL 5, and GEL 3 after five reswelling cycles decreased to 40, 49, 58, and 72%, respectively. These results infer the reason of hydrogen-bonding interactions among MOG and PVA that arisen during the drying process.

2.8. Salt-Sensitivity Analysis. To examine the salt-sensitivity factor (f) of the MOG/PVA hydrogels, the swelling capacity at equilibrium (W_s) in various salt solutions (NaCl, KCl, NH_4Cl, $CaCl_2$, and $AlCl_3$) at 0.15 mol/L concentration is given in Table 1. In NaCl, KCl, and NH_4Cl solutions, the equilibrium swelling capacity W_s of all hydrogels increased in the following ascending order: $Na^+ < K^+ < NH_4^+$ (Table 1).

https://dx.doi.org/10.1021/acsomega.0c01023
ACS Omega XXXX, XXX, XXX−XXX

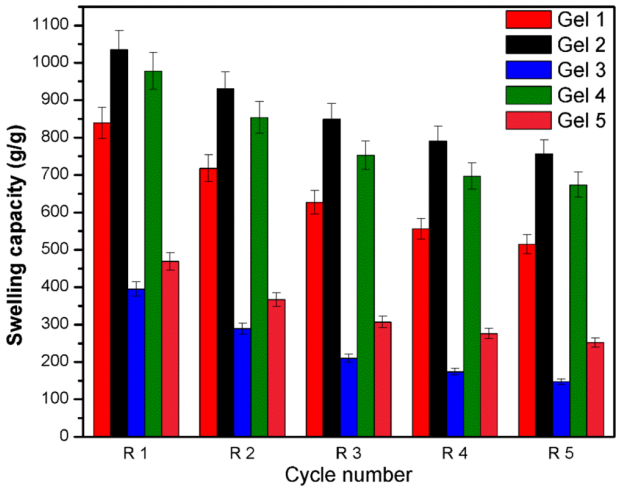


Figure 7. Reswelling capability of MOG/PVA hydrogels during consecutive swelling/drying cycles.

This may be because of the charge screening effect of these cations on the W_s values increased in the same order. The equilibrium swelling capacity in NaCl, $CaCl_2$, and $AlCl_3$, decreases with the increasing of cationic charge (Na^+ < Ca^{2+} < Al^{3+}) (Table 1), which might be elucidated by the Donnan equilibrium theory.[29,30] According to Donnan, the shrinking and swelling behavior were ascribed to the decrease osmotic pressure difference between the interior hydrogel and the external solution.[31]

2.9. Network Parameters. The gel content influences the cross-link density (CD) of the hydrogel network and it rely upon the composition of the reaction mixture. Higher swelling capacity is because of low CD of hydrogels, and at lower CD, the water permeation into the hydrogel matrix becomes easier, resulting in the increment in gel-swelling capacity. Table 2 shows with the increase in borax concentration, the CD of the forming gel gets increased while M_c decreased, this may be because of the number of efficient cross-links per unit volume increases in the hydrogel network, leading to a reduction in pore size of the voids present in the network chains. CD of gel 2 is lower than gel 4 and gel 5 because of the presence of hydrophilic functional groups and adequate pore size. Moreover, the network parameters influence the hygroscopicity, strength, and integrity of hydrogels.

3. CONCLUSIONS

In this research work, a series of borax cross-linked MOG/PVA hydrogels were developed and optimized as the water conservation material in agriculture. The procedure for hydrogel synthesis is ecofriendly as it employs biodegradable, catalyst-free, and inexpensive materials. The SRs of MOG/PVA hydrogels changed dramatically from 526 to 1163 g/g, depending upon the composition of the reaction mixture (borax and PVA). The swelling and shrinking behavior of MOG/PVA hydrogels were influenced by inorganic salt aqueous solutions. The water retention test confirms the good water holding property of the MOG/PVA hydrogel. This biobased SH has great potential for large-scale commercial applications including agriculture and hygienic products.

4. EXPERIMENTAL SECTION

4.1. Materials. MOG (Nutramine Life Sciences, Delhi), polyvinyl alcohol (SD Fine Chemicals India, M_w 85,000−125,000, alcoholysis degree 98−99%), and borax (Qualigens Fine Chemicals) are used. Deionized water was used throughout the experimentation. The solvents and chemicals were used without any further purification.

4.2. Preparation of the MOG/PVA Hydrogel. For the preparation of MOG/PVA hydrogel, definite concentration of MOG (4% w/v) was hydrated and stirred at 200 rpm with a overhead stirrer for 6−8 h at room temperature. PVA solutions of different concentrations (8% (w/v), 10% (w/v), and 12% (w/v)) were prepared separately by dissolving required quantity of PVA in deionized water at 80 °C. The cross-linker borax was separately dissolved in the calculated amount in deionized water (Table 2) and was added to the reacting solution mixture and again stirred for 30 min and kept aside for another 4 h without stirring for proper cross-linking. To remove unreacted borax, the hydrogels were rinsed with deionized water and soaked overnight in acetone for dehydration. Dehydrated hydrogels were filtered through a nylon sieve and dried in an oven at 45 °C till constant weight was obtained. The procedure for the synthesis of MOG/PVA hydrogel is described in Figure 8a, whereas dried and swollen hydrogels are depicted in Figure 8c,d, respectively.

4.3. Characterization of the MOG/PVA Hydrogel. FTIR spectra of MOG, PVA, and MOG/PVA hydrogel were recorded on a FTIR spectrophotometer (model: Tensor 37 spectrophotometer), within the frequency range of 4000−500 cm^{-1} by the KBr pellet method. XRD patterns of MOG, PVA, and MOG/PVA hydrogel were collected using an X-ray diffractometer (model: Bruker) with Cu Kα radiation (λ = 1.54 Å) at a voltage of 40 kV and a current of 40 mA in the 2θ range of 10−80°. The surface morphology of MOG, PVA, and MOG/PVA hydrogel was observed by SEM (model: S-4800 Hitachi, Japan). Before the analyses, a thin layer of gold was sputter-coated on the sample surface.

4.4. Swelling Measurements. A series of accurately weighed bone-dried hydrogel samples (0.1 ± 0.0001 g) was dipped entirely in deionized water or various saline solutions [urea solution: 50 g urea in 1000 mL deionized water; 0.9 wt % NaCl aqueous (saline) solution; D-glucose solution: 50 g D-glucose in 1000 mL deionized water; synthetic urine: 0.6 g

Table 1. Effects of Salt Type on Equilibrium Swelling Capacity (W_s) and Salt-Sensitivity Factor (f) of MOG/PVA Hydrogels

	gel 1		gel 2		gel 3		gel 4		gel 5	
salt	W_s	f	W_s	F	W_s	f	W_s	f	W_s	f
NaCl	175	0.827	213	0.816	83	0.843	201	0.821	99	0.836
KCl	191	0.811	231	0.801	91	0.827	219	0.805	107	0.822
NH_4Cl	207	0.795	249	0.785	101	0.808	237	0.789	119	0.802
$CaCl_2$	83	0.918	106	0.909	39	0.926	97	0.914	46	0.924
$AlCl_3$	39	0.961	51	0.956	14	0.973	45	0.959	19	0.968

https://dx.doi.org/10.1021/acsomega.0c01023
ACS Omega XXXX, XXX, XXX−XXX

Table 2. Reaction Parameters and Properties of Cross-linked Hydrogels

code	MOG % (w/v)	PVA % (w/v)	borax (mol/L)	physical condition of gel	M_c (g/mol)	mean pore size (μm)	SR (g/g)
gel 1	4	10	0.081×10^{-3}	weak	3.98×10^6	513	1011.6 ± 0.7
gel 2	4	10	1.05×10^{-3}	moderate	3.76×10^6	497	1163.5 ± 1.1
gel 3	4	10	1.31×10^{-3}	strong	1.43×10^6	305	526.9 ± 0.4
gel 4	4	8	1.05×10^{-3}	moderate	3.14×10^6	453	1124.1 ± 0.9
gel 5	4	12	1.05×10^{-3}	moderate	2.02×10^6	387	601.3 ± 0.6

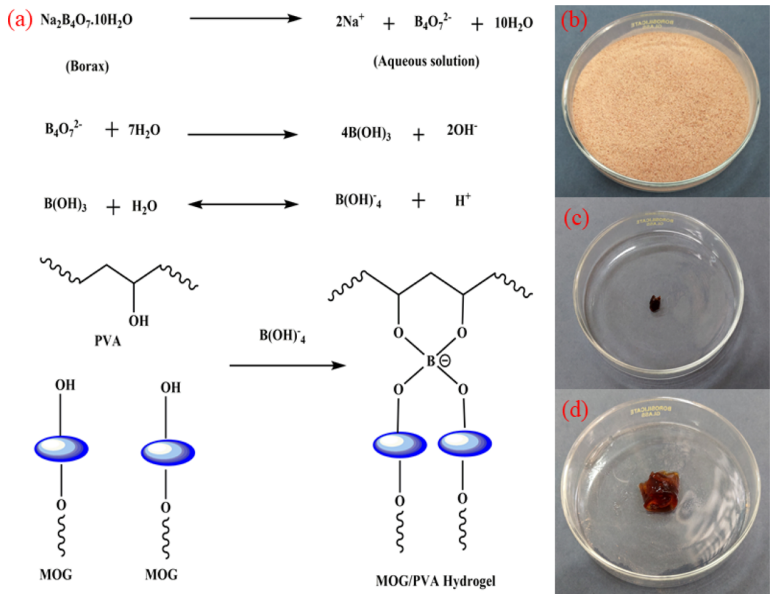


Figure 8. (a) Schematic representation for the synthesis of MOG/PVA SHs, (b) MOG powder, (c) dried hydrogel, and (d) swollen hydrogel.

$CaCl_2$ + 1 g $MgSO_4$ + 8 g NaCl + 20 g urea in 1000 mL deionized water] and allowed to achieve swelling equilibrium at room temperature.

At certain time intervals, swollen hydrogels were withdrawn from the swelling medium, and later, excess surface water (or salt solution) was removed by filter paper and weighed accurately (W_t). This process was repeated until the weight of the swollen hydrogel sample remained constant (W_e), which demonstrated that the hydrogel had reached its equilibrium swelling capacity. The water absorbency of hydrogel (SR; g/g) was calculated by using eq 1, where (W_d) is the weight of the dry hydrogel samples. The equilibrium swelling capacity (W_{eq}, g/g) was calculated using eq 2, where (W_d) is the same as eq 1.

$$\mathrm{SR} = \frac{W_t - W_d}{W_d} \tag{1}$$

$$W_{eq} = \frac{W_e - W_d}{W_d} \tag{2}$$

4.5. Water Retention Measurement. To study the percentage water retention of hydrogel samples under constant temperature, a certain quantity of water-swollen hydrogel was placed in a vacuum oven at high temperature (60 °C) and weighed every 1 h. The percentage water retention (W_R) was detected by the following eq 3

$$W_R = \frac{W_i - W_d}{W_e - W_d} \times 100\% \tag{3}$$

where W_i is the weight of the swollen hydrogel and W_d and W_e are the same as eq 2.

4.6. Reswelling of Cross-Linked Hydrogels. A preweighted dry hydrogel sample was immersed in deionized water to ensure that the swelling equilibrium was achieved and weighted. The swollen hydrogels were bone-dried in a vacuum oven at 80 °C. The dried hydrogel sample was soaked again in deionized water to determine its equilibrium swelling capacity. This swelling/drying cycle was repeated five times using the same hydrogel, and finally losses of equilibrium swelling capacity were measured for each hydrogel sample.

4.7. Salt-Sensitivity Factor (*f*). The effect of ions (anions and cations) on the swelling capacity of the hydrogels was evaluated in various salt solutions as described by Rodrigues et al.[32] The dimensionless salt-sensitivity factor (f) was determined using eq 4:

$$f = 1 - \left(\frac{W_s}{W_w}\right) \tag{4}$$

where W_s and W_w are the equilibrium swelling capacity in saline solution and in deionized water, respectively. All salts were used at concentration of 0.15 mol/L.

4.8. Network Parameters. Some important network parameters used to characterize the network structure of hydrogel were determined. The hydrogel cross-linking density (CD) is obtained by the average molar mass (M_c) entrapped between two neighboring cross-linking points of the polymer chain contained in the gel. The CD value is inversely proportional to M_c, as the higher the M_c, the lower the CD. According to the Flory–Huggins theory[33] and a previously developed method,[34,35] the M_c value of any hydrogel can be determined using eq 5

https://dx.doi.org/10.1021/acsomega.0c01023
ACS Omega XXXX, XXX, XXX−XXX

$$M_c = Q^{5/3}(D_2V_1/(0.5 - X_1)) \quad (5)$$

where Q represents the equilibrium water absorption of the hydrogel, D_2 denotes the hydrogel density, V_1 is the solvent volume used for swelling, and X_1 is the Flory−Huggins interaction parameter between the solvent and hydrogel. The mean pore size or mesh size of the polymeric network of the hydrogels (ξ) broadly defines the linear distance between the macromolecular chains in a cross-linked network and is estimated using eq 6.[36−38]

$$\xi = 0.071(M_c)^{1/2}/(\phi)^{1/3} \quad (6)$$

Here, (ϕ) is the polymer volume fraction in the swollen state.

AUTHOR INFORMATION

Corresponding Author

Saiqa Ikram − *Department of Chemistry, Faculty of Natural Sciences, Jamia Millia Islamia, New Delhi 110025, India;* orcid.org/0000-0003-2274-7669; Phone: +91-1126981717 ext. 3255; Email: sikram@jmi.ac.in

Authors

Suhail Ahmad − *Department of Chemistry, Faculty of Natural Sciences, Jamia Millia Islamia, New Delhi 110025, India; Applied Chemistry & Polymer Technology, Delhi Technological University, New Delhi 110042, India*

Kaiser Manzoor − *Department of Chemistry, Faculty of Natural Sciences, Jamia Millia Islamia, New Delhi 110025, India*

Roli Purwar − *Applied Chemistry & Polymer Technology, Delhi Technological University, New Delhi 110042, India*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.0c01023

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors are highly thankful to University Grants Commission, India and Scheme for Promotion of Academic & Research Collaboration (MHRD-SPARC) under Ministry of Human Resources, India.

REFERENCES

(1) Zhao, C.; Zhang, M.; Liu, Z.; Guo, Y.; Zhang, Q. Salt-Tolerant Superabsorbent Polymer with High Capacity of Water-Nutrient Retention Derived from Sulfamic Acid-Modified Starch. *ACS Omega* **2019**, *4*, 5923.

(2) Ahmad, S.; Ahmad, M.; Manzoor, K.; Purwar, R.; Ikram, S. A Review on Latest Innovations in Natural Gums Based Hydrogels: Preparations & Applications. *Int. J. Biol. Macromol.* **2019**, *136*, 870.

(3) Alam, M. N.; Christopher, L. P. Natural Cellulose-Chitosan Cross-Linked Superabsorbent Hydrogels with Superior Swelling Properties. *ACS Sustainable Chem. Eng.* **2018**, *6*, 8736.

(4) Cheng, D.; Liu, Y.; Yang, G.; Zhang, A. Water- and Fertilizer-Integrated Hydrogel Derived from the Polymerization of Acrylic Acid and Urea as a Slow-Release N Fertilizer and Water Retention in Agriculture. *J. Agric. Food Chem.* **2018**, *66*, 5762.

(5) Ali, A.; Ahmed, S. Recent Advances in Edible Polymer Based Hydrogels as a Sustainable Alternative to Conventional Polymers. *J. Agric. Food Chem.* **2018**, *66*, 6940.

(6) Kabir, S. M. F.; Sikdar, P. P.; Haque, B.; Bhuiyan, M. A. R.; Ali, A.; Islam, M. N. Cellulose-Based Hydrogel Materials: Chemistry, Properties and Their Prospective Applications. *Prog. Biomater.* **2018**, *7*, 153.

(7) Peng, N.; Hu, D.; Zeng, J.; Li, Y.; Liang, L.; Chang, C. Superabsorbent Cellulose-Clay Nanocomposite Hydrogels for Highly Efficient Removal of Dye in Water. *ACS Sustainable Chem. Eng.* **2016**, *4*, 7217.

(8) Ashraf, M. U.; Hussain, M. A.; Muhammad, G.; Haseeb, M. T.; Bashir, S.; Hussain, S. Z.; Hussain, I. A Superporous and Superabsorbent Glucuronoxylan Hydrogel from Quince (Cydonia Oblanga): Stimuli Responsive Swelling, on-off Switching and Drug Release. *Int. J. Biol. Macromol.* **2017**, *95*, 138.

(9) Ahmed, E. M. Hydrogel: Preparation, Characterization, and Applications: A Review. *J. Adv. Res.* **2015**, *6*, 106.

(10) Salleh, K. M.; Zakaria, S.; Sajab, M. S.; Gan, S.; Chia, C. H.; Jaafar, S. N. S.; Amran, U. A. Chemically Crosslinked Hydrogel and Its Driving Force towards Superabsorbent Behaviour. *Int. J. Biol. Macromol.* **2018**, *118*, 1422.

(11) Demeter, M.; Virgolici, M.; Vancea, C.; Scarisoreanu, A.; Kaya, M. G. A.; Meltzer, V. Network Structure Studies on γ-Irradiated Collagen-PVP Superabsorbent Hydrogels. *Radiat. Phys. Chem.* **2017**, *131*, 51.

(12) Guilherme, M. R.; Aouada, F. A.; Fajardo, A. R.; Martins, A. F.; Paulino, A. T.; Davi, M. F. T.; Rubira, A. F.; Muniz, E. C. Superabsorbent Hydrogels Based on Polysaccharides for Application in Agriculture as Soil Conditioner and Nutrient Carrier: A Review. *Eur. Polym. J.* **2015**, *72*, 385.

(13) Shah, L. A.; Khan, M.; Javed, R.; Sayed, M.; Khan, M. S.; Khan, A.; Ullah, M. Superabsorbent Polymer Hydrogels with Good Thermal and Mechanical Properties for Removal of Selected Heavy Metal Ions. *J. Cleaner Prod.* **2018**, *201*, 78.

(14) Chen, Y.; Zhang, Y.; Wang, F.; Meng, W.; Yang, X.; Li, P.; Jiang, J.; Tan, H.; Zheng, Y. Preparation of Porous Carboxymethyl Chitosan Grafted Poly (Acrylic Acid) Superabsorbent by Solvent Precipitation and Its Application as a Hemostatic Wound Dressing. *Mater. Sci. Eng. C* **2016**, *63*, 18.

(15) Peng, N.; Wang, Y.; Ye, Q.; Liang, L.; An, Y.; Li, Q.; Chang, C. Biocompatible Cellulose-Based Superabsorbent Hydrogels with Antimicrobial Activity. *Carbohydr. Polym.* **2016**, *137*, 59.

(16) Yang, Y.; Tong, Z.; Geng, Y.; Li, Y.; Zhang, M. Biobased Polymer Composites Derived from Corn Stover and Feather Meals as Double-Coating Materials for Controlled-Release and Water-Retention Urea Fertilizers. *J. Agric. Food Chem.* **2013**, *61*, 8166.

(17) Ranote, S.; Ram, B.; Kumar, D.; Chauhan, G. S.; Joshi, V. Functionalization of Moringa Oleifera Gum for Use as Hg2+ Ions Adsorbent. *J. Environ. Chem. Eng.* **2018**, *6*, 1805.

(18) Bhattacharya, S. B.; Das, A. K.; Banerji, N. Chemical Investigations on the Gum Exudate from Sajna (Moringa Oleifera). *Carbohydr. Res.* **1982**, *102*, 253.

(19) Raja, W.; Bera, K.; Ray, B. Polysaccharides from: Moringa Oleifera Gum: Structural Elements, Interaction with β-Lactoglobulin and Antioxidative Activity. *RSC Adv.* **2016**, *6*, 75699.

(20) Baptista, A. T. A.; Silva, M. O.; Gomes, R. G.; Bergamasco, R.; Vieira, M. F.; Vieira, A. M. S. Protein Fractionation of Seeds of Moringa Oleifera Lam and Its Application in Superficial Water Treatment. *Sep. Purif. Technol.* **2017**, *180*, 114.

(21) Zhang, H.; Xia, H.; Zhao, Y. Poly(Vinyl Alcohol) Hydrogel Can Autonomously Self-Heal. *ACS Macro Lett.* **2012**, *1*, 1233.

(22) Kamoun, E. A.; Kenawy, E.-R. S.; Chen, X. A Review on Polymeric Hydrogel Membranes for Wound Dressing Applications: PVA-Based Hydrogel Dressings. *J. Adv. Res.* **2017**, *8*, 217.

(23) Mansur, H. S.; Oréfice, R. L.; Mansur, A. A. P. Characterization of Poly(Vinyl Alcohol)/Poly(Ethylene Glycol) Hydrogels and PVA-Derived Hybrids by Small-Angle X-Ray Scattering and FTIR Spectroscopy. *Polymer* **2004**, *45*, 7193.

(24) Spoljaric, S.; Salminen, A.; Luong, N. D.; Seppälä, J. Stable, Self-Healing Hydrogels from Nanofibrillated Cellulose, Poly(Vinyl Alcohol) and Borax via Reversible Crosslinking. *Eur. Polym. J.* **2014**, *56*, 105.

(25) Dahlan, N. A.; Pushpamalar, J.; Veeramachineni, A. K.; Muniyandy, S. Smart Hydrogel of Carboxymethyl Cellulose Grafted

Carboxymethyl Polyvinyl Alcohol and Properties Studied for Future Material Applications. *J. Polym. Environ.* **2018**, *26*, 2061.
(26) Dai, H.; Huang, H. Enhanced Swelling and Responsive Properties of Pineapple Peel Carboxymethyl Cellulose-g-poly(acrylic acid-co-acrylamide) Superabsorbent Hydrogel by the Introduction of Carclazyte. *J. Agric. Food Chem.* **2017**, *65*, 565.
(27) Patra, T.; Pal, A.; Dey, J. A Smart Supramolecular Hydrogel ofNα-(4-n-Alkyloxybenzoyl)-l-histidine Exhibiting pH-Modulated Properties. *Langmuir* **2010**, *26*, 7761.
(28) Gharekhani, H.; Olad, A.; Mirmohseni, A.; Bybordi, A. Superabsorbent Hydrogel Made of NaAlg-g-Poly(AA-Co-AAm) and Rice Husk Ash: Synthesis, Characterization, and Swelling Kinetic Studies. *Carbohydr. Polym.* **2017**, *168*, 1.
(29) Chang, C.; He, M.; Zhou, J.; Zhang, L. Swelling Behaviors of PH- and Salt-Responsive Cellulose-Based Hydrogels. *Macromolecules* **2011**, *44*, 1642.
(30) Mahdavinia, G. R.; Zohuriaan-Mehr, M. J.; Pourjavadi, A. Modified Chitosan III, Superabsorbency, Salt- and PH-Sensitivity of Smart Ampholytic Hydrogels from Chitosan-g-PAN. *Polym. Adv. Technol.* **2004**, *15*, 173.
(31) Zhao, Y.; Kang, J.; Tan, T. Salt-, PH- and Temperature-Responsive Semi-Interpenetrating Polymer Network Hydrogel Based on Poly(Aspartic Acid) and Poly(Acrylic Acid). *Polymer* **2006**, *47*, 7702.
(32) Rodrigues, F. H. A.; Fajardo, A. R.; Pereira, A. G. B.; Ricardo, N. M. P. S.; Feitosa, J. P. A.; Muniz, E. C. Chitosan-Graft-Poly(Acrylic Acid)/Rice Husk Ash Based Superabsorbent Hydrogel Composite: Preparation and Characterization. *J. Polym. Res.* **2012**, *19*, 1.
(33) Flory, P. J. *Principles of Polymer Chemistry*; Springer, 1953.
(34) Li, X.; Xu, S.; Wang, J.; Chen, X.; Feng, S. Structure and Characterization of Amphoteric Semi-IPN Hydrogel Based on Cationic Starch. *Carbohydr. Polym.* **2009**, *75*, 688.
(35) Alam, M. N.; Islam, M. S.; Christopher, L. P. Sustainable Production of Cellulose-Based Hydrogels with Superb Absorbing Potential in Physiological Saline. *ACS Omega* **2019**, *4*, 9419.
(36) Chung, J. T.; Vlugt-Wensink, K. D. F.; Hennink, W. E.; Zhang, Z. Effect of Polymerization Conditions on the Network Properties of Dex-HEMA Microspheres and Macro-Hydrogels. *Int. J. Pharm.* **2005**, *288*, 51.
(37) Li, X.; Wu, W.; Wang, J.; Duan, Y. The Swelling Behavior and Network Parameters of Guar Gum/Poly(Acrylic Acid) Semi-Interpenetrating Polymer Network Hydrogels. *Carbohydr. Polym.* **2006**, *66*, 473.
(38) Jhaveri, S. J.; Hynd, M. R.; Dowell-Mesfin, N.; Turner, J. N.; Shain, W.; Ober, C. K. Release of Nerve Growth Factor from HEMA Hydrogel-Coated Substrates and Its Effect on the Differentiation of Neural Cells. *Biomacromolecules* **2009**, *10*, 174.

https://dx.doi.org/10.1021/acsomega.0c01023
ACS Omega XXXX, XXX, XXX−XXX

Int. j. inf. tecnol.
https://doi.org/10.1007/s41870-020-00462-4

ORIGINAL RESEARCH

Multi-dimensional aggregation: a viable solution for interval data

Shailender Kumar[1]

Received: 22 November 2019 / Accepted: 21 April 2020
© Bharati Vidyapeeth's Institute of Computer Applications and Management 2020

Abstract Now a day multi-dimensional data modeling and aggregate query processing which are key assets of business intelligence solutions are being frequently realized to the unorthodox data. For interval values which are recorded when the data is on hold, multidimensional aggregation is the only viable solution and the author emphasizes over this aspect in this paper. Actually, such intervals reflect the state of reality of either current data or such data which were part of the present database. Every possible challenge which interval data throws upon is resolved in this paper through introduction of aggregation operator. Although the intervals are unknown at first but they eventually depend on the actual data and it turns out to be quiet handy while associating them with the resulting tuples. Only those result groups are selected for this purpose, which are specified partially. The interval data signifies that data holds either for each interim in the interval or entire interval and in both of these two cases it faces contention with the operators. In this paper, the author presents the empirical analysis of the aggregation operator after its implementation over the huge industrial data sets and claims that it holds an edge over the other temporal aggregation algorithms.

Keywords Aggregates · Databases · Query processing · Static · Stable

1 Introduction

In most of the present reality database employments, e.g., in the fiscal, medicinal, and systematic spheres, handle time-oriented data, which is raw facts with linked time-phases that holds certain time-varying feature of the facts, classically at the point when the certainties were or is right in the displayed veracity or when the realities was or is section of the contemporary condition of the database. In distinction to this, contemporary database administration delivers valuable tiny integral query language support for time-varying facts administration [1]. In phase through the growing dispersion of professional aptitude, collective calculation turn out to be significant. An aggregate operator transfigures a reflection temporal arrangement into a prompt upshot arrangement [2]. Ordinarily, this is accomplished by principal part the contention transient connection into bunches of lines with undistinguishable esteems for more than one components, then appertain an aggregate procedure, e.g., max or min, to every cluster in one go. For time-traverse esteemed time-arranged databases, total is exceptionally testing on the grounds that the time-traverse esteems can likewise be utilized for portraying the bunches of contention columns [3].

In this paper, the author proposes a novel time-oriented aggregation operator which is employed to perform multi-dimensional aggregation [4]. It improves a decent variety of prior proposed conglomeration operator and gives multi-dimensional help to two periods of total: (a) the elucidation of point assemblies for which to report more than one amassed esteems and (b) the elucidation of amassing assemblies, i.e., gathering of transient contention pushes that are associated with the item bunches and over which the time-situated total methods are assessed [5]. The author constructs the elementary structural chunk on

✉ Shailender Kumar
shailenderkumar@dce.ac.in

[1] Department of Computer Science and Engineering, Delhi Technological University, Delhi, India

Published online: 02 May 2020

contemporary developments in multi-dimensional query process and is the central exertion to dynamism these approaches to phase-esteemed data that are proven by time-oriented facts [6]. The paper covenants a well-organized employment of the aggregation operator with an average density of order $n \log z.$, where z signifies the number of product clusters and n signifies the figure of time-oriented aspects rows [7]. In investigational estimations on huge records, the time-oriented aggregation operator display nearly undeviating performance.

Aggregation of time-oriented facts presents complex trials [8]. Most significantly, the stint-phases of product rows can hinge on the authentic records and are not acknowledged in advancement [9]. As a result, the clustering of the artifact tuples can only be quantified moderately. Following, accumulation should provision what is named persistent, flexible, and fissile semantics of the connotation among records and time-intervals [10]. For instance, the quintessence of a retro juncture with a record adjust is static, hugeness that the adjust grasps for every single sub-periods of the traverse [11]. In deviation, contemplate the association of an exacting month with the numeral of periods operated by a demonstrative amid that period, e.g., 150 periods. The stint-relationship is malleable, by way of 150 periods are well-thoroughly deliberated a total of the periods functioned by the wage earner amid every day amid that month. A stint-relationship is fissionable if the realities can't be connected with changed time–values [12]. All techniques so far offer help for static practices. In conclusion, a period arranged accumulation final product may be more noteworthy than the contention time-situated connection [13]. Explicitly, for instant time-oriented aggregates that are gathered to form a static span, the product stint-relation scope can be double that of the time-oriented reflection [14]. To enumerate the product scope, the paper outlines the concept of a time-oriented aggregation aspect; and to regulate the time-oriented aggregation, the capability to state immobile time-span for the product rows is comprised [15].

The paper is organized in the following manner: Sect. 2 crisply explicates the literatures available in the domain of temporal databases and temporal aggregate operators. Section 3 includes key concepts and basic parameters. In Sect. 4, the author explains and introduce aggregation operator. Section 5 gives a brief overview about the employment of the operator. Section 6 reports on an empirical evaluation and Sect. 7 sum up the things by providing the concision the beneficiation of this article.

2 Background

The notion of hasty as well as amassed time-oriented aggregate operators has been, as of now, expressed by a few authors [16]. The importance of a rapid time-oriented amassed at stint 'st' is assumed by capturing the cluster of rows that confiscation at 'st' [17]. The implication of an amassed time-oriented accumulative at stint 'st' is assumed by capturing the cluster of rows that seizure in the progressive span [st − s, st], s ≥ 0. Misty amassed upshots with uninterrupted stints are amalgamated into static time spans. Most analysis has been improved the situation moment time-arranged totals, e.g. [4, 6], and accumulative aggregates [5, 18] and time-oriented aggregates with auxiliary succession predicates [7, 14] have acknowledged slender contemplation.

An introductory plan by Tuma [8, 19] for presuming time-oriented aggregates entails two examinations of the response stint-oriented schema—one for the assessment of the stint-interims of the item lines and the assessment of the time-arranged totals. A scheme by Kline and Snodgrass [4, 20] look at the info time-arranged connection just a single time, development of an accumulation sequencing tree in principle memory. In the interim, the sequencing tree isn't steady, the most pessimistic scenario time intricacy is $O(n^2)$ where cardinality of a relation is n. A rise, giving under the indistinguishable most pessimistic scenario unpredictability, is the k-ordered conglomeration sequencing tree [21]. A methodology deeds fractional congregation of columns for junk get together of timeworn protuberances.

A steady tree, for collection in principle memory, is grounded on the requested entire sign of time [4, 22]. This elucidation performs for the total capacities like summation, average, and count; for minimum and maximum, a merge-sort like procedures are planned. The two calculations have a most pessimistic scenario. For minor stockpiling, a procedure of bucket calculation is advocated over the information columns to mic-buckets giving to a dividing of the time reflections and furthermore figures out how to pay for enduring lines particular administration. Conglomeration is then accomplished on every single mic-bucket in partition. The calculation includes right to utilize the complete databank thrice.

The disk-grounded calculation, preservation of prompt and cumulative time-oriented conglomeration [6, 23] has already been carried out. A SB-tree encompasses a step of traverses related with modestly figured totals. With the SB-tree, total questions are practical to a complete base assembly—it isn't conceivable to grasp gathering establishes. The multilevel-styled SB-tree [23] emphases on arrangement total explorations amalgamated with go

predicates. The two strategies are focused on to spread grounds, the stint-interim of the info row is erased; thus the outcome isn't a stint situated connection.

The overall strategies portions three aspects. To begin with, the time-arranged gathering process combines the subdividing of the course of events with the array of the info lines [21]. The course of events is isolated into ranges, and an info push fit into an express board if its entire sign of time associate that panel. Second, the item pushes are very much characterized for stint-focuses and not after some stint-interims. Third, they permit the exploitation of at utmost one non-time-situated trait for time-arranged congregation.

Our opportunity arranged collection administrator, which spreads the multi-dimensional crease administrator to engage time-situated total, overpowers these limits and rearranges the conglomeration administrators pondered [3, 16]. It de-couples the isolating of the timetable from the collection of the info columns, consequently allowing to require item pushes over likely covering ranges and to control the degree of the item connection. Furthermore, it bolsters various characteristic highlights [17]. For a viable commercial the creators deed the classification of the info connection.

3 Preliminaries

3.1 Temporal aggregates

The paper commences a divergent stint sphere, where the modules are entitled stints (or stint-facts), well-found with a whole request $<$ T, and the author employs these as our stint-sphere. A whole mark of stint (or stint-facts) is a curled cluster over the stint-sphere and is signified by two stints, $[T_{start}, T_{end}]$, signifying its universal initial and absolute end facts, consistently. The author employs T as shorthand for [T_start, T_end]. For whole mark of time, the author presents numerous schemas: st $\in$ sT represents the stint 'st' is incorporated over the whole mark of stint sT. For two whole mark of stint sT and sT′, sT′ $\subseteq$ sT iff every stints in sT′ are also included in sT, and sT $\cap$ sT′ yields the group of stints over the whole mark of time. If sT $\cap$ sT′ $\neq \varnothing$, the paper states that the binary segments transect.

A time-oriented schema is a multi-tuple Schema = (ψ, ς, dom), where ψ is a determinate regular group of time-oriented attributes, ς is a determinate group of spheres, and sphere: $\rightarrow \varsigma$ is a procedure that links a sphere with every time-oriented attribute. A time-oriented schema is a schema with single whole mark of stint valued attribute (the sphere of whole mark of time corresponds to ς). For the determination of this paper, the author expresses time-oriented stint-relation schemas $V = (w_1, w_2, w_3, \ldots, w_n, \text{Timestamp})$ and $Z = (u_1, u_2, u_3, \ldots, u_n)$. Note that the statement that time-oriented schemas have a whole mark of time attribute. Timestamp is only for appropriateness. There is no concealed time-oriented property, and all clarifications are parameterized with a settled characteristic. Generally, the retitle administrator μ can be utilized to direct schemas as reasonable.

3.2 Attribute characteristics

The paper differentiates amongst three semantics of the stint-relationship of a standard attribute with a whole mark of time attribute. For a time-oriented stint-relation with schema $(w_1, w_2, w_3, \ldots, w_n, Timestamp)$ the attribute features. Timestamp are assumed as $vt = (v_1, v_2, v_3, \ldots, v_n)$, where $v_i \in \{v, m, a\}$. The values v, a, and m represent persistent, flexible, and nuclear features, correspondingly. For example, $vt = (v, m)$ for the schema (X, Y, Timestamp) represents that X has a persistent features and Y has a flexible features. If numerous time-oriented properties are employed, e.g., substantial stint and exchange stint, a standard characteristics enhanced the highlights over more than one stint. The rest of the paper utilizes just a single stint property T, and V denotes to the features associated with sT.

Definition 1 (*Modification of aspect values*). Let $t = (v_1, v_2, v_3, \ldots, v_n|T)$. be a super tuple over time-oriented schema $(w_1, w_2, w_3, \ldots, w_n, \text{Timestamp})$., U be a whole mark of time, and let $Z = (u_1, u_2, u_3, \ldots, u_n)$. be the aspect value features. The alteration in the aspect values is demarcated as streams:

$$\begin{aligned} &\text{adjct}\ (t, U, Z) = (\text{adjct}\ (t.w_1, t.T, U, u_1), \\ &\ldots, \text{adjct}(t.w_n, t.T, U, u_n), U). \end{aligned} \tag{1}$$

Considering the features Z, the adjct procedure alters every standard aspect value of t to the stint-sphere U and produces the tuple which associates T. For example, for the tuple (Apr, 1000, [2004/03, 2004/12]), the features (v, m), and the time-sphere [2004/05, 2004/06] the adjct procedure yields (Apr, 1000, [2004/05, 2004/06]).

4 Multi-dimensional aggregate operator

4.1 Aggregation clusters

A common time-oriented conglomeration operator should handle the dimension of two orthogonal individualities of conglomeration. Categorization of the artifact assemblages

is done on the basis of fallouts in time-oriented aggregate. The clusters of rows which are named as aggregation cluster significantly represents the stint-relationship with product cluster. Every product cluster can be characterized as a row in a time-oriented stint-relation R with schema $(k_1, k_2, k_3, \ldots, k_n|T)$.. Each k_i. is an element of the time-oriented stint-relation which reflects the reflection of the conglomeration operator; tuples undertake standards from k_i. that ensue in the reflection schema. For the whole mark of time, there are two scenarios: static spans and stable spans. With static spans, the whole mark of time attribute undertakes as values the utmost, non-intersecting spans over which the group of reflection rows is persistent. Intervals to aggregate are represented in Fig. 1.

Definition 2 (*Static spans*). Let v be a time-oriented stint-relation with whole mark of time attribute T. The paper describes the static span of v as:

$$\mathrm{SS(v)} = \{\mathrm{sT}|\forall \mathrm{v} \in \mathrm{v}(\mathrm{v.sT} \supseteq \mathrm{sT''} \vee \mathrm{r.sT} \cap \mathrm{sT'} = \emptyset) \wedge \forall \mathrm{sT'} \supset \mathrm{sT}\left(\exists \mathrm{v} \in \mathrm{v}\left(\mathrm{v.sT} \not\supseteq \mathrm{sT''} \wedge \mathrm{r.sT} \cap \mathrm{sT'} \neq \emptyset\right)\right)\Big\}. \tag{2}$$

The leading portion of the equation confirms that product spans do not cross margins of reflection spans and the subsequent portion of the equation confirms that the product spans are uttermost. The static spans for any time-oriented stint-relation are clustered by its sub-group.

Theorem 1 (Cardinality of static spans). *For a time-oriented stint-relation v with m rows, $m > 0$, the cardinality of static spans is partial by the following equation*:

$$|SS(v)| \leq 2m - 1 \tag{3}$$

Proof The end-points of the rows with whole mark of time can be undeviatingly wl-ordered giving at most 2 m time-points. 'm' time-points yields at most m − 1 successive time spans.

Definition 3 (*Stable spans*). Let v be a time-oriented stint-relation with whole mark of time attribute T. The Stable-span(v), has been described as a user-specified group of time-stamps, that fulfills thetandard temporal

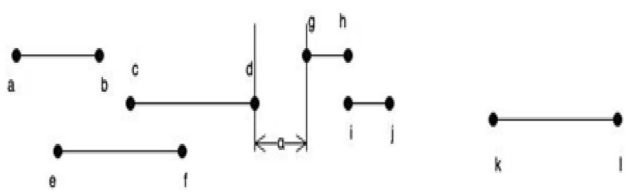


Fig. 1 Intervals to aggregate

statements. These statements describe that stable spans must overlaps with other spans in the time-oriented reflection stint-relation. The categorical ordering of product clusters with stable spans permits control of the cardinality of the artifact schema.

Definition 4 (*Partially quantified product clusters*). A product cluster with schema $(k_1, k_2, k_3, \ldots, k_n)$ is moderately quantified iff the standard of the whole mark of time attribute is not quantified properly. The paper characterizes a partially quantified product row as $v = (k_1, k_2, k_3, \ldots, k_n, [*, *])$.

With partially quantified product clusters in place, the author pushes the conclusion of the product clusters with static spans into the algorithm for the estimation of the time-oriented multi-dimensional conglomeration operator. The incompletely evaluated item pushes are simulated to every single static span for the predictable conglomeration clusters. The cross connection from proclamation that assigns the huge information columns to the static spans is utilized certainly by the estimation of a calculation.

5 Employment of the aggregation operator

5.1 Notion and outline

The work of the collection administrator for static spans is grounded on the various articulations: on the off chance that we analyze the time-situated contention connection, which is very much requested by the span begin time estimations of the lines, whenever point 'pt' assess the item pushes in advance to the point 'pt' (supposing, no columns that begin next to 'pt' added to the item group). Consequently, time-arranged contention connection is being scrutinized, item bunch are framed, and obsolete-columns are also isolated from primary memory. Just the columns those are legitimate at point-stint 'pt'. For the utilization with stable spans, the entire characteristic of time of the item group is measured in the gathering connection. Thus, it isn't compulsory to protect the information pushes in primary memory, however can assess them and refresh the total esteems as we filter the information connection. Data flow of the employment is clearly represented in Fig. 2.

5.2 Algorithm for static span

The algorithm employed for aggregation has five input elements: the cluster stint-relation c, the time-oriented reflection stint-relation t, a list of aggregate procedures $P = \{ p_1, p_2, \ldots, p_n. \}$, a assortment base α, and trait features F. The yield is a period situated connection that incorporates c delayed by the estimations of the total

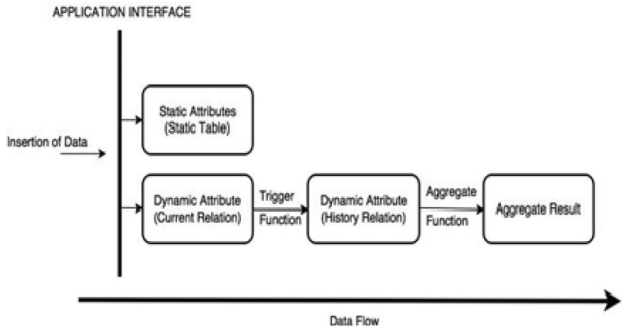


Fig. 2 Data flow

techniques in P. The calculation utilizes two sorts of information assemblies.

The initial step of the calculation is to instate the cluster reflection ct. In the event that c is a proj-reflection over relation, the gathering reflection is at first vacant and will be occupied while scrutinizing the time-situated contention rows. Else, ct is introduced with c, with the begin stint of the records settled to $-\infty$. In conclusion, a record over the non-fleeting properties is created. T succeeding advance is to assess the time-situated contention connection r sequentially on the premise of begin time of the lines. On the off chance that the gathering connection is a social variable based math articulation, we may need to spread the assemblage table with another section before as far as possible all product groups to which information tuple r contributes.

5.3 Algorithm for stable span

The algorithm for the assessment of administrator CT with stable traverse semantics. The key information configuration is the assemblage relation ct that holds the cluster connection c and has a supplementary section named Pi for each time-situated total methods Pi $\in$ P. The item groups, involving their entire sign of time, are entirely measured. The information rows cannot be kept in an end-point assemblage tree.

5.4 Complexity

For the complexity examination of time-oriented aggregate operator, the treatment of the information connection r is proper, which is isolated into four stages: query in the record, manufacture of result lines, refresh of the list and inclusion of the tuple at last point assemblage tree. A refresh list and the query related to the clter reflection have multifaceted nature and order of complexity $\log n_c$, the number of tuples present in the cluster relation is reflected by n_c. The development of a performance outcome push is undeviating as related to the uncovered columns. The support for different trait highlights infers at a cost. For every outcome, it needs a sweep over the entire E-point assemblage tree as well as a change in the characteristic esteems. On the off chance that exclusive static qualities were utilized, the total esteems could be assessed incrementally.

6 Experimental results and analysis

The author executed a number of trials with the static span and stable span algorithms, investigative their presentation for numerous backgrounds. All trials were track on an Intel core i3 with a 1.8 GHz processor and 2 GB memory. The author uses facts stint-schema that comprise from 20,000 to 1, 00,000 entries. The lifecycle of the facts stint-relations is about [0, 2^{20}]. The bunch connection holds one of a kind record. This is a vilest incident since all entire characteristic of time wind up on the comparable E-point assemblage tree. A cluster reflection with extra columns constructed over minor E-point assemblage trees and perform superior. Group partitioning scheme is represented in Fig. 3.

The primary trial examines the versatility of the static spans. The time intricacy depends on the quantity of unclog-tuples and the accumulation encouragement. Connection r-most noticeably awful with the major comprehensible number of unclog-tuples and the most extreme accumulation encouragement for static span has a handling stint that is quadratic in the extent of the information connection. For every other datum gatherings, the preparing stint divulges an undeviating execution. Connection r-measure up to have a collection perspective near 0 despite the fact that the quantity of open tuples is most extreme.

The routine of stable spans isn't misrepresented by crossing columns since the total result is computed augmentation. The indispensable limitation as far as execution is the quantity of upshot assemblages and the efficacy of the current query hone (for every single tuple in r, we should discover all bunches that gratify the standard state

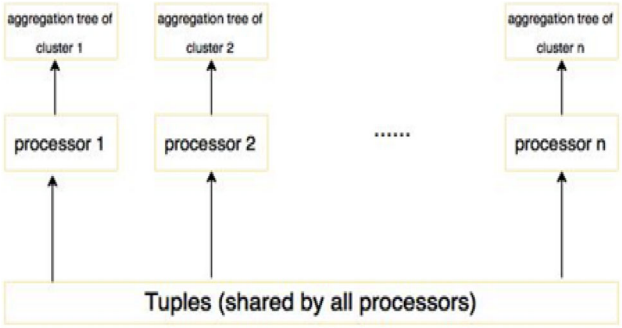


Fig. 3 Group partitioning scheme

and along these lines must be refreshed). Since we utilize AVL-trees for ordering item bunches, the execution increments steeply alongside the quantity of groups. In the event that we utilize hashing, the query time would be stable. In any case, the hashing just backs the fairness conditions.

The execution impact of variable key components on the algorithms utilized by the information connection r-rand is explored. The preparing time of dynamic number of open-tuples is high when contrasted with static one. The execution decays bit by bit since with adaptable properties, all open-tuples must be kept at last point tree. When a static span has been build up, the end-point tree is navigated, the eminence esteems are attuned, and the preceding total strategy is assessed. We have likewise fused the execution of a digressed static spans that backings static characteristics as it were. Static spans are incrementally figures out the totals, and its accomplishment is self-representing the quantity of unclog-tuples as in Fig. 4.

6.1 Static span versus stable span

The result of evaluating aggregates with static span semantics is represented in two ways: (1) with moderately quantified outcome clusters, and (2) a priori calculation of static spans by consuming SQL. The results confirm that static span with moderately quantified outcome clusters is certainly an effective way of evaluating aggregates with static span semantics. The performance of stable span declines as the number of clusters rises as clearly shown in Table 1. Nonetheless, up to an aggregation part of very nearly 46% Stable span outflanks static spans. In this manner, stable spans are effective for sensible aggregation elements, and it licenses to unequivocally control the aggregation factor.

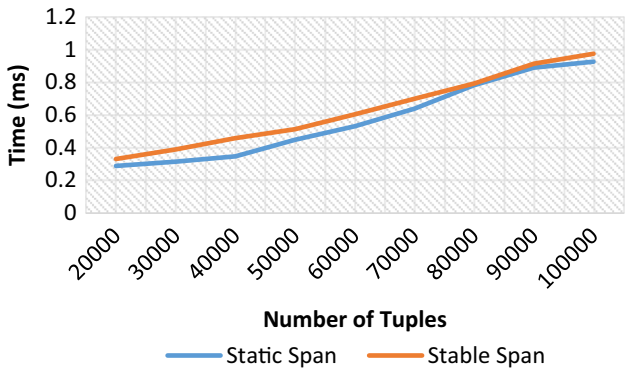


Fig. 4 Performance comparison of static span versus stable span

Table 1 Perfromance comparison of static span vs stable span at each tuple value

Static span		Stable span	
No. of tuples	Time (ms)	No. of tuples	Time (ms)
20,000	0.286	20,000	0.329
30,000	0.312	30,000	0.388
40,000	0.344	40,000	0.457
50,000	0.447	50,000	0.511
60,000	0.529	60,000	0.603
70,000	0.638	70,000	0.698
80,000	0.782	80,000	0.793
90,000	0.889	90,000	0.914
100,000	0.926	100,000	0.975

6.2 Static span versus balanced tree

The very last trial equates static span over the weighted balanced assemblage tree quantified in [3, 5]. The utmost operational procedure come into play, but prompt that it provisions summation, count, or average—it provisions stretchy aspects. Static span supports only static aspects and consents incremental calculation of aggregate standards. While stable span has the similar presentation as the balanced tree algorithm, the trials demonstrate that the provision for numerous stint-aspects features in static span is expensive and the DBMS should recognize circumstances which do not include flexible aspects and stable spans are employed. The recollection investigation of static span hinges on the numeral of unclog-rows.

7 Conclusion and future scope

This paper exhibits a novel time-oriented total administrator that controls most recent headway in multiple dimension query processing to relate with span-esteemed information. The conglomeration administrator streamlines an assorted variety of formerly proposed time-arranged collection administrators. It splits the portrayal of artifact stint-clusters from the explanation of conglomeration assemblages, i.e., the congregations of stint-oriented valued tuples that are linked with the artifact clusters and over which the time-oriented conglomeration procedures are evaluated. This prompts a definite informative system that licenses additionally to direct the extent of the result connection. Consequently, the conglomeration operator provision multiple time-oriented aspect features, embracing stretchy values which embrace aspect standards have to be attuned if the rows time-value altered. Lastly, the author delivers two dissimilar algorithms for the assessment of the

aggregation operator with static span and stable span, correspondingly. Comprehensive trial evaluated over the procedures which are climbable with reverence to facts group extent and associate well with other time-oriented aggregation algorithms. The appraisal additionally exposes that the help for several property determines at a cost. Future work includes various enhancement phases in static and stable span calculations, containing the following ones: improvement procedures for stint-relational algebra expressions that interrelate with the conglomeration administrator, the introduction of the cluster relation on the fly, indexes on the cluster relation, and a variety of the E-point assemblage tree that does not entail a completely arranged information connection.

References

1. Aljawarneh Shadi A, Radhakrishna V, Kumar V, Vinjamuri J (2017) G-SPAMINE: an approach to discover temporal association patterns andtrends in internet of things. Future Gener Comput Syst 74:430–443
2. Carafoli L, Mandreoli F, Martoglia R, Penzo W (2017) Streaming tables: native support to streaming data in DBMSs. IEEE Trans Syst Man Cybern: Syst 47(10):2768–2782
3. Cheng K (2016) Approximate temporal aggregation with nearby coalescing. In: International conference on database and expert systems applications. Springer, Berlin, pp 426–433
4. Ding W, Zhang S, Zhao Z (2017) A collaborative calculation on real-time stream in smart cities. Simul Model Pract Theory 72:23–33
5. Kumar S, Sharma N, Duggal R (2017) Veiled endpoints: a novel dimension for aggregation. Adv Comput Sci Technol 10(5):945–963
6. Dyreson CE (2003) Temporal coalescing with now granularity, and incomplete information. In: Proceedings of the 2003 ACM SIGMOD international conference on Management of data. ACM, pp 169–180
7. Freytag JC, Goodman N (1986) Translating aggregate queries into iterative programs. In: VLDB, vol 86, pp 25–28
8. Gordevičius J, Gamper J, Böhlen M (2012) Parsimonious temporal aggregation. VLDB J Int J Very Large Databases 21(3):309–332
9. Guzzo A, Pugliese A, Rullo N, Sacca D, Piccolo A (2017) Malevolent activity detection with hypergraph-based models. IEEE Trans Knowl Data Eng 29(5):1115–1128
10. Halawani SM, AlBidewi I, Ahmad ABR, Al-Romema NA (2012) Retrieval optimization technique for tuple timestamp historical stint-relation temporal data model. J Comput Sci 8(2):243
11. Kline N, Snodgrass RT (1995) Computing temporal aggregates. In: Data engineering, 1995. Proceedings of the eleventh international conference on, pp 222–231. IEEE
12. Kumar S, Rishi R (2016) Retrieval of meteorological data using temporal data modeling. Ind J Sci Technol 9(37):1–10
13. Kvet M, Matiako K, Kvet M (2016) Managing and storing function results in temporal approach. In: Open and big data (OBD), international conference on, pp 87–94. IEEE
14. Lopez IFV, Snodgrass RT, Moon B (2005) Spatiotemporal aggregate computation: a survey. IEEE Trans Knowl Data Eng 17(2):271–286
15. Marcellino M (1999) Some consequences of temporal aggregation in empirical analysis. J Bus Econ Stat 17(1):129–136
16. Moon B, Lopez IFV, Immanuel V (2003) Efficient algorithms for large-scale temporal aggregation. IEEE Trans Knowl Data Eng 15(3):744–759
17. Navathe SB, Ahmed R (1989) A temporal stint-relational model and a query language. Inf Sci 49(1–3):147–175
18. Snodgrass RT, Gomez S, McKenzie LE (1993) Aggregates in the temporal query language TQuel. IEEE Trans Knowl Data Eng 5(5):826–842
19. Terenziani Paolo (2016) Irregular indeterminate repeated facts in temporal stint-relational databases. IEEE Trans Knowl Data Eng 28(4):1075–1079
20. Yang J, Widom J (2001) Incremental computation and maintenance of temporal aggregates. In: Data engineering, 2001. Proceedings 17th international conference on. IEEE, pp 51–60
21. Zhang D, Markowetz A, Tsotras VJ, Gunopulos D, Seeger B (2008) On computing temporal aggregates with range predicates. ACM Trans Database Syst 33(2):12
22. Zhou X, Wang F, Zaniolo C (2006) Efficient temporal coalescing query support in stint-relational database systems. In: International conference on database and expert systems applications. Springer, Berlin, pp 676-686
23. Kumar S, Rishi R (2016) A relative analysis of modern temporal data models. In: Computing for sustainable global development (INDIACom), 2016 3rd international conference on. IEEE, pp 2851–2855

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Node Classification in Complex Networks using Network Embedding Techniques

Sanjay Kumar
Computer Science and Engineering
Delhi Technological University
Delhi,India
Sanjay.kumar@dtu.ac.in

Sanidhya Chaudhary
Computer Science and Engineering
Delhi Technological University
Delhi,India
sanidhya.chaudhary97@gmail.com

Saksham Kumar
Computer Science and Engineering
Delhi Technological University
Delhi,India
saksham.kmr.21@gmail.com

Raj Kumar Yadav
Computer Science and Engineering
Delhi Technological University
Delhi,India
rajkumar.raj.yadav.2015@gmail.com

***Abstract*—Node classification is a technique to assign node labels using the network topology and various node features. Many techniques like random walk-based method, label propagation graph regularizations are proposed in this research. Network embedding is a popular method to capture graph structure in low dimensional vector space and can capture the features of the nodes properly. In this paper, the popular embedding techniques are visualized on various data-sets and introduce the idea of network embedding techniques to perform the node classification task.**

***Index Terms*—Label propagation, Network embedding, Node classification, Online Social Networks, Random walk.**

I. INTRODUCTION

Node classification [1,2,3] is one of the most popular techniques, which is aimed to determine the labels of the vertices (nodes) of the network according to the other labelled vertices (nodes) and the structure of network graph. Many times, in a network, only a part of total nodes is labelled. There can be many reasons behind the unknown labels for a larger number of nodes so it becomes extremely essential to classify these nodes to acquire knowledge about their characteristics and behaviour. The labelled nodes and the links can be used to infer the missing labels.

The Node Classification problem can be stated as: providing a social network (or any other network structure) where some of the nodes are labeled and rest unlabeled, in what way estimate a good quality labeling for each node of the network. This problem is called the "node classification problem", fundamental problem being providing labeling for each node in a network.

Problem Statement:

Given a graph **G (V, E, W)**

V is the set of all nodes in the graph (—V— = n)

E is the set of edges in the graph

W is the weight matrix of the graph

Identify applicable funding agency here. If none, delete this.

Let **Vl** represent the labelled subset of V

Vu = (V - Vl) represents the set of unlabelled nodes

Let **Y** represent the set of all possible labels (be m)

Let **Yl** be the set of initial labels given on nodes of set Vl. To solve the problem, our method uses the information already available in the partly labeled network to predict labels for the rest of the nodes. Our goal is to train the classifier using the nodes that are pre labeled. Then the trained classifier is used to label the unlabeled nodes and also to nodes that already have some labels to augment their existing labels.

II. NETWORK OR GRAPH EMBEDDING

Transforming graph properties into vector space is Graph Embedding. Embedding must capture the geometric structure of graph, its vertex-to-vertex connection, and all other appropriate information about the graph and its vertices.

A. Methods used for Graph Embedding

1) DeepWalk[9]

DeepWalk is a random-walk based network embedding procedure. Random walks begin from a chosen node then moving to random neighbors at each step for a defined number of times. DeepWalk consists of 3 basic steps:

a) Sampling (Sampling of graph with random walks)

b) Training the skip-gram network (Skip-Gram network accepts node as one-hot-vector and maximizes the chances of predicting its neighbor nodes)

c) Computing the Embeddings (the hidden layer of the trained network is used as the embedding)

2) LINE [10]

Large-scale Information Network Embedding (LINE) defines two basic functions for first and second order proximity. These are explained below:

a) First Order Proximity refers to similarity between two nodes if they are connected by an edge.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:51:46 UTC from IEEE Xplore. Restrictions apply.

b) Second Order Proximity refers to similarity of two nodes' neighborhood's structure.

3) Node2Vec [11]

Node2Vec is similar to DeepWalk with the addition of biasing in random walks over the discovery of larger neighborhood or discovery of microscopic view around the node. Two parameters P and Q were introduced to control the probability of returning to the previous node and discovering the undiscovered part of the network respectively. All other steps (Training, computing Embeddings) is same as DeepWalk.

4) SDNE [12]

Structural Deep Network Embedding (SDNE) does not use random walks and is designed to preserve the first and second-order proximity on nodes in the embeddings.

SDNE uses an autoencoder neural net consisting of two parts Two (left and right) Vanilla auto encoders which accepts node adjacency vectors and learn the second order proximity.

5) Struc2Vec [13]

Struc2Vec embeds a network using node's structural identity. The idea is to construct a meta-graph which encodes the structural information of the graph of interest, and then use random walks from every node in this meta-graph to determine the contexts.

Struc2Vec has four basic steps:

a) Determine the structural similarity between every pair of nodes of the network, over different neighborhood sizes.

b) Construct a weighted multi layered graph where each layer of this graph relates to a level in hierarchy determining the structural similarity.

c) Use this multi layered graph to generate context for every node in the network based on biased random walks.

d) Use standard techniques to learn a latent representation from this context.

B. Graph Embedding Visualization

1) DeepWalk Visualization

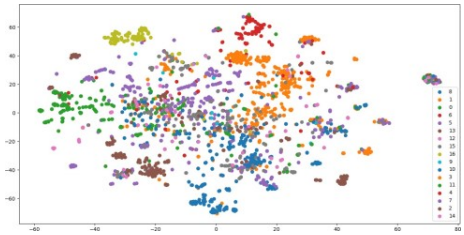


Fig 1. DeepWalk on wiki Dataset.

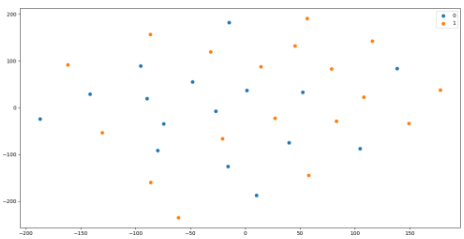


Fig 2. DeepWalk on karate Dataset.

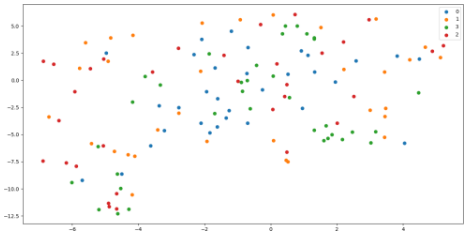


Fig 3. DeepWalk on Flight(Brazil) Dataset.

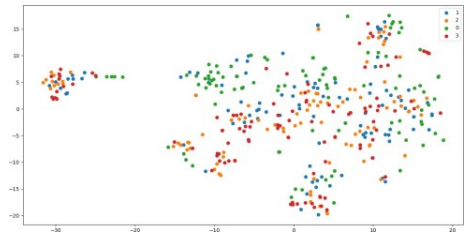


Fig 4. DeepWalk on Flight(Europe) Dataset.

2) LINE Visualization

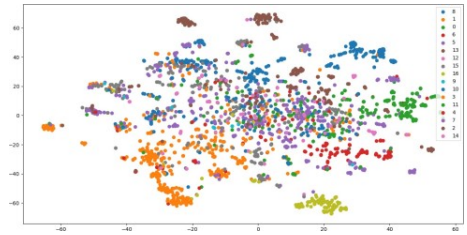


Fig 5. LINE on wiki Dataset.

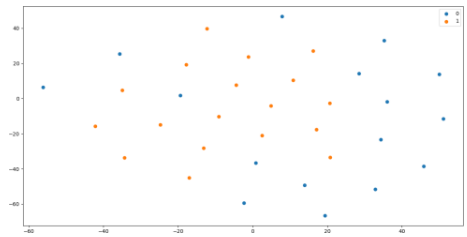
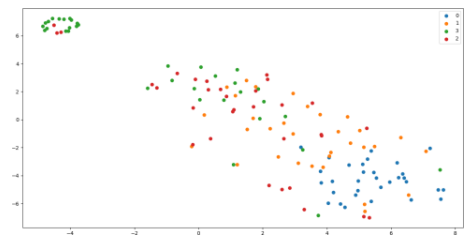


Fig 6. LINE on karate Dataset.



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:51:46 UTC from IEEE Xplore. Restrictions apply.

Fig 7. LINE on Flight(Brazil) Dataset.

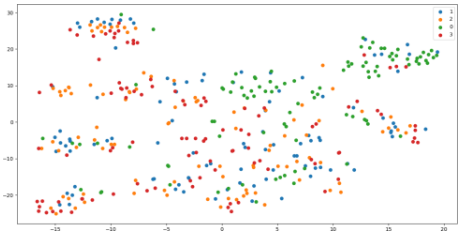


Fig 8. LINE on Flight(Europe) Dataset.

3) Node2Vec Visualization

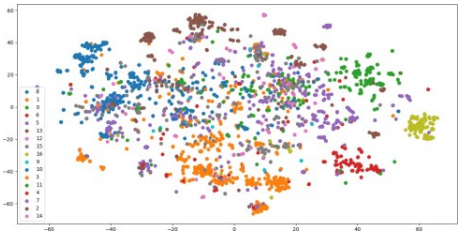


Fig 9. Node2Vec on wiki Dataset.

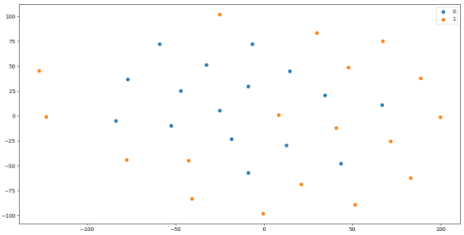


Fig 10. Node2Vec on karate Dataset.

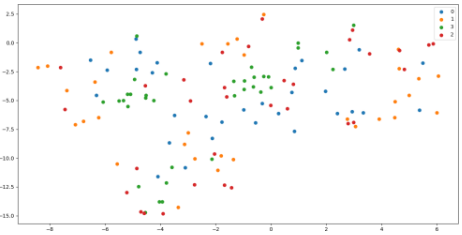


Fig 11. Node2Vec on Flight(Brazil) Dataset.

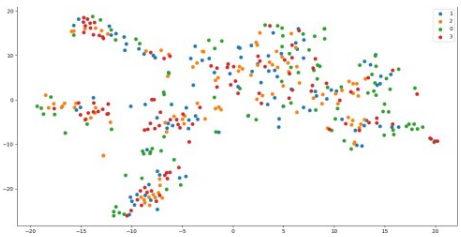


Fig 12. Node2Vec on Flight(Europe) Dataset.

4) SDNE Visualization

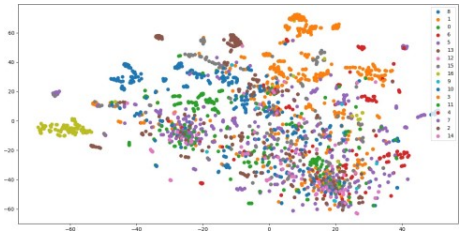


Fig 13. SDNE on wiki Dataset.

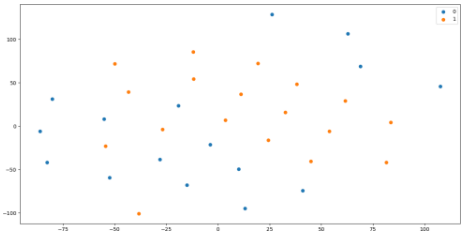


Fig 14. SDNE on karate Dataset.

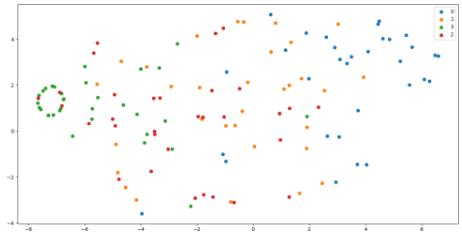


Fig 15. SDNE on Flight(Brazil) Dataset.

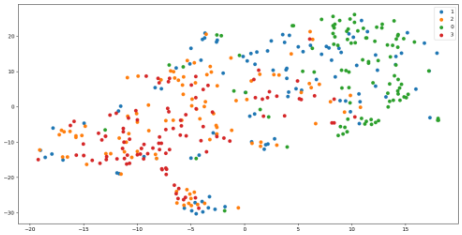


Fig 16. SDNE on Flight(Europe) Dataset.

5) Struc2Vec Visualization

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:51:46 UTC from IEEE Xplore. Restrictions apply.

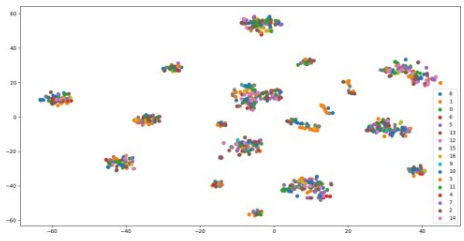


Fig 1. Struc2Vec on wiki Dataset.

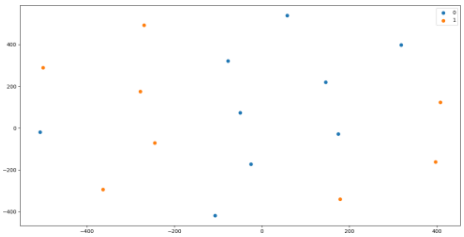


Fig 1. Struc2Vec on karate Dataset.

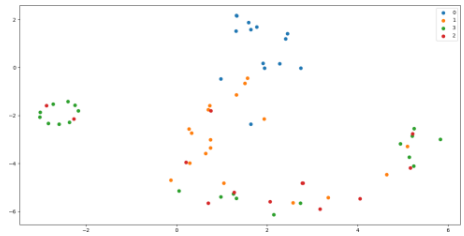


Fig 1. Struc2Vec on Flight(Brazil) Dataset.

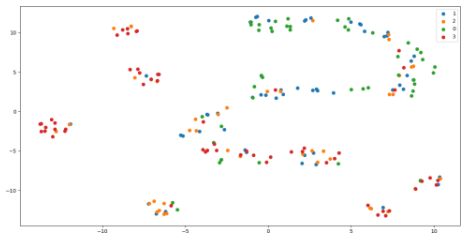


Fig 1. Struc2Vec on Flight(Europe) Dataset.

III. METHODOLOGY

A. Algorithm

Using the idea of network embedding techniques followed by label propagation and regression for classification and a proper mixture of test and training data, an algorithm for node classification is proposed which works as follows:

Input: Graph (G)
Output: Accuracy Score, F1_score

NodeClassification(G)
G ← DiGraph(edge_list)
model ← EmbeddingTechnique(G)
model.train(window_size, iter)
embeddings ← model.get_embeddings()
for embedding in embeddings
evaluate_embedding(embedding)
plot_embedding(embedding)
evaluate_embeddings(embedding):
X, Y ← read_node_labels(label _list)
tr_frac ← 0.8
clf ← classifier(embedding, clf = LogisticRegression())
clf.evaluate(Y_test, Y_predicted, trac_frac)
label_propagation(G)
label_prop_model←labelPropagation()
label_prop_model.fit(X_train,y_train)
prediction ← label_prop_model.predict(X_test)
evaluate(Y_test, prediction)

The graph is given as an input to the algorithm in step 2, which is converted to lower dimensions using different graph embedding techniques in step 3 and 4. Step 4 trains the model to evaluate embeddings. Plot_embedding is used to plot the reduced vector form in pictorial form. Evaluate_embeddings function in step 9 is used to classify nodes by logistic regression in which tr_fact is the fraction of pre-labelled dataset that is the number of nodes which are already labelled. Step 13 evaluates the predicted labels and gives accuracy and f1_score as output. Similarly, in function 'label_propagation' (step 14), 80% nodes are used to predict the labels of remaining nodes and again the predicted labels are evaluated in step 18.

A directed graph is constructed using the edge list for various datasets like karate, wiki etc and have used 80% of the nodes for training our model and then used it to classify the remaining 20% nodes using various methods. Initially, classified using label propagation in which each node checks for the labels of its neighbors and then adopts the label with highest frequency.

Then tried to classify nodes using feature- based methods and then applying logistic regression to classify them. In non-embedded graph the neighbor list is considered as the features of the node, however in embedded graph the dimensionality is used a feature list and then after the again using 80% of the classified nodes have tried to classify the rest 20% of the nodes. In order to classify using logistic regression, used OneVsAll classifier in which one classifier is fitted per class. Interpretability is the main advantage of this classifier as the inspection of classifier will give the knowledge of the class as each class is represented by one classifier.

A comprehensive comparison of various classification techniques is shown and suitability of different embedding techniques for different datasets and have also proved the scalability problem of old classifying methods as the size of dataset increases the gap between the evaluation score of embedded and non-embedded graph also increases substantially, however the evaluation score was quite compa-rable in smaller datasets.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:51:46 UTC from IEEE Xplore. Restrictions apply.

B. Evaluation Metrices

- **Accuracy score:** In classification, this function computes subset accuracy: the set of labels predicted for a sample must exactly match the corresponding set of labels in y true i.e. true labels.
- **F1 score** F-score or F-measure are the equivalent terms of F1 score. The weighted average of the recall and precision is interpreted as F1 score. The best value of F1_ score is 1 whereas worse is 0. Equal weightage is given precision and recall. Mathematically, it can be stated as:

F1 = (recall * precision) * 2 / (recall + precision)

a) Micro Metrics are calculated over the global domain by counting the total number of false negatives, false positives and true positives.
b) Macro Metrics are calculated for every label and their unweighted mean is found. The label imbalance is not taken into account

IV. Result

Evaluation and comparisons of node classification using both non-embedded and embedded graphs. The graph was embedded using different techniques mentioned above in Section 2.1. The metrics mentioned in Section 3.2 are used to evaluate and compare the results obtained.

Application of Label Propagation on various datasets without the use of any embedding techniques produced results shown in Table 1. The values obtained are one of the lowest showing limitations of Label Propagation on big datasets. Using Logistic Regression on non-embedded graphs shows slight improvement over previous Label Propagation, shown in Table 2.Various graph embedding techniques are used in the proposed algorithm and results are obtained showing improvements over previous approaches. Results with graph embeddings are shown from Table 3 through Table 8.

Traditional methods of node classification may not scale well with the increasing size of the network. This is also evident by observing the value of F1_score and accuracy below as they are comparable in smaller datasets such as karate and flight but as the size of the dataset increases the F1_score and accuracy of embedded graphs is much better than that of non-embedded graphs like in wiki data set, classification using logistic regression gave 32.6% better results when applied on embedded graph(LINE embedding technique) than non-embedded graph by considering neighbors as features. Classification of wiki data set using logistic regression on embedded graph gave 86.4% better results than label propagation on non-embedded graph. Similarly, on comparing the struc2vec embedding of flight (Brazil) dataset and found that the accuracy score is 61.36% better than that of non-embedded one.

TABLE I
Using Label Propagation on non-Embedded Graphs

DATASET	*MICRO*	*MACRO*	*ACCURACY*
Wiki	0.3775	0.13189	0.3776
Karate	0.3775	0.13189	0.3776
Flight(Brazil)	0.3775	0.13189	0.3776
Flight(Europe)	0.3775	0.13189	0.3776

TABLE II
Using Logistic Regression on non-Embedded Graph where neighbours are considered as features

DATASET	*MICRO*	*MACRO*	*ACCURACY*
Wiki	0.5218	0.3560	0.5218
Karate	0.5714	0.3636	0.5714
Flight(Brazil)	0.3529	0.1666	0.4444
Flight(Europe)	0.2702	0.2941	0.4875

TABLE III
Results obtained using DeepWalk for graph embedding

DATASET	*MICRO*	*MACRO*	*ACCURACY*
Wiki	0.6902	0.6070	0.6902
Karate	0.4285	0.3000	0.4285
Flight(Brazil)	0.1481	0.1491	0.1481
Flight(Europe)	0.4750	0.4706	0.4750

TABLE IV
Results obtained using LINE for graph embedding

DATASET	*MICRO*	*MACRO*	*ACCURACY*
Wiki	0.6299	0.5023	0.6299
Karate	0.4285	0.3000	0.4285
Flight(Brazil)	0.5185	0.4495	0.5185
Flight(Europe)	0.4375	0.4351	0.4375

TABLE V
Results obtained using Node2Vec for graph embedding

DATASET	*MICRO*	*MACRO*	*ACCURACY*
Wiki	0.6881	0.5945	0.6881
Karate	0.4285	0.3000	0.4286
Flight(Brazil)	0.3704	0.3216	0.3703
Flight(Europe)	0.4000	0.3983	0.4000

TABLE VI
Results obtained using SDNE for graph embedding

DATASET	*MICRO*	*MACRO*	*ACCURACY*
Wiki	0.6216	0.4859	0.6216
Karate	0.8571	0.8571	0.8571
Flight(Brazil)	0.5925	0.5527	0.5925
Flight(Europe)	0.5750	0.5660	0.5750

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:51:46 UTC from IEEE Xplore. Restrictions apply.

TABLE VII
RESULTS OBTAINED USING STRUC2VEC FOR GRAPH EMBEDDING

DATASET	*MICRO*	*MACRO*	*ACCURACY*
Wiki	0.1950	0.1100	0.1950
Karate	0.7500	0.7333	0.7500
Flight(Brazil)	0.7143	0.7222	0.7143
Flight(Europe)	0.4500	0.4428	0.4500

CONCLUSION AND FUTURE WORK

In this paper, various approaches to node classification techniques are analyzed using label propagation on non-embedded graphs and logistic regression on both embedded and non-embedded graphs. Node classification using logistic regression on embedded graphs gives better results than both logistic regression and label propagation on non-embedded graphs. Various embedding techniques work best for different data sets because of their methods of feature generation i.e., dimensionality. For example, struc2vec works best on flight networks which are identified by the network structure however LINE, SDNE and node2vec gave better results on large data sets such as wiki dataset. The results obtained after embedding are quite significant and give commendable results. The embedding techniques can be further used for link prediction i.e., predicting the missing links which can be used for building a recommendation engine, Graph reconstruction. The node classification efficiency can be further enhanced using a neural network, which can greatly capture the characteristic dynamics of the network, by training the network data-sets.

REFERENCES

[1] Goyal, P., Ferrara, E.: Graph embedding techniques, applications, and performance: A survey. Knowledge-Based Systems, 151, 78-94.(2018).
[2] Bhagat, S., Cormode, G., Muthukrishnan, S. : Node classification in social networks. In Social network data analytics (pp. 115-148). Springer, Boston, MA. (2011).
[3] Tang, J., Aggarwal, C., Liu, H.: Node classification in signed social networks. In Proceedings of the 2016 SIAM international conference on data mining (pp. 54-62). Society for Industrial and Applied Mathematics. (2016, June).
[4] Grover, A., Zweig, A., Ermon, S.: Graphite: Iterative generative modeling of graphs. arXiv preprint arXiv:1803.10459. (2018)
[5] Zhang, X. K., Ren, J., Song, C., Jia, J., Zhang, Q. : Label propagation algorithm for community detection based on node importance and label influence. Physics Letters A, 381(33), 2691-2698. (2017).
[6] Zhu, X., Ghahramani, Z.: Learning from labeled and unlabeled data with label propagation.(2002).
[7] Jokar, E., Mosleh, M.: Community detection in social networks based on improved Label Propagation Algorithm and balanced link density. Physics Letters A, 383(8), 718-727. (2019).
[8] Zhang, T., Popescul, A., Dom, B.: Linear prediction models with graph regularization for web-page categorization. In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 821-826). (2006, August).
[9] Perozzi, B., Al-Rfou, R., Skiena, S.: Deepwalk: Online learning of social representations. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 701-710).(2014).
[10] Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., Mei, Q.: Line: Large-scale information network embedding. In Proceedings of the 24th international conference on world wide web (pp. 1067-1077).(2015, May).
[11] Grover, A., Leskovec, J.node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 855-864).(2016, August).
[12] Ribeiro, L. F., Saverese, P. H., Figueiredo, D. R.: struc2vec: Learning node representations from structural identity. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 385-394). (2017, August).
[13] Ribeiro, L. F., Saverese, P. H., Figueiredo, D. R.: struc2vec: Learning node representations from structural identity. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 385-394). (2017, August).
[14] Kipf, T. N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907. (2016).
[15] Wang, D., Cui, P., Zhu, W: Structural deep network embedding. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1225-1234).(2016, August).
[16] Sun, G., Zhang, X.: Graph Embedding with Rich Information through Bipartite Heterogeneous Network. CoRR. (2017).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:51:46 UTC from IEEE Xplore. Restrictions apply.

Nonlinear Dyn
https://doi.org/10.1007/s11071-020-05762-9

ORIGINAL PAPER

Nonlinear dynamics of a time-delayed epidemic model with two explicit aware classes, saturated incidences, and treatment

Kanica Goel · Abhishek Kumar · Nilam

Received: 28 April 2020 / Accepted: 10 June 2020
© Springer Nature B.V. 2020

Abstract Whenever a disease emerges, awareness in susceptibles prompts them to take preventive measures, which influence individuals' behaviors. Therefore, we present and analyze a time-delayed epidemic model in which class of susceptible individuals is divided into three subclasses: unaware susceptibles, fully aware susceptibles, and partially aware susceptibles to the disease, respectively, which emphasizes to consider three explicit incidences. The saturated type of incidence rates and treatment rate of infectives are deliberated herein. The mathematical analysis shows that the model has two equilibria: disease-free and endemic. We derive the basic reproduction number R_0 of the model and study the stability behavior of the model at both disease-free and endemic equilibria. Through analysis, it is demonstrated that the disease-free equilibrium is locally asymptotically stable when $R_0 < 1$, unstable when $R_0 > 1$, and linearly neutrally stable when $R_0 = 1$ for the time delay $\varrho > 0$. Further, an undelayed epidemic model is studied when $R_0 = 1$, which reveals that the model exhibits forward and backward bifurcations under specific conditions, which also has important implications in the study of disease transmission dynamics. Moreover, we investigate the stability behavior of the endemic equilibrium and show that Hopf bifurcation occurs near endemic equilibrium when we choose time delay as a bifurcation parameter. Lastly, numerical simulations are performed in support of our analytical results.

Keywords Full and partial awareness · Time delay · Nonlinear incidences and treatment rates · Bifurcations · Stability · Numerical simulations

Mathematics Subject Classification 34D20 · 92B05 · 37M05

K. Goel · Nilam (✉)
Department of Applied Mathematics, Delhi Technological University, Delhi 110042, India
e-mail: nilam.research18@gmail.com; rathi.nilam@gmail.com

K. Goel
e-mail: kanica.dtu2016@gmail.com

A. Kumar
Department of Mathematics, School of Basic Sciences and Research, Sharda University, Greater Noida 201310, India
e-mail: abhishek.dtu14@gmail.com; abhishek.kumar6@sharda.ac.in

1 Introduction

The last two decades have seen several large-scale epidemics outbreaks such as Ebola, SARS, Zika virus, and swine flu, which leads to low socioeconomic status and inadequate access to health care. People get information about these outbreaks quite quickly due to significant advances in social media, which can have an insightful effect on the actual epidemic dynamics [17,31]. Therefore, at the beginning of an epidemic outbreak, the initial step is to make the individuals aware of the disease and its preventive methods. Awareness programs can alert the susceptible population toward the

Published online: 04 July 2020

contagious disease [23]. If susceptible individuals have adequate knowledge and proper information about the infection, then they take precautionary measures such as regular hand sanitization, the use of face masks, wear hand gloves, vaccination, and even quarantine, which can reduce the impact of illness, for instance, ranging from the plague outbreak in the English village of Eyam in 1665–1666 [9], where the town completely sealed itself off to prevent further transmission of plague, to more recent outbreaks of swine influenza [17] and Ebola [31]. In the situation of rural health services, the challenge of government health-care scheme is that there are numerous gaps in primary health services, and the health-care facilities are mainly urban centric. The rural residents may have low health knowledge awareness, and their receiving way of health knowledge can be traditional and straightforward. Knowledge level increases with higher education levels; therefore, one of the main factors is their education level. Therefore, a lot of people are not fully but partially aware of the spread and control of infectious diseases. Due to partial awareness, some individuals often medicate themselves adopting antibiotics, even when advised against doing so which weakens their immune system makes them at a high risk of catching the infection [18]. Therefore, complete awareness about the cycle of disease, along with the utilization of appropriate precautions and adequate decontamination procedures, is vital. Full awareness of the disease in humans develops a habit of taking precautionary measures against it, and they follow the instructions given by health workers and government and lower their risk of becoming infected.

In the study of the transmission of infectious diseases, the incidence rate has a vital role as it determines the number of infectives per unit of time. In 1927, Kermack Mckendrick considered the bilinear incidence rate, which follows the law of mass action [1]. In the bilinear incidence rate, the number of infectives linearly increases, which might be real for a small population of infected individuals, but it is impractical for a large number of infectives. Therefore, several studies are devoted to consider nonlinear incidence rate for disease transmission dynamics (e.g., [2,3,15,16,26,34–40]. In the present study, we have incorporated the nonlinear saturated incidence rate, where the interaction term is of the form $Sg(I)$, $g(I) = \beta I/(1 + \varepsilon I)$. Here, βI measures the force of infection, and $1/(1 + \varepsilon I)$ measures the inhibition effect of the behavioral change of the susceptibles when their number increases and crowding effect of the infective individuals [2,22]. In this incidence rate, the number of infectives depends on the nonlinear bounded map $g(I)$, which tends to saturation level β/ε when I gets large. It prevents the unboundedness of the contact rate as it includes the psychological effects and, thus, more realistic than the bilinear incidence rate.

Treatment is vital to cure the infection and prevent the development of resistant bacteria. Therefore, consideration of the treatment rate in the epidemic model is of great importance. In 2004, Wang and Ruan [10] studied the SIR epidemic model with the constant treatment rate and showed various bifurcations. The consideration of constant treatment rate might be real when there are small infected populations because there are limited medical resources available in society. Therefore, in 2012, Zhou and Fan [24] improved the treatment rate by considering Holling type II functional response as given below:

$$h(I) = \frac{\beta I}{1 + \alpha I}, \quad I \geq 0, \quad \beta \geq 0, \quad \alpha \geq 0.$$

and explore the SIR epidemic model to understand the effect of the limited medical resources and their supply efficiency on the transmission of infectious diseases. Motivated by the work mentioned above, we deliberate the Holling type II treatment rate (also called saturated treatment rate) and study the effect of treatment rate with limited medical resources in the current epidemic model.

The inclusion of time delay in the study of epidemiology is an important aspect. Persons with asymptomatic infections play an essential role in the spread of infectious disease, especially as they are unaware of their infection and therefore take no special hygiene precautions. Thus, the study of epidemiology involves time delay, which needs to be considered for practical purposes. The inclusion of time delay may arise due to delays caused by the latency in a vector and delay caused by a latent period in the host [4]. Recently, many mathematical models studied the impact of time delay in their epidemic model [7,13,35–37,40].

The impact of information and awareness on the spread of epidemics has been studied by many authors [14,19–21,27–29,33,34]. Funk et al. [19] considered the aware susceptible and aware infected populations in their epidemic model, on the assumption that the

aware susceptibles will be at a lower rate of catching the infection than the unaware susceptibles, and studied that disease dynamics in a well-mixed population. Kiss et al. [20] studied the effect of information transmission on the dynamics of sexually transmitted diseases. They assumed that the entire population is aware of the risk of infection; however, only a specific portion decides to react by constraining their contact with contaminated people. Misra et al. and Dubey et al. [21,34] investigated the impact of awareness programs on the transmission dynamics of infectious diseases in their nonlinear epidemic models. Some researchers studied the impact of awareness on the disease transmission dynamics along with the influence of time delay. Zuo et al. [29] introduced a time delay in the media variable to emphasize the delay in reporting cases of infections. Zhao et al. [27] studied the SIRS epidemic model by incorporating time delay in media coverage. Zuo and Liu [28] studied the effect of awareness programs driven by media and the delay on the prevalence of the infectious disease. In all these models, the study reveals that the disease-free equilibrium (DFE) is stable when a basic reproduction number R_0 is less than one, unstable when $R_0 > 1$, irrespective of the value of the time delay, and has a stable endemic equilibrium for time delay equals to zero. In the study of Zuo et al. [29], Zhao et al. [27], and Misra et al. [21], the occurrence of Hopf bifurcation is shown for the particular value of the time delay. Greenhalgh et al. [33] included two delays, one in reporting of infected cases and another delay representing the loss of disease awareness after a fixed period. Their study reveals the reduction in infected individuals with an increment in the duration of awareness. It is also shown that both the time delays can produce oscillations and destabilize the endemic equilibrium.

The public is a coalition of many subgroups of individuals with vastly different social, educational, and economic backgrounds. During an outbreak, people adopt full or partial awareness, depending on the understanding of how they perceive risks, and communicate about the effectiveness of protective measures. Therefore, instead of going directly from susceptible to infected class, we introduce fully aware and partial aware susceptible compartments into the SIR epidemic model due to heterogeneous protection level and extend the epidemic model to include the behavioral change of susceptibles, which can change the transmission patterns and reduce the prevalence of disease to the more extent. The precaution level of susceptible individuals is heterogeneous as they may take different levels of precautions to protect themselves from being infected based on the severity of epidemics or their characteristics. We consider three specific nonlinear incidence rates of unaware susceptibles, fully aware susceptibles, and partially aware susceptible, respectively, with the inclusion of time delay as a latent phase having a fixed duration. Also, with awareness, treatment to infectives is essential to mitigate the infection. Therefore, we consider the nonlinear saturated treatment rate, which includes the fact of limitation in the availability of resources. We formulate the nonlinear time-delayed mathematical epidemic model and perform the stability analysis to demonstrate the eradication or persistence of the disease with the help of the basic reproduction number R_0 for the time delay $\varrho \geq 0$. The bifurcation theory approach using the center manifold theory is performed, revealing the existence of backward and forward bifurcations, which shows that reducing R_0 below unity is not sufficient to eradicate the disease due to the presence of backward bifurcation, which has an important implication in the study of disease dynamics. Further, choosing time delay as a bifurcation parameter, the periodic and oscillatory solutions appear via Hopf bifurcation. Moreover, the numerical experiments show the importance of considering full and aware susceptible individuals and nonlinear terms such as time delay, incidences, and treatment.

The rest of the manuscript is organized as follows: The model and its basic properties are presented in Sect. 2. In Sect. 3, the existence of disease-free and endemic equilibria is presented, and stability analysis is performed using the basic reproduction number. The conditions for the backward and forward bifurcations are derived. Also, the existence of stability switches is shown via the presence of Hopf bifurcation by choosing the time delay as a bifurcation parameter. In Sect. 4, numerical simulations are presented to signify the theoretical results. Lastly, Sect. 5 is devoted to a brief discussion.

2 Mathematical model and basic properties

We assume that N denotes the total constant population, and it is divided into five compartments according to the disease status: unaware susceptible class $S(t)$, fully aware class $A_F(t)$, partially aware class $A_P(t)$, infected individuals class $I(t)$, and removed individu-

als class $R(t)$. The unaware susceptible class consists of those individuals who are vulnerable to the disease and taking no precautions against it. A fully aware susceptible class involves those individuals who have adequate knowledge and proper information about the spread of the disease and taking precautionary measures against it. Partially aware susceptibles consist of those individuals who have an incomplete understanding of the spread and prevention of the disease and have low resources available to escape from the infection. Including a behavioral response among unaware susceptibles would unavoidably call for response among fully aware susceptibles and partially aware susceptibles. Therefore, we have considered three explicit same functional types of saturated incidence rates described below:

(i) $F_1(S(t-\varrho), I(t-\varrho)) = (\beta_1 S(t-\varrho) I(t-\varrho))/(1+\varepsilon I(t-\varrho))$, representing the saturated incidence rate among unaware susceptibles,

(ii) $F_2(A_F(t-\varrho), I(t-\varrho)) = (\beta_2 A_F(t-\varrho) I(t-\varrho))/(1+\varepsilon I(t-\varrho))$, representing the saturated incidence rate among fully aware susceptibles,

(iii) $F_3(A_P(t-\varrho), I(t-\varrho)) = (\beta_2 A_F(t-\varrho) I(t-\varrho))/(1+\varepsilon I(t-\varrho))$, representing the saturated incidence rate among partially aware susceptibles,

where ϱ denotes the time delay, representing the latent period. The latent period is defined as the period between exposure and infection, since the pathogen is present in a latent stage, without clinical symptoms or signs of infection in the host. The delay term in $S(t)$, $A_F(t)$, and $A_P(t)$ is introduced because people, who are unaware, fully aware, or partially aware of the disease may consider themselves in their respective classes after becoming infected. We assume that such individuals are in the latent period.

Let κ denote the constant rate of inflow of new unaware susceptibles due to the recruitment of new members by the current members or immigration. Let δ_1 denote the rate at which unaware susceptibles $S(t)$ become fully aware. δ_2 denotes the rate at which $S(t)$ becomes partially aware of the disease. The parameters β_1, β_2, and β_3 are the transmission rates of infection of unaware, fully aware, and partially aware susceptible populations, respectively. We suppose that all the susceptible classes can become infected by contact with infected individuals, but the fully aware class has less chance to be infected as compared to the unaware susceptible and partially aware susceptible individuals [33]. Therefore, it is assumed that β_1 and β_3 are greater β_2. The parameter ε represents the behavioral changes or measures of inhibition adopted by infectives. ϑ denotes the natural mortality rate of all individuals, whereas d and θ denote the disease-induced mortality rate and recovery rate of the infectives, respectively. The function $h(I(t)) = aI/(1+bI)$ represents the treatment rate of infectives, where a and b denote the cure rate and limitation rate in the availability of resources, respectively. The symbols and description of the parameters and state variables are given in Table 1 briefly.

The flow diagram of the proposed epidemic model is given in Fig. 1, and model is represented mathematically under the following system of delay differential equations:

$$\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= \kappa - \delta_1 S - \delta_2 S - \frac{\beta_1 S(t-\varrho) I(t-\varrho)}{1+\varepsilon I(t-\varrho)} - \vartheta S,\\
\frac{\mathrm{d}A_F}{\mathrm{d}t} &= \delta_1 S - \frac{\beta_2 I(t-\varrho) A_F(t-\varrho)}{1+\varepsilon I(t-\varrho)} - \vartheta A_F,\\
\frac{\mathrm{d}A_P}{\mathrm{d}t} &= \delta_2 S - \frac{\beta_3 I(t-\varrho) A_P(t-\varrho)}{1+\varepsilon I(t-\varrho)} - \vartheta A_P,\\
\frac{\mathrm{d}I}{\mathrm{d}t} &= \frac{\beta_1 S(t-\varrho) I(t-\varrho)}{1+\varepsilon I(t-\varrho)} + \frac{\beta_2 I(t-\varrho) A_F(t-\varrho)}{1+\varepsilon I(t-\varrho)}\\
&\quad + \frac{\beta_3 I A_P(t-\varrho)}{1+\varepsilon I(t-\varrho)} - (d+\theta+\vartheta) I - \frac{aI}{1+bI},\\
\frac{\mathrm{d}R}{\mathrm{d}t} &= \theta I + \frac{aI}{1+bI} - \vartheta R.
\end{aligned} \tag{1}$$

Suppose that the initial conditions of the model (1) take the form

$$\begin{aligned}
&S(\nu) = \chi_1(\nu),\ A_F(\nu) = \chi_2(\nu),\ A_P(\nu) = \chi_3(\nu),\\
&I(\nu) = \chi_4(\nu),\ R(\nu) = \chi_5(\nu),\\
&\chi_i(\nu).\ \nu \in [-\varrho,\ 0],\ \chi_i(0) > 0\ (i = 1, 2, 3, 4, 5),
\end{aligned} \tag{2}$$

where $(\chi_1(\nu), \chi_2(\nu), \chi_3(\nu), \chi_4(\nu), \chi_5(\nu)) \in C([-\varrho, 0], \mathbb{R}^5_+)$, which is the Banach space of continuous functions mapping the interval $[-\varrho, 0]$ into $\mathbb{R}^5_+$, where

$$\begin{aligned}
\mathbb{R}^5_+ = \{&(S, A_F, A_P, I, R)\,/S \geq 0, A_F \geq 0,\\
&A_P \geq 0, I \geq 0, R \geq 0\}.
\end{aligned}$$

Using Proposition 2.3 in [32] and Proposition 2.1 in [25], it follows that all state variables of the model (1) are nonnegative, i.e., $(S, A_F, A_P, I, R) \in \mathbb{R}^5_+$. For

Table 1 Notations of model variables and parameters

Symbol	Description
N	Total constant population
$S(t)$	Unaware susceptibles
$A_F(t)$	Fully aware susceptibles
$A_P(t)$	Partially aware susceptibles
$I(t)$	Infected population
$R(t)$	Removed population
ϱ	Latent period (time delay)
κ	Constant recruitment rate of unaware susceptibles
β_1	Transmission rate of susceptibles to infected individuals
β_2	Transmission rate of fully aware to infected individuals
β_3	Transmission rate of partially aware to infected individuals
ε	Inhibition measures by infectives
δ_1	Rate of full awareness in unaware susceptibles
δ_2	Rate of partial awareness in unaware susceptibles
ϑ	Natural death rate
d	Disease-induced death rate
θ	Recovery rate
a	Cure rate
b	Rate of limitations in treatment availability

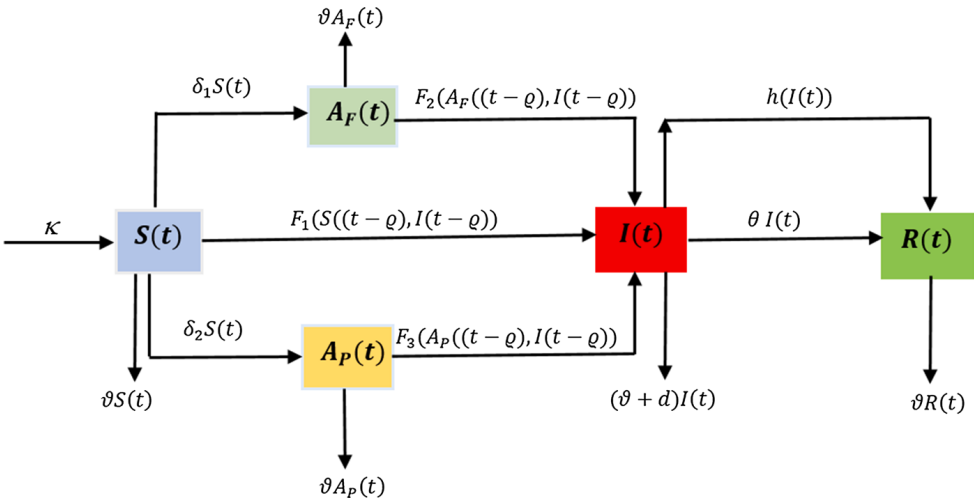


Fig. 1 Flow diagram of the model (1)

biological reasons, we assume that the model parameters κ, δ_1, δ_2 β_1, β_2, β_3, ε, ϑ, θ, d, a, b are positive. By the fundamental theory of functional differential equations [6], the model (1) has a unique solution $(S(t)\, A_F(t), A_P(t), I(t), R(t))$ satisfying the initial conditions (2).

Lemma 1 *The compact set*

$$\Omega = \left\{(S(t), A_F(t), A_P(t), I(t), R(t)) \in \mathbb{R}^5_+ : S(t) + A_F(t) + A_P(t) + I(t) + R(t) \leq \frac{\kappa}{\vartheta}\right\}$$

is invariant for the solutions of the model (1).

Proof Since the right-hand side of the model (1) and its derivatives are continuous, it assures the well-posedness of the model (1).

On adding the equations of the model (1), we get

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\left(S(t)+A_F(t)+A_P(t)+I(t)+R(t)\right)\\&=\kappa-\vartheta\left(S(t)+A_F(t)+A_P(t)\right.\\&\quad\left.+I(t)+R(t)\right)-dI\\&\leq\kappa-\vartheta\left(S(t)+A_F(t)+A_P(t)+I(t)+R(t)\right).\end{aligned}\tag{3}$$

Thus, the invariant region for the existence of the solutions is given as

$$0<\lim_{t\to\infty}\left(S(t)+A_F(t)+A_P(t)+I(t)+R(t)\right)\leq\frac{\kappa}{\vartheta}.\tag{4}$$

Hence, the solution of the model (1) is compact. □

In the next section, we obtain the model's equilibria and perform stability analysis to investigate the behavior of the equilibrium points.

3 Mathematical Analysis

Since $R(t)$ does not appear in the first four equations of the model (1), without loss of generality, it is sufficient to consider first four equations of the model for the analysis purpose:

$$\begin{aligned}\frac{\mathrm{d}S}{\mathrm{d}t}&=\kappa-\delta_1S-\delta_2S-\frac{\beta_1S(t-\varrho)I(t-\varrho)}{1+\varepsilon I(t-\varrho)}-\vartheta S,\\\frac{\mathrm{d}A_F}{\mathrm{d}t}&=\delta_1S-\frac{\beta_2I(t-\varrho)A_F(t-\varrho)}{1+\varepsilon I(t-\varrho)}-\vartheta A_F,\\\frac{\mathrm{d}A_P}{\mathrm{d}t}&=\delta_2S-\frac{\beta_3I(t-\varrho)A_P(t-\varrho)}{1+\varepsilon I(t-\varrho)}-\vartheta A_P,\\\frac{\mathrm{d}I}{\mathrm{d}t}&=\frac{\beta_1S(t-\varrho)I(t-\varrho)}{1+\varepsilon I(t-\varrho)}+\frac{\beta_2I(t-\varrho)A_F(t-\varrho)}{1+\varepsilon I(t-\varrho)}\\&\quad+\frac{\beta_3IA_P(t-\varrho)}{1+\varepsilon I(t-\varrho)}-(d+\theta+\vartheta)I-\frac{aI}{1+bI}.\end{aligned}\tag{5}$$

For the existence of model equilibrium, we set the right-hand side terms of the (5) to zero and obtain that the model (5) has two equilibria given below:

1. Disease-free equilibrium $E_0(S_0, A_{F_0}, A_{P_0}, 0)$ (represents the eradication of infected individuals, i.e., $I\equiv 0$ for all $t>0$), discussed in Sect. 3.1.
2. Endemic equilibrium $E_e(S^*, A_F^*, A_P^*, I^*)$ (represents the persistence of infected individuals above a certain positive level, i.e., $I>0$ for all $t>0$), discussed in Sect. 3.2.

3.1 Disease-free equilibrium

The system (5) has a disease-free equilibrium of the form $E_0(S_0, A_{F_0}, A_{P_0}, 0)$, where

$$S_0=\frac{\kappa}{\delta_1+\delta_2+\vartheta},\ A_{F_0}=\frac{\delta_1\kappa}{\vartheta(\delta_1+\delta_2+\vartheta)},$$
$$A_{P_0}=\frac{\delta_2\kappa}{\vartheta(\delta_1+\delta_2+\vartheta)}.$$

The characteristic equation of the system (5) at disease-free equilibrium is obtained as

$$\begin{aligned}&(-\vartheta-\lambda)^2\left(-\delta_1-\delta_2-\vartheta-\lambda\right)\left(\lambda+(\vartheta+d+\theta+a)\right.\\&\left.-\left(\beta_1S_0+\beta_2A_{F_0}+\beta_3A_{P_0}\right)\mathrm{e}^{-\lambda\varrho}\right)=0.\end{aligned}\tag{6}$$

Equation (6) is fourth-degree transcendental equation in λ with roots $\lambda_1=-\vartheta$, $\lambda_2=-\vartheta$, $\lambda_3=-(\delta_1+\delta_2+\vartheta)$, and the other roots are the solution of

$$\begin{aligned}&\lambda+(\vartheta+d+\theta+a)\\&-\left(\beta_1S_0+\beta_2A_{F_0}+\beta_3A_{P_0}\right)\mathrm{e}^{-\lambda\varrho}=0.\end{aligned}\tag{7}$$

On simplification, Eq. (7) can be written as

$$\begin{aligned}X(\lambda):=\lambda+(\vartheta+d+\theta+a)\\-\frac{\kappa\left(\beta_1\vartheta+\beta_2\delta_1+\beta_3\delta_2\right)}{\vartheta\left(\delta_1+\delta_2+\vartheta\right)}\mathrm{e}^{-\lambda\varrho}=0.\end{aligned}\tag{8}$$

The term $\frac{\kappa(\beta_1\vartheta+\beta_2\delta_1+\beta_3\delta_2)}{\vartheta(\vartheta+d+\theta+a)(\delta_1+\delta_2+\vartheta)}\mathrm{e}^{-\lambda\varrho}$ at $\varrho=0$ is defined as the basic reproduction number of the system (5). Thus, the basic reproduction number of the system (5) is

$$R_0=\frac{\kappa\left(\beta_1\vartheta+\beta_2\delta_1+\beta_3\delta_2\right)}{\vartheta\left(\vartheta+d+\theta+a\right)\left(\delta_1+\delta_2+\vartheta\right)}.$$

The basic reproduction number R_0 is the average number of secondary infections caused by a single infected agent, during his/her entire infectious period, in an entirely susceptible population [8].

3.1.1 Analysis for $\mathbf{R_0\neq 1}$

As mentioned above, Eq. (6) has three negative real roots $\lambda_1=-\vartheta$, $\lambda_2=-\vartheta$, $\lambda_3=-(\delta_1+\delta_2+\vartheta)$, and other roots are the solutions of Eq. (8).

Note that

$$X(0)=(\vartheta+d+\theta+a)\,(1-R_0).\tag{9}$$

If $R_0 > 1$, then $X(0) < 0$. Also,

$$X'(\lambda) = 1 + \frac{\kappa\,(\beta_1\vartheta + \beta_2\delta_1 + \beta_3\delta_2)}{\vartheta\,(\delta_1 + \delta_2 + \vartheta)}\mathrm{e}^{-\lambda\varrho}\varrho > 0. \quad (10)$$

Thus, $\lim_{\lambda\to\infty} X(\lambda) = +\infty$.

Hence, $X(\lambda) = 0$, $X(0) < 0$, $X'(\lambda) > 0$, and $\lim_{\lambda\to\infty} X(\lambda) = +\infty$ indicate that there always exists a unique and positive real root of $X(\lambda) = 0$ if $R_0 > 1$.

Now, if $R_0 < 1$, then

$$\begin{aligned}\mathrm{Re}\ \lambda &= \frac{\kappa\,(\beta_1\vartheta + \beta_2\delta_1 + \beta_3\delta_2)\,\mathrm{e}^{-(\mathrm{Re}\ \lambda)\varrho}\cos(Im\lambda)\varrho}{\vartheta\,(\delta_1 + \delta_2 + \vartheta)} \\ &\quad - (\vartheta + d + \theta + a) \\ &< \frac{\kappa\,(\beta_1\vartheta + \beta_2\delta_1 + \beta_3\delta_2)}{\vartheta\,(\delta_1 + \delta_2 + \vartheta)} - (\vartheta + d + \theta + a) \\ &= (\vartheta + d + \theta + a)\,(R_0 - 1) \\ &< 0. \end{aligned} \quad (11)$$

Therefore, $R_0 < 1$ implies that λ is a root of Eq. (6) with negative real part.

3.1.2 Analysis at $\mathbf{R_0 = 1}$

In this subsection, the system (5) is analyzed at $E_0(S_0, A_{F_0}, A_{P_0}, 0)$ and $R_0 = 1$ for the time delay $\varrho \geq 0$.

Case (i) $\varrho > 0$

If $R_0 = 1$, then $\lambda = 0$ is a simple characteristic root of Eq. (8). Note that $R_0 = 1$ implies that $\kappa\,(\beta_1\vartheta + \beta_2\delta_1 + \beta_3\delta_2) = \vartheta\,(\vartheta + d + \theta + a)\,(\delta_1 + \delta_2 + \vartheta)$.

Let $\lambda = r + is$ be any of the other solutions, then (8) becomes:

$$\begin{aligned} &r + \mathrm{i}s + (\vartheta + d + \theta + a) \\ &\quad - (\vartheta + d + \theta + a)\,\mathrm{e}^{-r\varrho}\,(\cos s\varrho - \mathrm{i}\sin s\varrho) = 0. \end{aligned} \quad (12)$$

By applying Euler's formula and splitting real and imaginary parts, Eq. (12) can be written as

$$\begin{aligned} r + \vartheta + d + \theta + a &= (\vartheta + d + \theta + a)\cos s\varrho\ \mathrm{e}^{-r\varrho}, \\ s &= -(\vartheta + d + \theta + a)\sin s\varrho\ \mathrm{e}^{-r\varrho}. \end{aligned} \quad (13)$$

On squaring and adding both the equations of Eq. (13), we obtain

$$\begin{aligned} &(r + \vartheta + d + \theta + a)^2 + s^2 \\ &\quad = (r + \vartheta + d + \theta + a)^2\,\mathrm{e}^{-2r\varrho}. \end{aligned} \quad (14)$$

If there exists a root satisfying both the equations of (13), then this root also satisfies Eq. (14) obtained by squaring and adding these equations. For Eq. (14) to be verified, we must have $r \leq 0$. Therefore, E_0 is linearly neutrally stable.

Case (ii) $\varrho = 0$

We now analyze the qualitative behavior of the system (5) without time delay, i.e., we take $\varrho = 0$. This analysis has an interest in itself and will also allow getting some information on the stability of the coexistence equilibrium in the case with delay [30]. The bifurcation theory approach obtained in [11], which is based on the center manifold theory [5], has been used to decide the local stability of a nonhyperbolic equilibrium (i.e., linearization matrix has at least one eigenvalue with zero real part) near the criticality E_0 and $R_0 = 1$. It allows us to clarify the direction of the bifurcation and describes the local behavior of disease-free equilibrium near $R_0 = 1$. For simplicity, we redefine the state variables as $S = x_1$, $A_F = x_2$, $A_P = x_3$, and $I = x_4$. So, the system (5) takes the form:

$$\begin{aligned} \frac{\mathrm{d}x_1}{\mathrm{d}t} &= \kappa - \delta_1 x_1 - \delta_2 x_1 - \frac{\beta_1 x_1 x_4}{\varepsilon x_4 + 1} - x_1\vartheta \equiv g_1, \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} &= -\frac{\beta_2 x_2 x_4}{\varepsilon x_4 + 1} - \vartheta x_2 + \delta_1 x_1 \equiv g_2, \\ \frac{\mathrm{d}x_3}{\mathrm{d}t} &= -\frac{\beta_3 x_3 x_4}{\varepsilon x_4 + 1} - \vartheta x_3 + \delta_2 x_1 \equiv g_3, \\ \frac{\mathrm{d}x_4}{\mathrm{d}t} &= \frac{\beta_1 x_1 x_4}{\varepsilon x_4 + 1} + \frac{\beta_2 x_2 x_4}{\varepsilon x_4 + 1} + \frac{\beta_3 x_3 x_4}{\varepsilon x_4 + 1} \\ &\quad - (d + \theta + \vartheta)x_4 - \frac{a x_4}{b x_4 + 1} \equiv g_4. \end{aligned} \quad (15)$$

We observe that when $R_0 = 1$, the chosen bifurcation parameter β_2 takes the form

$$\begin{aligned} \beta_2 &= \beta_2^* \\ &= \frac{\delta_2\,(\vartheta(a + d + \theta + \vartheta) - \beta_3\kappa) + \vartheta\,(\delta_1 + \vartheta)\,(a + d + \theta + \vartheta) + \beta_1\kappa(-\vartheta)}{\delta_1\kappa}. \end{aligned}$$

The Jacobian matrix $J(E_0,\ \beta_2^*)$ of the system (15) obtained at criticality (that is, at E_0 and β_2^*) is

$$J(E_0, \beta_2^*) = \begin{pmatrix} -\vartheta - \delta_1 - \delta_2 & 0 & 0 & -\frac{\kappa\beta_1}{\vartheta + \delta_1 + \delta_2} \\ \delta_1 & -\vartheta & 0 & -\frac{\kappa\beta_2^*\delta_1}{\vartheta\,(\vartheta + \delta_1 + \delta_2)} \\ \delta_2 & 0 & -\vartheta & -\frac{\kappa\beta_3\delta_2}{\vartheta\,(\vartheta + \delta_1 + \delta_2)} \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The eigenvalues of the Jacobian matrix $J(E_0, \beta_2^*)$ are $\lambda_1 = -\vartheta,\ \lambda_2 = -\vartheta,\ \lambda_3 = -\vartheta - \delta_1 - \delta_2,\ \lambda_4 = 0$. We see that λ_1, λ_2, and λ_3 are real and negative eigenvalues and $\lambda_4 = 0$ is a simple zero eigenvalue (since the algebraic multiplicity of $\lambda_4 = 0$ is 1) of $J(E_0, \beta_2^*)$. Thus, when $R_0 = 1$, the DFE E_0 is a non-hyperbolic equilibrium.

The right eigenvector $v = (v_1, v_2, v_3, v_4)$ of $J(E_0, \beta_2^*)$ corresponding to $\lambda_4 = 0$ is obtained as

$$
\begin{aligned}
v_1 &= -\frac{\beta_1 \kappa}{(\delta_1 + \delta_2 + \vartheta)^2},\\
v_2 &= -\frac{\delta_1 \kappa\,(\beta_2\,(\delta_1 + \delta_2 + \vartheta) + \beta_1 \vartheta)}{\vartheta^2\,(\delta_1 + \delta_2 + \vartheta)^2},\\
v_3 &= -\frac{\delta_2 \kappa\,(\beta_3\,(\delta_1 + \delta_2 + \vartheta) + \beta_1 \vartheta)}{\vartheta^2\,(\delta_1 + \delta_2 + \vartheta)^2},\\
v_4 &= 1.
\end{aligned}
$$

The left eigenvector $w = (w_1, w_2, w_3, w_4)$ of $J(E_0,\ \beta_2^*)$ corresponding to $\lambda_4 = 0$ is obtained as

$$
\begin{aligned}
w_1 &= 0,\\
w_2 &= 0,\\
w_3 &= 0,\\
w_4 &= 1.
\end{aligned}
$$

Let g_k's, $k = 1, 2, 3, 4$, denote the right-hand side of the system (15). The bifurcation coefficients a_1 and b_1 defined in Theorem 4.1 of [33] are given by:

$$
\begin{aligned}
a_1 &= \sum_{k,i,j=1}^{4} w_k v_i v_j \left(\frac{\partial^2 g_k}{\partial x_i \partial x_j}\right)_{E_0},\\
b_1 &= \sum_{k,i=1}^{4} w_k v_i \left(\frac{\partial^2 g_k}{\partial x_i \partial \beta_2^*}\right)_{E_0}.
\end{aligned}
$$

The nonzero partial derivatives of the functions g_k's at E_0 are evaluated as

$$
\begin{aligned}
\left(\frac{\partial^2 g_4}{\partial x_4 \partial x_1}\right)_{E_0} &= \beta_1, \left(\frac{\partial^2 g_4}{\partial x_4 \partial x_2}\right)_{E_0} = \beta_2^*,\\
\left(\frac{\partial^2 g_4}{\partial x_4 \partial x_3}\right)_{E_0} &= \beta_3, \left(\frac{\partial^2 g_4}{\partial x_1 \partial x_4}\right)_{E_0} = \beta_1,\\
\left(\frac{\partial^2 g_4}{\partial x_2 \partial x_4}\right)_{E_0} &= \beta_2^*,\\
\left(\frac{\partial^2 g_4}{\partial x_3 \partial x_4}\right)_{E_0} &= \beta_3, \left(\frac{\partial^2 g_4}{\partial x_4^2}\right)_{E_0}\\
&= -\frac{2\varepsilon\beta_1\kappa}{\delta_1 + \delta_2 + \vartheta} - \frac{2\varepsilon\beta_2^*\delta_1\kappa}{\vartheta\,(\delta_1 + \delta_2 + \vartheta)}\\
&\quad -\frac{2\varepsilon\beta_3\delta_2\kappa}{\vartheta\,(\delta_1 + \delta_2 + \vartheta)} + 2ab, \textit{and}\\
\left(\frac{\partial^2 g_4}{\partial x_4 \partial \beta_2^*}\right)_{E_0} &= \frac{\delta_1\kappa}{\vartheta\,(\delta_1 + \delta_2 + \vartheta)}.
\end{aligned}
$$

The bifurcation coefficients a_1 and b_1 are calculated at the bifurcation parameter β_2^* as follows:

$$
\begin{aligned}
a_1 &= -\frac{2}{\delta_1\kappa\vartheta^2\,(\delta_1 + \delta_2 + \vartheta)}\Big(\delta_1(\delta_2(\vartheta^2\,(\kappa(a(\varepsilon - b)\\
&\quad +\varepsilon(d + \theta + \vartheta)) + 2(a + d + \theta + \vartheta)^2\Big)\\
&\quad + \beta_3\kappa\,(\beta_3\kappa - 2\vartheta(a + d + \theta + \vartheta)))\\
&\quad + \vartheta^3\,(\kappa(a(\varepsilon - b) + \varepsilon(d + \theta + \vartheta))\\
&\quad +2(a + d + \theta + \vartheta)^2\Big))\\
&\quad + \delta_1^2\vartheta^2\,(\kappa(a(\varepsilon - b) + \varepsilon(d + \theta + \vartheta))\\
&\quad +(a + d + \theta + \vartheta)^2\Big)\\
&\quad + (\delta_2\,(\vartheta(a + d + \theta + \vartheta) - \beta_3\kappa)\\
&\quad + \vartheta^2(a + d + \theta + \vartheta))^2\\
&\quad - \beta_1\kappa\vartheta\,(2\delta_2\,(\vartheta(a + d + \theta + \vartheta) - \beta_3\kappa)\\
&\quad +\vartheta\,(\delta_1 + 2\vartheta)\,(a + d + \theta + \vartheta))\\
&\quad + \beta_1^2\kappa^2\vartheta^2\Big)\\
&= -\frac{2}{\delta_1\kappa\vartheta^2\,(\delta_1 + \delta_2 + \vartheta)} \times G(\beta_3),\\
b_1 &= \frac{\kappa}{\delta_1 + \delta_2 + \vartheta}.
\end{aligned}
$$

where

$$
G(\beta_3) = \left(\zeta_2\,\beta_3^2 + \zeta_1\,\beta_3 + \zeta_0\right), \tag{16}
$$

where the coefficients $\zeta_0,\ \zeta_1$ and ζ_2 are

$$
\begin{aligned}
\zeta_2 &= \delta_2\,(\delta_1 + \delta_2)\,\kappa^2,\\
\zeta_1 &= 2\delta_2\kappa\vartheta\,(\beta_1\kappa - (\delta_1 + \delta_2 + \vartheta)\,(a + d + \theta + \vartheta)),\\
\zeta_0 &= \vartheta^2\Big(\big(\delta_1 + \delta_2 + \vartheta\big)\big(\delta_1\big(\kappa(a(\varepsilon - b)\\
&\quad + \varepsilon(d + \theta + \vartheta))\\
&\quad + (a + d + \theta + \vartheta)^2\big)
\end{aligned}
$$

$$+\big(\delta_2+\vartheta\big)(a+d+\theta+\vartheta)^2\big)$$
$$+\beta_1\delta_1\kappa(a+d+\theta+\vartheta)$$
$$+\beta_1\kappa\big(\beta_1\kappa-2\big(\delta_1+\delta_2+\vartheta\big)(a+d+\theta+\vartheta)\big)\big). \tag{17}$$

It can be seen that the bifurcation coefficient b_1 is always positive and the sign of a_1 depends the sign of $G(\beta_3)$, given in Eq. (16). If $G(\beta_3) < 0$, then $a_1 > 0$, whereas if $G(\beta_3) > 0$, then $a_1 < 0$.

The discriminant of quadratic polynomial $G(\beta_3)$ is obtained as

$$D = a\big(\delta_1+\delta_2\big)\kappa(b-\varepsilon)\big(\delta_1+\delta_2+\vartheta\big)$$
$$+\big(-\delta_1-\delta_2-\vartheta\big)\big(\big(\delta_1+\delta_2+\vartheta\big)(a+d+\theta+\vartheta)^2$$
$$+\varepsilon\big(\delta_1+\delta_2\big)\kappa(d+\theta+\vartheta)\big)$$
$$+\beta_1\kappa\big(\delta_1+\delta_2+2\vartheta\big)(a+d+\theta+\vartheta)$$
$$+\beta_1^2\big(-\kappa^2\big). \tag{18}$$

Let β_3^* and β_3^{**} be two roots of Eq. (16), then we get

$$\beta_3^* = \frac{-\zeta_1+D}{2\zeta_2}, \text{ and } \beta_3^{**} = \frac{-\zeta_1-D}{2\zeta_2}. \tag{19}$$

Using Theorem 4.1(iv) in [11], the type of bifurcation is governed by the sign of a_1 and hence by the sign of $G(\beta_3)$. If $G(\beta_3)$ is of positive sign, then forward bifurcation occurs, whereas if the sign of $G(\beta_3)$ is negative, then the system (15) reveals a backward bifurcation. The study of forward and backward bifurcation is important in the study of epidemiology as they play a relevant role in disease control and eradication. In the case of forward bifurcation, the condition $R_0 < 1$ is a necessary and sufficient condition for disease eradication, whereas reducing R_0 below unity is no longer sufficient for disease eradication when a backward bifurcation occurs. These behavior differences are essential in planning how to control a disease; a backward bifurcation at $R_0 = 1$ makes control more difficult. These two cases are discussed below separately.

(I) Forward bifurcation: When there is a forward bifurcation at $R_0 = 1$, it is not possible for a disease to invade a population if $R_0 < 1$ because the system will return to the disease-free equilibrium $I = 0$ if some infectives are introduced into the population. For values of R_0 slightly greater than 1, E_0 changes its stability from stable to unstable and the model admits a unique endemic equilibrium, which is locally asymptotically stable [11]. Therefore, it is imperative to find the range for which forward bifurcation occurs. The range of forward bifurcation is governed by the positivity of $G(\beta_3)$. Thus, there are two cases in which $G(\beta_3)$ is found to be positive. These are given as follows:

$$\begin{cases} \zeta_1 > 0, \\ \zeta_0 > 0. \end{cases} \tag{20}$$

or

$$\begin{cases} D > 0, \\ \text{either of } (\zeta_1 > 0,\ \zeta_0 < 0),\ (\zeta_1 < 0.\ \zeta_0 > 0),\ \text{or } (\zeta_1 < 0,\ \zeta_0 < 0) \text{ holds}, \\ \beta_3 < \beta_3^* \text{ or } \beta_3 > \beta_3^{**}. \end{cases} \tag{21}$$

(II) Backward bifurcation: The backward bifurcation is characterized as when $R_0 < 1$; a small unstable endemic equilibrium appears while the disease-free equilibrium and a larger endemic equilibrium are locally asymptotically stable. When $R_0 > 1$, then an unstable disease-free equilibrium and a stable endemic equilibrium exist [11]. It is illustrated in Fig. (4). The range of existence of backward bifurcation (i.e., when $G(\beta_3) < 0$) as follows:

$$\begin{cases} D > 0, \\ \text{either of } (\zeta_1 > 0,\ \zeta_0 < 0),\ (\zeta_1 < 0.\ \zeta_0 > 0),\ \text{or } (\zeta_1 < 0,\ \zeta_0 < 0) \text{ holds}, \\ \beta_3^{**} < \beta_3 < \beta_3^*. \end{cases} \tag{22}$$

Based on the analysis above, we state the following theorems:

Theorem 1 *The disease-free equilibrium* $E_0(S_0, A_{F_0}, A_{P_0}, 0)$ *of the delayed system (5) is asymptotically stable if* $R_0 < 1$ *and unstable if* $R_0 > 1$ *for* $\varrho > 0$.

Theorem 2 *The disease-free equilibrium* $E_0(S_0, A_{F_0}, A_{P_0}, 0)$ *of the delayed system (5) at* $R_0 = 1$ *is linearly neutrally stable for* $\varrho > 0$.

Theorem 3 *When* $R_0 = 1$, *then the undelayed system (15) reveals a backward (forward) bifurcation at disease-free equilibrium* $E_0(S_0, A_{F_0}, A_{P_0}, 0)$ *if and only if* $G(\beta_3) < 0\ (> 0)$.

The graphical presentations of the forward and backward bifurcations are shown in Figs. 2, 3, and 4 for the experimental data mentioned below: $\kappa = 2$, $\varepsilon = 0.1$, $\vartheta = 0.01$, $d = 0.001$, $b = 0.8$, $\theta = 0.01$, $a = 0.9$, $\beta_1 = 0.004$, $\delta_1 = 0.01$, $\delta_2 = 0.009$. At these values of parameters, we evaluate that the range of β_3 is $\beta_3^{**} = 0.00098164 \leq$

$\beta_3 \leq 0.0088652 = \beta_3^*$. The cases of occurrence of forward and backward bifurcations, given in the inequalities (20), (21), and (22), are illustrated from (1)–(3) as below:

(1) If we take $\beta_3 = 0.00008$, then we obtain that $\zeta_2 = 0.000522$, $\zeta_1 = 1.242 \times 10^{-6} > 0$, and $\zeta_0 = 7.75344 \times 10^{-9} > 0$. This case illustrates the inequality (20), and the graph is shown in Fig. 2.
(2) On considering $\beta_3 = 0.01$, we obtain that $\zeta_2 = 0.000522$, $\zeta_1 = 1.242 \times 10^{-}6 > 0$, $\zeta_0 = -5.44699 \times 10^{-8} < 0$, and the discriminant $D = 1.15276 \times 10^{-10} > 0$. It illustrates the inequality (21), and the graph is shown in Fig. 3.
(3) If we take $\beta_3 = 0.006$, which lies between β_3^{**} and β_3^*, then we obtain that the coefficients $\zeta_2 = 0.000522$, $\zeta_1 = 1.242 \times 10^{-6} > 0$ and $\zeta_0 = -4.05047 \times 10^{-8} < 0$, and the discriminant $D = 8.61164 \times 10^{-11} > 0$. This case illustrates the inequality (22), and the graph is shown in Fig. 4.

In Figs. 2, 3, and 4, the solid lines show stability and the dashed lines show instability.

3.2 Endemic equilibrium and stability

Assuming that $S,\ A_F,\ A_P,\ I \neq 0$. To determine the conditions for the existence of endemic (positive) equilibrium $E_e(S^*, A_F^*, A_P^*, I^*)$, we put the right-hand side of the system (5) to zero. We get:

$$\kappa - \delta_1 S - \delta_2 S - \frac{\beta_1 S I}{1+\varepsilon I} - S\vartheta = 0, \tag{23}$$

$$-\frac{\beta_2 A_F I}{1+\varepsilon I} - \vartheta A_F + \delta_1 S = 0, \tag{24}$$

$$-\frac{\beta_3 A_P I}{1+\varepsilon I} - \vartheta A_P + \delta_2 S = 0, \tag{25}$$

$$\frac{\beta_1 S I}{1+\varepsilon I} + \frac{\beta_2 A_F I}{1+\varepsilon I} + \frac{\beta_3 A_P I}{1+\varepsilon I} - (d+\theta+\vartheta)I - \frac{aI}{1+bI} = 0. \tag{26}$$

Solving for S, A_F and A_P in terms of I, from Eqs. (23), (24), and (25), respectively, and substituting the resulting expression in Eq. (26), after simplification, we obtain the following equation in I:

$$P(I) := A_0 + A_1 I + A_2 I^2 + A_3 I^3 + A_4 I^4 = 0, \tag{27}$$

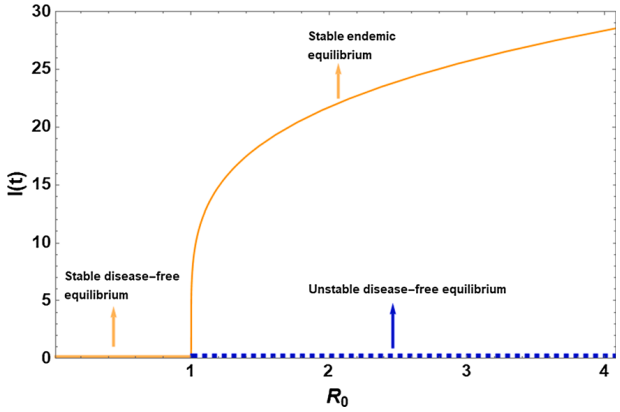


Fig. 2 Plot of R_0 versus $I(t)$, showing the occurrence of forward bifurcation

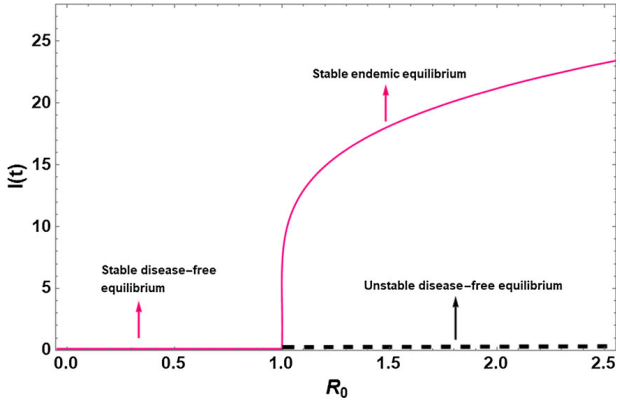


Fig. 3 Plot of R_0 versus $I(t)$, showing the occurrence of forward bifurcation

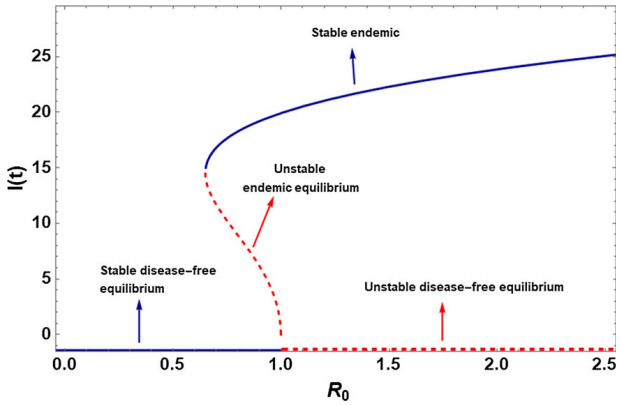


Fig. 4 Plot of R_0 versus $I(t)$, showing the presence of backward bifurcation

where

$$
\begin{aligned}
A_0 &= \vartheta\,(\delta_1+\delta_2+\vartheta)\,(a+d+\theta+\vartheta)(1-R_0),\\
A_1 &= \vartheta\,((\delta_1+\delta_2+\vartheta)\,(3\varepsilon\vartheta+\beta_2+\beta_3)+\beta_1\vartheta)\\
&\quad (a+d+\theta+\vartheta)-\beta_2\delta_1\kappa\vartheta\,(2\varepsilon+b)\\
&\quad -\beta_3\delta_2\kappa\vartheta\,(2\varepsilon+b)\\
&\quad -\beta_1\kappa\vartheta^2(2\varepsilon+b)+b\vartheta^2\,(\delta_1+\delta_2+\vartheta)\\
&\quad (d+\theta+\vartheta)-\beta_2\beta_3\delta_1\kappa-\beta_2\beta_3\delta_2\kappa\\
&\quad -\beta_1\,(\beta_2+\beta_3)\,\kappa\vartheta,\\
A_2 &= \vartheta\Big(\beta_3\,(\delta_2(2a\varepsilon-\varepsilon\kappa(\varepsilon+2b)+(2\varepsilon+b)(d+\theta+\vartheta))\\
&\quad +(\delta_1+\vartheta)\,(2a\varepsilon+(2\varepsilon+b)(d+\theta+\vartheta)))\\
&\quad +3\varepsilon\vartheta\,(\delta_1+\delta_2+\vartheta)\,(a\varepsilon+(\varepsilon+b)(d+\theta+\vartheta))\Big)\\
&\quad +\beta_2\Big(\beta_3\Big(\,(\delta_1+\delta_2)\,(a-\kappa(\varepsilon+b)+d+\theta+\vartheta)\\
&\quad +\vartheta(a+d+\theta+\vartheta)\Big)+\vartheta\big(\delta_1(2a\varepsilon-\varepsilon\kappa(\varepsilon+2b)\\
&\quad +(2\varepsilon+b)(d+\theta+\vartheta))\\
&\quad +(\delta_2+\vartheta)\,(2a\varepsilon+(2\varepsilon+b)(d+\theta+\vartheta))\big)\Big)\\
&\quad -\beta_1\Big(\beta_2\,(-\vartheta(a+d+\theta+\vartheta)+\kappa\vartheta(\varepsilon+b)+\beta_3\kappa)\\
&\quad +\vartheta\,(\vartheta(-2a\varepsilon+\varepsilon\kappa(\varepsilon+2b)\\
&\quad -(2\varepsilon+b)(d+\theta+\vartheta))\\
&\quad -\beta_3(a-\kappa(\varepsilon+b)+d+\theta+\vartheta))\Big),\\
A_3 &= -\,(\varepsilon\vartheta+\beta_3)\,((\varepsilon\vartheta+\beta_2)\,(-a\varepsilon\vartheta-\beta_1(a-b\kappa))\\
&\quad -\varepsilon\delta_1\,(a\varepsilon\vartheta+\beta_2(a-b\kappa)))+\varepsilon\delta_2\,(\varepsilon\vartheta+\beta_2)\,\big(a\varepsilon\vartheta\\
&\quad +\beta_3(a-b\kappa)\big)-(d+\theta+\vartheta)\big(\beta_1\,(\vartheta\,(-\beta_3(\varepsilon+b)\\
&\quad -\varepsilon\vartheta(\varepsilon+2b))-\beta_2\,(\vartheta(\varepsilon+b)+\beta_3))\\
&\quad -(\delta_1+\delta_2+\vartheta)\,(\beta_2\,(\beta_3(\varepsilon+b)+\varepsilon\vartheta(\varepsilon+2b))\\
&\quad +\varepsilon\vartheta\,(\beta_3(\varepsilon+2b)+\varepsilon\vartheta(\varepsilon+3b)))\big),\\
A_4 &= b\,(\varepsilon\vartheta+\beta_2)\,(\varepsilon\vartheta+\beta_3)\,(d+\theta+\vartheta)\\
&\quad (\varepsilon\,(\delta_1+\delta_2+\vartheta)+\beta_1)\,.
\end{aligned}
\tag{28}
$$

For a positive root of I^* of the polynomial $P(I)$, we have

$$S^* = \frac{\kappa+\varepsilon\kappa I^*}{(\delta_1+\delta_2+\vartheta)\,(\varepsilon I^*+1)+\beta_1 I^*}, \tag{29}$$

$$A_F^* = \frac{\delta_1 S^*(\varepsilon I^*+1)}{\varepsilon I^*\vartheta+\beta_2 I^*+\vartheta}, \tag{30}$$

$$A_P^* = \frac{\delta_2 S^*(\varepsilon I^*+1)}{\varepsilon I^*\vartheta+\beta_3 I^*+\vartheta}. \tag{31}$$

So, $E_e(S^*, A_F^*, A_P^*, I^*)$ is a positive equilibrium of the system (5).

Theorem 4 *When $R_0 > 1$, then there is either a unique or three positive endemic equilibria if all equilibria are simple roots.*

Proof Let $R_0 > 1$. We see that the coefficient A_4 is always positive. On the other hand, $A_0 < 0$ when $R_0 > 1$. From Eq. (27), we have a fourth-degree polynomial in I, given below:

$$P(I) := A_0 + A_1 I + A_2 I^2 + A_3 I^3 + A_4 I^4 = 0.$$

The following possibilities for the signs of A_1, A_2, and A_3 exist:

$$
\begin{aligned}
&\mathbf{V}_1 : A_1 > 0,\ A_2 > 0,\ \text{and } A_3 > 0,\\
&\mathbf{V}_2 : A_1 < 0,\ A_2 < 0,\ \text{and } A_3 > 0,\\
&\mathbf{V}_3 : A_1 < 0,\ A_2 > 0,\ \text{and } A_3 > 0,\\
&\mathbf{V}_4 : A_1 < 0,\ A_2 < 0,\ \text{and } A_3 < 0,\\
&\mathbf{V}_5 : A_1 > 0,\ A_2 > 0,\ \text{and } A_3 < 0,\\
&\mathbf{V}_6 : A_1 > 0,\ A_2 < 0,\ \text{and } A_3 > 0,\\
&\mathbf{V}_7 : A_1 > 0,\ A_2 < 0,\ \text{and } A_3 < 0,\\
&\mathbf{V}_8 : A_1 < 0,\ A_2 > 0,\ \text{and } A_3 < 0.
\end{aligned}
$$

Using Descartes' rule of signs [12], $P(I)$ can have either a unique or three positive roots. If any of the conditions $\mathbf{V}_1$–$\mathbf{V}_4$ holds, then there is unique endemic equilibrium, whereas for the existence of three endemic equilibria, any one of the conditions $\mathbf{V}_5$–$\mathbf{V}_8$ must satisfy. □

For the present study, we consider the case of unique endemic equilibrium only. **H1:** Suppose that any of the conditions ($\mathbf{V}_1$–$\mathbf{V}_4$, and $R_0 > 1$) holds, then the system (5) admits a unique endemic equilibrium.

Now, we study the local stability behavior of endemic equilibrium $E_e(S^*, A_F^*, A_P^*, I^*)$. For this, we obtain the characteristic equation of the system (5) at $E_e(S^*, A_F^*, A_P^*, I^*)$ as given below:

$$P_0(\lambda)+\mathrm{e}^{-\lambda\varrho}P_1(\lambda)+\mathrm{e}^{-2\lambda\varrho}P_2(\lambda)+\mathrm{e}^{-3\lambda\varrho}P_3(\lambda)=0, \tag{32}$$

where,

$$
\begin{aligned}
P_0(\lambda) &= \lambda^4+K_1\lambda^3+K_2\lambda^2+K_3\lambda+K_4,\\
P_1(\lambda) &= K_5\lambda^3+K_6\lambda^2+K_7\lambda+K_8,\\
P_2(\lambda) &= K_9\lambda^2+K_{10}\lambda+K_{11},\\
P_3(\lambda) &= K_{12}\lambda+K_{13}.
\end{aligned}
\tag{33}
$$

where the coefficients K_i, $i = 1$ to 13 are given in Appendix. For $\varrho = 0$, the characteristic equation becomes:

$$\lambda^4 + Q_3\lambda^3 + Q_2\lambda^2 + Q_1\lambda + Q_0 = 0, \tag{34}$$

where $Q_i's$, $i = 0, 1, 2, 3$, are given in Appendix.

Based on Routh–Hurwitz criterion, it can be concluded that all the roots of Eq. (34) have negative real parts if the following inequalities hold:

H2: $Q_i > 0$ $(i = 0, 1, 2, 3)$, $Q_3Q_2 - Q_1 > 0$, and
$$Q_3Q_2Q_1 - Q_1^2 - Q_3^2Q_0 > 0. \tag{35}$$

Thus, we state the following theorem:

Theorem 5 *At $\varrho = 0$, the endemic equilibrium $E_2(S^*, A_F^*, A_P^*, I^*)$ is locally asymptotically stable if **H2** holds.*

Now multiplying by $e^{-\lambda\varrho}$ on both sides of the characteristic equation (32), we get:

$$P_0(\lambda)e^{\lambda\varrho} + P_1(\lambda) + e^{-\lambda\varrho}P_2(\lambda) + e^{-2\lambda\varrho}P_3(\lambda) = 0. \tag{36}$$

Change of stability and hence the existence of oscillatory solution may appear if the roots of the characteristic equation are purely imaginary. Therefore, to study the stability of E_e, assuming that $\lambda = i\omega$, $(\omega > 0)$ is a root of Eq. (36). Then, Eq. (36) becomes:

$$P_0(i\omega)e^{i\omega\varrho} + P_1(i\omega) + e^{-i\omega\varrho}P_2(i\omega) + e^{-2i\omega\varrho}P_3(i\omega) = 0. \tag{37}$$

Equation (37) can be rewritten as

$$\begin{aligned}&(M_0 + iN_0)(\cos\omega\varrho + i\sin\omega\varrho) + (M_1 + iN_1)\\ &+ (M_2 + iN_2)(\cos\omega\varrho - i\sin\omega\varrho)\\ &+ (M_3 + iN_3)(\cos 2\omega\varrho - i\sin 2\omega\varrho) = 0,\end{aligned} \tag{38}$$

where M_0, M_1, M_2, M_3, N_0, N_1, N_2 and N_3 denote the real and imaginary parts of $P_0(i\omega)$, $P_1(i\omega)$, $P_2(i\omega)$ and $P_3(i\omega)$, respectively, given as:

$$\begin{aligned}&M_0 = \omega^4 - K_2\omega^2 + K_4,\ N_0 = -K_1\omega^3 + K_3\omega,\\ &M_1 = -K_6\omega^2 + K_8,\ N_1 = -K_5\omega^3 + K_7\omega,\\ &M_2 = -K_9\omega^2 + K_{11},\ N_2 = K_10\omega,\\ &M_3 = K_{13},\ N_3 = K_{12}\omega.\end{aligned}$$

When Eq. (38) splits into its real and imaginary parts, we get

$$M_1 + M_4\cos\omega\varrho + V_4\sin\omega\varrho = -N_3\sin 2\omega\varrho - M_3\cos 2\omega\varrho, \tag{39}$$

$$N_1 + N_5\cos\omega\varrho + M_5\sin\omega\varrho = -N_3\cos 2\omega\varrho + M_3\sin 2\omega\varrho, \tag{40}$$

where $M_4 = M_0 + M_2$, $N_4 = N_2 - N_0$, $M_5 = M_0 - M_2$ and $N_5 = N_2 + N_0$.

On squaring Eqs. (39) and (40) and then adding, we obtain

$$(M_1 + M_4\cos\omega\varrho + N_4\sin\omega\varrho)^2 + (N_1 + N_5\cos\omega\varrho) + M_5(\sin\omega\varrho)^2 - N_3^2 - M_3^2 = 0. \tag{41}$$

On substituting $\sin\omega\varrho = \sqrt{1 - \cos^2\omega\varrho}$ in Eq. (41), we obtain

$$L_1\cos^4\omega\varrho + L_2\cos^3\omega\varrho + L_3\cos^2\omega\varrho + L_4\cos\omega\varrho + L_5 = 0, \tag{42}$$

where

$$\begin{aligned}&L_1 = p_1^2 + 4p_5^2,\ L_2 = 4p_1p_2 + 8p_4p_5,\\ &L_3 = 4p_2^2 + 2p_1p_3 + 4p_4^2 - 4p_5^2,\\ &L_4 = 4p_2p_3 - 8p_4p_5,\\ &L_5 = p_3^2 - 4p_4^2,\\ &p_1 = M_4^2 + N_5^2 - N_4^2 - M_5^2,\ p_2 = M_1M_4 + N_1N_4,\\ &p_3 = M_1^2 + N_4^2 + N_1^2 + M_5^2 - N_3^2 - M_3^2,\\ &p_4 = M_1N_4 + N_1M_5,\\ &p_5 = M_4N_4 + M_5N_5.\end{aligned}$$

Let $\cos\omega\varrho = x$, then Eq. (42) can be written as

$$f(x) := L_1x^4 + L_2x^3 + L_3x^2 + L_4x + L_5 = 0. \tag{43}$$

From Eq. (43), we obtain

$$f'(x) = 4L_1x^3 + 3L_2x^2 + 2L_3x + L_4 = 0. \tag{44}$$

For convenience, we assume that $x = y - L_2/4L_1$. Then, Eq. (44) becomes:

$$y^3 + Ay + B = 0, \tag{45}$$

where

$$A = \frac{L_3}{2L_1} - \frac{3}{16}\left(\frac{T_2}{T_1}\right)^2,$$

$$B = \frac{1}{32}\left(\frac{L_2}{L_1}\right)^3 - \frac{L_2L_3}{8L_1^2} + \frac{L_4}{4L_1}.$$

Now, roots of Eq. (45) are given as

$$y_1 = \sqrt[3]{-\frac{B}{2} + \sqrt{O}} + \sqrt[3]{-\frac{B}{2} - \sqrt{O}},$$

$$y_2 = \eta_1 \sqrt[3]{-\frac{B}{2} + \sqrt{O}} + \eta_2 \sqrt[3]{-\frac{B}{2} - \sqrt{O}},$$
$$y_3 = \eta_2 \sqrt[3]{-\frac{B}{2} + \sqrt{O}} + \eta_1 \sqrt[3]{-\frac{B}{2} - \sqrt{O}},$$

where $\eta_1 = \frac{-1+\sqrt{3}\mathrm{i}}{2}$, $\eta_2 = \frac{-1-\sqrt{3}\mathrm{i}}{2}$ and $O = \frac{B^2}{4} + \frac{A_3}{27}$.

It follows from $\cos \omega\varrho = x$ and $x = y - L_2/4L_1 = F_1(\omega)$ that

$$\cos \omega\varrho = F_1(\omega). \tag{46}$$

On substituting Eq. (46) in Eq. (41) and solving for $\sin \omega\varrho$, we obtain

$$\sin \omega\varrho = F_2(\omega). \tag{47}$$

From Eqs. (46) and (47), we get

$$F_1^2(\omega) + F_2^2(\omega) = 1. \tag{48}$$

Assume that **H3:**, Eq. (48) has at least one positive root ω_0, such that the characteristic equation (32) has a pair of purely imaginary roots $\mathrm{i}\omega_0$.

For ω_0, the corresponding critical value of time delay ϱ_k can be obtained as:

$$\varrho_k = \frac{1}{\omega_0}\left(\arccos F_1(\omega_0) + 2k\pi\right), \ k = 0, 1, 2, \ldots. \tag{49}$$

Assume that ϱ is a bifurcation parameter and ϱ_0=min ϱ_k, $k = 0, 1, 2, \ldots$ is the critical value.

To establish the Hopf bifurcation, we show that $Re\left[\frac{\mathrm{d}\lambda}{\mathrm{d}\varrho}\right]^{-1}\Big|_{\lambda=\mathrm{i}\omega_0} \neq 0$.

Differentiating Eq. (32) with respect to ϱ, we obtain:

$$\begin{aligned}
&\frac{\mathrm{d}\lambda}{\mathrm{d}\varrho}\left(P_0'(\lambda) + \mathrm{e}^{-\lambda\varrho}P_1'(\lambda) - \varrho\mathrm{e}^{-\lambda\varrho}P_1(\lambda) + \mathrm{e}^{-2\lambda\varrho}P_2'(\lambda) - 2\varrho P_2(\lambda)\mathrm{e}^{-2\lambda\varrho} + \mathrm{e}^{-3\lambda\varrho}P_3'(\lambda) - 3\varrho P_3(\lambda)\mathrm{e}^{-3\lambda\varrho}\right)\\
&\quad - \lambda\mathrm{e}^{-\lambda\varrho}P1(\lambda) - 2\lambda P_2(\lambda)\mathrm{e}^{-2\lambda\varrho} - 3\lambda P_3(\lambda)\mathrm{e}^{-3\lambda\varrho} = 0.\\
&\Longrightarrow \frac{\mathrm{d}\lambda}{\mathrm{d}\varrho}\left(P_0'(\lambda) + \mathrm{e}^{-\lambda\varrho}P_1'(\lambda) - \varrho\mathrm{e}^{-\lambda\varrho}P_1(\lambda) + \mathrm{e}^{-2\lambda\varrho}P_2'(\lambda) - 2\varrho P_2(\lambda)\mathrm{e}^{-2\lambda\varrho} + \mathrm{e}^{-3\lambda\varrho}P_3'(\lambda) - 3\varrho P_3(\lambda)\mathrm{e}^{-3\lambda\varrho}\right)\\
&\quad = \lambda\mathrm{e}^{-\lambda\varrho}P1(\lambda) + 2\lambda P_2(\lambda)\mathrm{e}^{-2\lambda\varrho} + 3\lambda P_3(\lambda)\mathrm{e}^{-3\lambda\varrho}.\\
&\Longrightarrow \frac{\mathrm{d}\lambda}{\mathrm{d}\varrho} = \frac{\lambda\left(P_1(\lambda) + 2P_2(\lambda)\mathrm{e}^{-\lambda\varrho} + 3P_3(\lambda)\mathrm{e}^{-2\lambda\varrho}\right)}{\left(P_0'(\lambda)\mathrm{e}^{\lambda\varrho} + P_1'(\lambda) - \varrho P_1(\lambda) + \mathrm{e}^{-\lambda\varrho}\left(P_2'(\lambda) - 2\varrho P_2(\lambda)\right) + \mathrm{e}^{-2\lambda\varrho}\left(P_3'(\lambda) - 3\varrho P_3(\lambda)\right)\right)}.
\end{aligned}$$

Thus, we get

$$\left[\frac{\mathrm{d}\lambda}{\mathrm{d}\varrho}\right]^{-1} = \frac{\left(P_0'(\lambda)\mathrm{e}^{\lambda\varrho} + P_1'(\lambda) + \mathrm{e}^{-\lambda\varrho}P_2'(\lambda) + \mathrm{e}^{-2\lambda\varrho}P_3'(\lambda)\right) - \left(\varrho P_1(\lambda) + 2\varrho P_2(\lambda)\mathrm{e}^{-\lambda\varrho} + 3\varrho P_3(\lambda)\mathrm{e}^{-2\lambda\varrho}\right)}{\lambda\left(P_1(\lambda) + 2P_2(\lambda)\mathrm{e}^{-\lambda\varrho} + 3P_3(\lambda)\mathrm{e}^{-2\lambda\varrho}\right)}.$$

$$\text{i.e.,} \left[\frac{\mathrm{d}\lambda}{\mathrm{d}\varrho}\right]^{-1} = \frac{\left(P_0'(\lambda)\mathrm{e}^{\lambda\varrho} + P_1'(\lambda) + \mathrm{e}^{-\lambda\varrho}P_2'(\lambda) + \mathrm{e}^{-2\lambda\varrho}P_3'(\lambda)\right)}{\lambda\left(P_1(\lambda) + 2P_2(\lambda)\mathrm{e}^{-\lambda\varrho} + 3P_3(\lambda)\mathrm{e}^{-2\lambda\varrho}\right)} - \frac{\varrho}{\lambda}.$$

$$Re\left[\frac{d\lambda}{d\varrho}\right]^{-1}\Big|_{\lambda=\mathrm{i}\omega_0} = \frac{X_1X_2 + Y_1Y_2}{X_2^2 + Y_2^2},$$

where

$$
\begin{aligned}
X_1 &= -\sin\omega_0\varrho_0\left(-4\omega_0^3+2K_2\omega_0\right)\\
&\quad+\cos\omega_0\varrho_0\left(-3K_1\omega_0^2+K_3\right)+\left(-3K_5\omega_0^2+K_7\right)\\
&\quad+\left(K_{10}\cos\omega_0\varrho_0+2\omega_0K_9\sin\omega_0\varrho_0\right)\\
&\quad+K_{16}\cos 2\omega_0\varrho_0,\\
X_2 &= K_5\omega_0^4-K_7\omega_0^2-2K_{10}\omega_0^2\cos\omega_0\varrho_0\\
&\quad+2\left(-K_9\omega_0^2+\omega_0K_{11}\right)\sin\omega_0\varrho_0\\
&\quad-3\omega_0^2K_{12}\cos 2\omega_0\varrho_0\\
&\quad+3\omega_0K_{13}\sin 2\omega_0\varrho_0,\\
Y_1 &= \cos\omega_0\varrho_0\left(-4\omega_0^3+2K_2\omega_0\right)\\
&\quad+\sin\omega_0\varrho_0\left(-3K_1\omega_0^2+K_3\right)+2\omega_0K_6\\
&\quad+2\omega_0K_9\cos\omega_0\varrho_0-K_{10}\sin\omega_0\varrho_0\\
&\quad-K_{16}\sin 2\omega_0\varrho_0,\\
Y_2 &= -K_6\omega_0^3+\omega_0K_8+2\left(-K_9\omega_0^3+\omega_0K_{11}\right)\cos\omega_0\varrho_0\\
&\quad+3\left(\omega_0^2K_{12}\sin 2\omega_0\varrho_0+\omega_0K_{13}\cos 2\omega_0\varrho_0\right)\\
&\quad+2\omega_0^2K_{10}\sin\omega_0\varrho_0.
\end{aligned}
$$

Obviously, if **H4:** $X_1X_2+Y_1Y_2\neq 0$, then $Re\left[\frac{d\lambda}{d\varrho}\right]^{-1}\Big|_{\lambda=i\omega_0}\neq 0$.

Thus, we state the following theorem:

Theorem 6 *For the system (5), if the conditions (**H1–H4**) hold, then the endemic equilibrium* $E_e=(S^*, A_F^*, A_P^*, I^*)$ *is locally asymptotically stable when* $\varrho\in[0,\varrho_0)$*; the system (5) undergoes a Hopf bifurcation at* $E_e=(S^*, A_F^*, A_P^*, I^*)$ *when* $\varrho=\varrho_0$*, and a family of periodic solutions bifurcate from* $E_e=(S^*, A_F^*, A_P^*, I^*)$.

4 Numerical simulation

In this section, numerical experiments are presented to show the analytical results using Mathematica 11.

We have considered the following set of experimental data: $\kappa=2$, $\vartheta=0.01$, $\beta_1=0.02$, $\beta_2=0.006$, $\beta_3=0.008$, $\delta_1=0.1$, $\delta_2=0.02$, $a=0.04$, $d=0.009$, $\theta=0.004$, $\varepsilon=0.5$, $b=0.09$.

At these parameters values, the endemic equilibrium is $E_e(S^*, A_F^*, A_P^*, I^*)=(11.903, 55.602, 9.443, 38.504)$ with $R_0>1$.

Figure 5 shows the behavior of different subpopulations for the time delay $\varrho=1$. Evidently, as time increases, unaware susceptibles decrease, and the fully aware and partially aware, infected, and removed individuals population increase and then start decaying and settle down to steady state $E_e(11.903, 55.602, 9.443, 38.504)$.

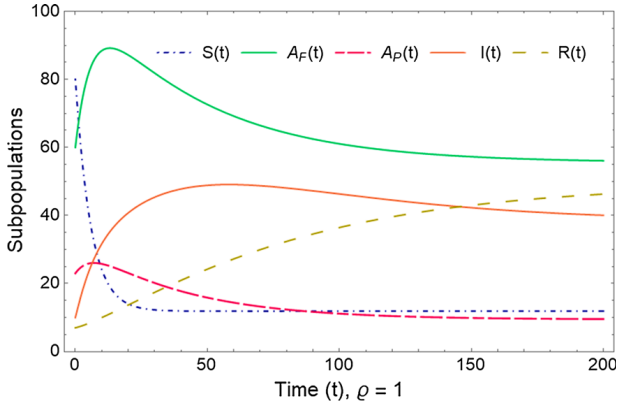


Fig. 5 Subpopulations $S-A_P-A_F-I-R$ at $\varrho=1$

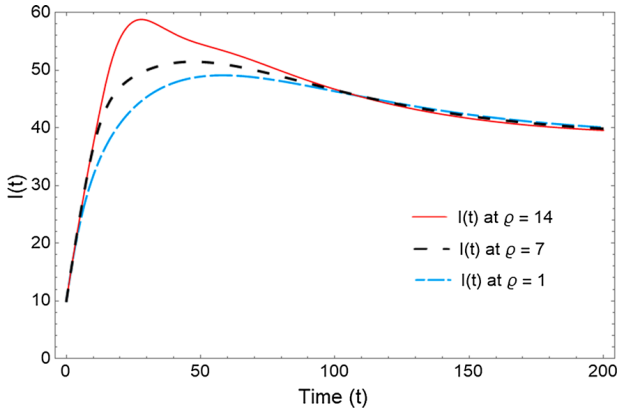
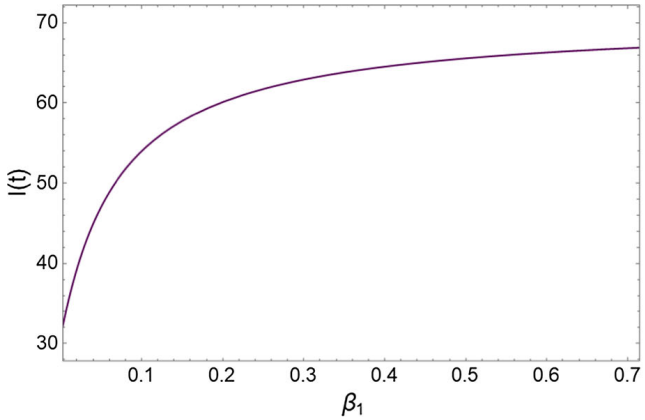


Fig. 6 Infected population $I(t)$ at different values of time delay ϱ

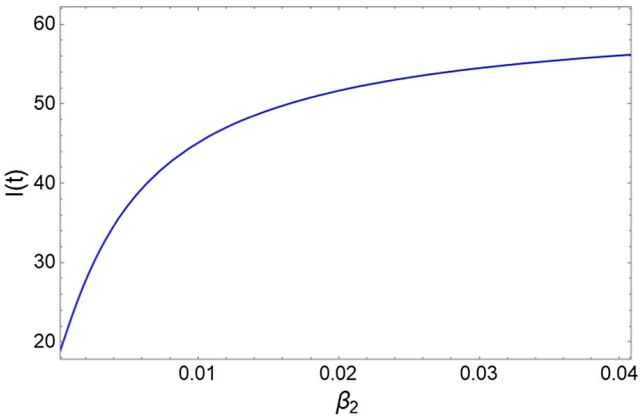
Figure 6 depicts the effect of time delay on the infected population. We plot the infected population for different values of time delay $\varrho=1, 7$, and 14, respectively. It reveals that the large value of time delay causes an increment in the infected population.

Figure 7a–c shows the influence of different transmission rates on infected population $I(t)$. It validates the increment in the number of infected population as the transmission rates increase, which is biologically true.

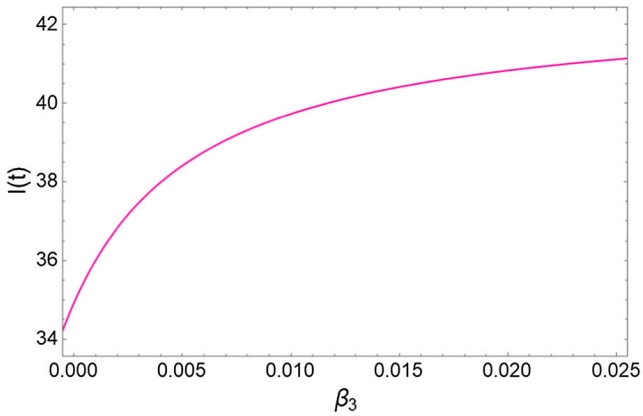
Figure 8a, b shows the impact of full and partial awareness rates (δ_1 and δ_2) on the infected population for the time delay $\varrho=1$. It is evident that if the full awareness rate is high, then infection diminishes at a high level. Partially awareness in humans also helps



(a) Infected population $I(t)$ when β_1 varies.

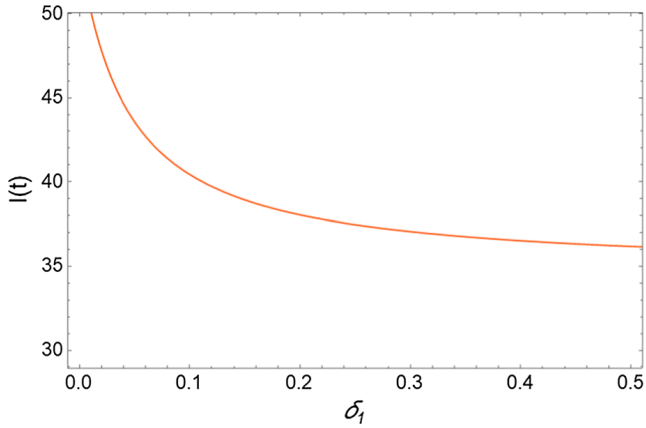


(b) Infected population $I(t)$ when β_2 varies.

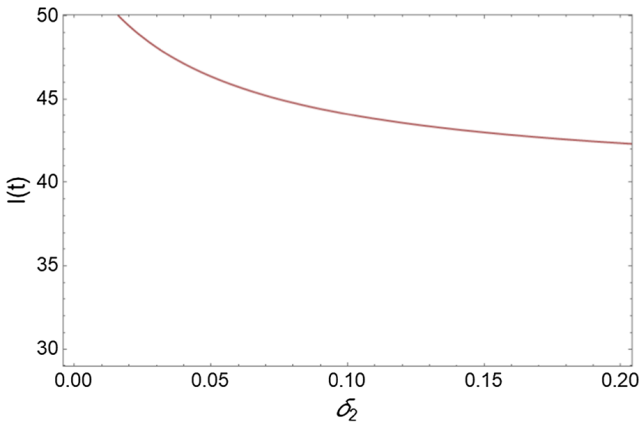


(c) Infected population $I(t)$ when β_3 varies.

Fig. 7 Infected population for the transmission rates of unaware, fully aware, and partially aware susceptibles for the time delay $\varrho = 1$



(a) Infected population $I(t)$ for δ_1.



(b) Infected population $I(t)$ for δ_2.

Fig. 8 Impact of full and partial awareness rates on Infected population for the time delay $\varrho = 1$

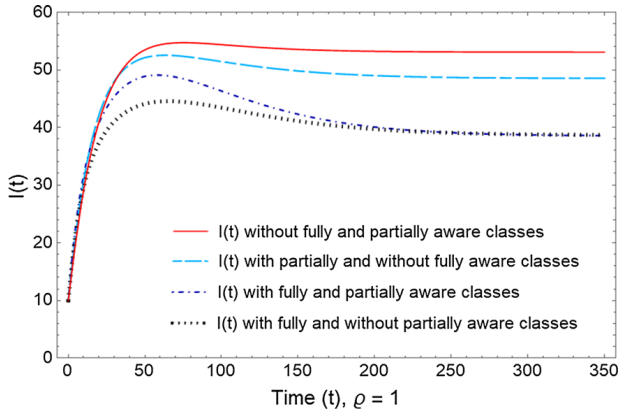
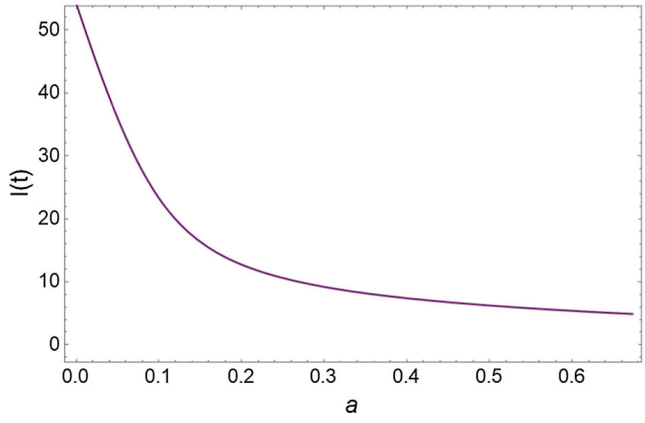


Fig. 9 Dynamics of infectious diseases showing the impact of aware classes on infected individuals $I(t)$ for the time delay $\varrho = 1$

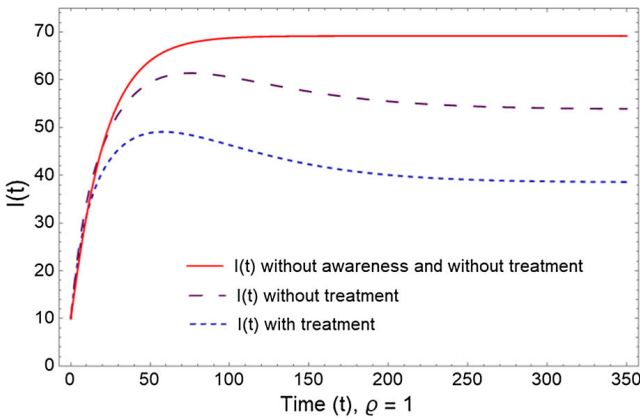
them to escape from the infection, as depicted in Fig. 8b. Thus, more efforts should be put to spread full awareness among people.

Figure 9 examines the potential of fully and partially aware classes in minimizing the impact of an epidemic. From the graph, it is evident that when people are not aware of the spread of disease at all (shown by a solid red line), then the infection occurs at a higher rate. Awareness in humans motivates them to escape from the infection. By the full and right information about a disease, individuals change their attitudes and actions to reduce their chances of becoming infected, spreading the disease further, or experiencing prolonged periods of medical treatment. Attempts to raise awareness of an infectious disease may also build a sense of threat in individuals who are inadequately informed. When there is only partial information available, then infection reduces, but a low level (as shown by the dashed green line). Because of weak, inaccessible, and absence of education, uneducated individuals are rarely formally trained in the process of handling diseases, its prognosis, and diagnosis, preventions, and cures. They take preventive measures by word of mouth or by social media, which leads to a reduction in the infection at a low level. Also, if people are partially mindful, then there are chances that they can take unnecessary medications due to fear of catching the infection, which can weaken their immune system. So those people are at more risk of getting infected (shown by the dot-dashed blue line). Therefore, the absence of partial awareness and presence of full awareness about the spread and prevention of disease are leading to a reduction in the range of illness at the high rate (shown by the dotted black line).

Figure 10a shows that the increase in the cure rate can increase the reduction in infected individuals. Figure 10b shows the influence of saturated treatment rate on infected population $I(t)$. We have plotted the infected population when there is neither treatment nor awareness available. Clearly, in this case, infection is spreading at a very high rate. Purple dashed line shows the infected population when people are aware, but treatment is not available, and the solid red line shows the infected population in the absence of awareness and treatment. The major difference can be seen between these two lines. Blue dashed line shows



(a) Effect of cure rate a on Infected population $I(t)$ for the time delay $\varrho = 1$.



(b) Infected individuals $I(t)$ with and without saturated treatment rate and aware classes for the time delay $\varrho = 1$.

Fig. 10 Impact of cure rate, awareness, and saturated treatment on the infected population for $\varrho = 1$

the infected population when both awareness and treatment are present, which is stabilizing at the lowest level among three lines. Thus, the treatment rate with awareness helps in reducing infection at a faster rate.

To show the presence of Hopf bifurcation, we take the following experimental data: $\kappa = 10$, $\varepsilon = 0.92$, $\beta_1 = 0.2$, $\beta_2 = 0.02$, $\beta_3 = 0.04$, $\delta_1 = 0.05$, $\delta_2 = 0.03$, $\vartheta = 0.1$, $a = 0.2$, $b = 0.5$, $\theta = 0.01$, $d = 0.01$.

We obtain that $E_e(S^*, A_F^*, A_P^*, I^*) = (25.5152, 10.5267, 5.37598, 42.1257)$, at these values of parameters.

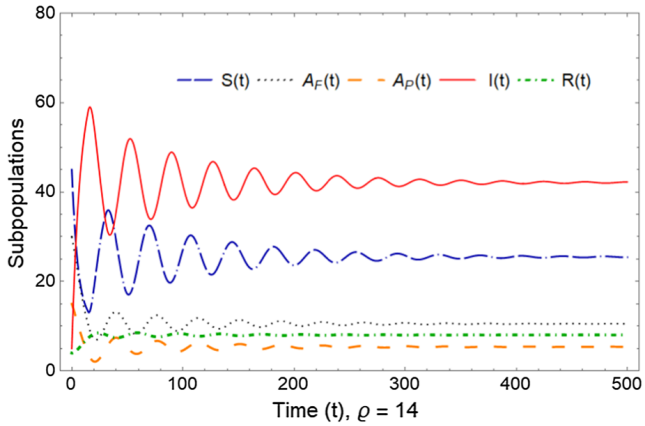
Figure 11 depicts the time series solutions of the model (1) with their respective SIR phase plot for distinct values of time delay ϱ. It is observed that the spread of infectious disease can be controlled for $\varrho = 14$, and $\varrho = 21$. That is, Fig. 11a–d shows that initially periodic solutions appear but after a time; the endemic equilibrium E_e reached to its steady state. Figure 11e–h shows the unstable limit cycle around the endemic equilibrium when time delay $\varrho > \varrho_0 = 24$.

5 Discussion

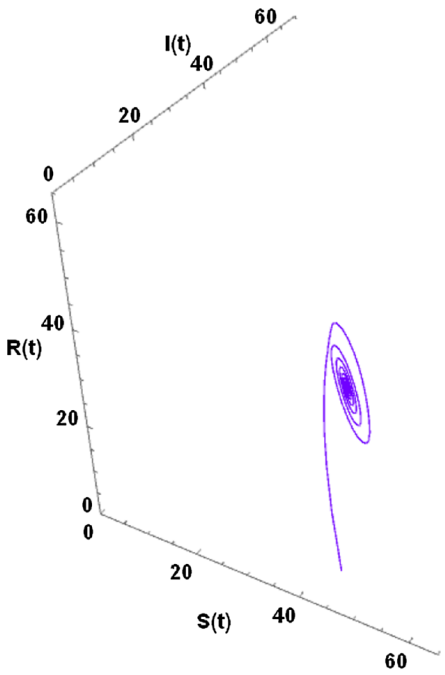
In the present article, we divide the total population into five compartments: unaware susceptibles, fully aware susceptibles, partially aware susceptibles, infected, and removed individuals and study a time-delayed epidemic model with the inclusion of saturated incidences and treatment rates. We analyze the model mathematically and study the dynamic behaviors of the epidemic model with and without time delay ϱ. The mathematical analysis of the model shows that it exhibits two equilibria: disease-free and endemic. By deriving the basic reproduction number R_0, we prove that the disease-free equilibrium is locally asymptotically stable when $R_0 < 1$, unstable when $R_0 > 1$, and linearly neutrally stable when $R_0 = 1$ for the time delay $\varrho > 0$. When we don't consider the time delay, then using the center manifold theory, it is obtained that the forward or backward bifurcation occurs when $R_0 = 1$. We obtain the bifurcation range of forward and backward bifurcations, given in the inequalities (20), (21), and (22), respectively. In the presence of forward bifurcation, as R_0 increases through unity, the disease-free equilibrium loses its stability, and a stable endemic equilibrium appears. In contrast, the presence of backward bifurcation shows that a stable endemic equilibrium coexists with a stable DFE when $R_0 < 1$. It has an important implication as it fails the ideal condition of reducing R_0 below unity to eradicate the diseases from society. Thus, the control programs must reduce R_0 further than below unity to eliminate the disease. Schematic diagrams of forward and backward bifurcation are depicted in Figs. (2), (3), and (4). Further, the local stability of the endemic equilibrium has been studied, which demonstrates that by choosing time delay as a bifurcation parameter, the Hopf bifurcation occurs near the endemic equilibrium, revealing the presence of oscillatory and periodic solutions.

The numerical simulations show the graphical representation of the effectiveness of theoretical results. We see that the consideration of time delay has a significant role as it has an impact on the number of infectives. It is seen that when the time delay is high, then infection spreads at a higher rate. The periodic and oscillatory solutions have been plotted near endemic equilibrium, which shows the presence of Hopf bifurcation at different values of time delay ϱ. As the delay ϱ passes through the critical value ϱ_0, the endemic equilibrium loses its stability, and an unstable limit cycle appears. When there is no treatment available, then knowledge about the spread of disease is the main focus. The role of full and partial awareness in susceptibles, with and without saturated treatment, has been shown numerically. When treatment is not given to infected individuals, then only susceptibles' full or partial awareness shows the significant difference in the number of infectives (Figs. 9, 10b), whereas if we consider saturated treatment rate along with awareness in the susceptibles, then the transmission pattern of infectious diseases changes more effectively, and the reduction in the prevalence of the disease can be seen to the more extent.

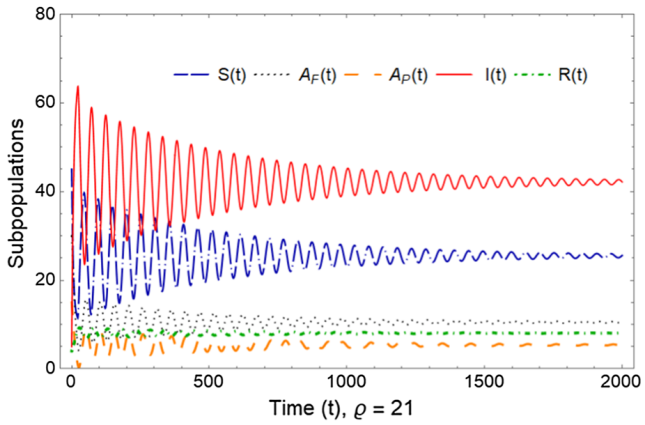
The findings of the model, consisting of explicit saturated incidences with latent period and saturated treatment rate, are capable of demonstrating the significant role of the latent period, the behavior of susceptibles through different subclasses, and limitation in available facilities of treatment. The results are capable of understanding the role of varying protection levels of susceptibles in the transmission pattern of the infectious diseases and hence further suggesting the control strategies to prevent the spread of infections at a massive scale. The full awareness about the spread of infectious disease increases public perception to avert infection and willingness to adopt prevention methods, which mitigates the transmission of disease. Due to the



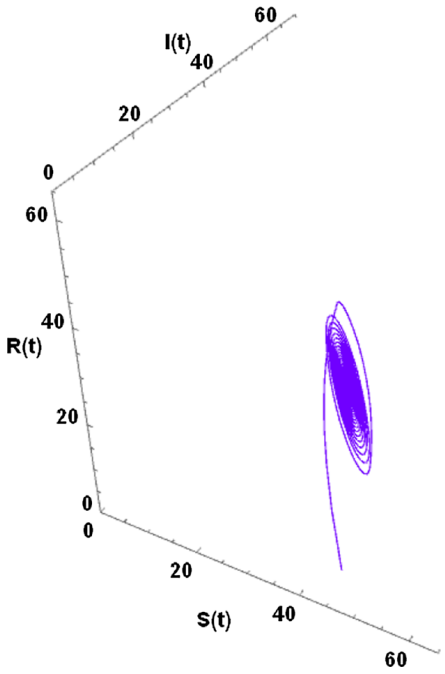
(a) Time series solution of the model (1) for $\varrho = 14$.



(b) Phase plot of susceptible, infected and removed population when $\varrho = 14$.

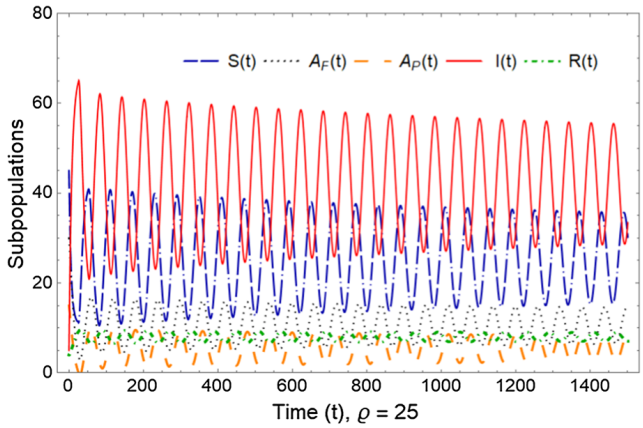


(c) Time series solution of the model (1) for $\varrho = 21$.

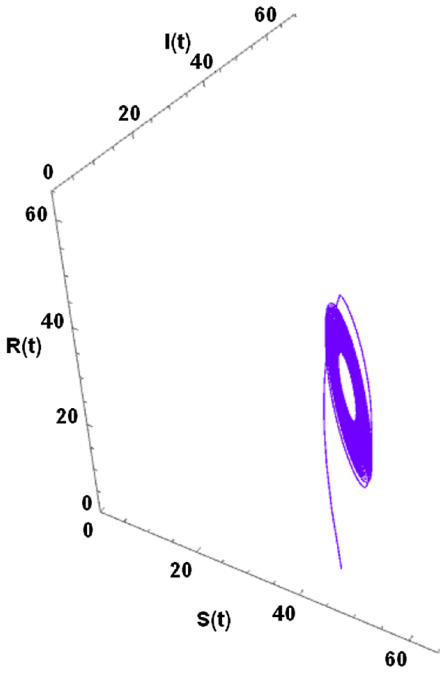


(d) Phase plot of susceptible, infected and removed population when $\varrho = 21$.

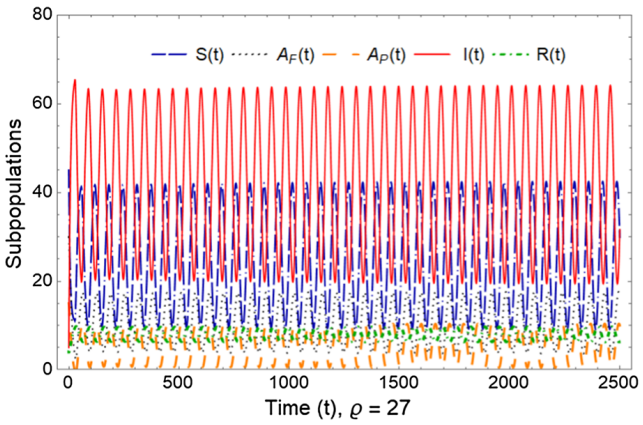
Fig. 11 Graphs depicting the presence of Hopf bifurcation for different values of time delay ϱ



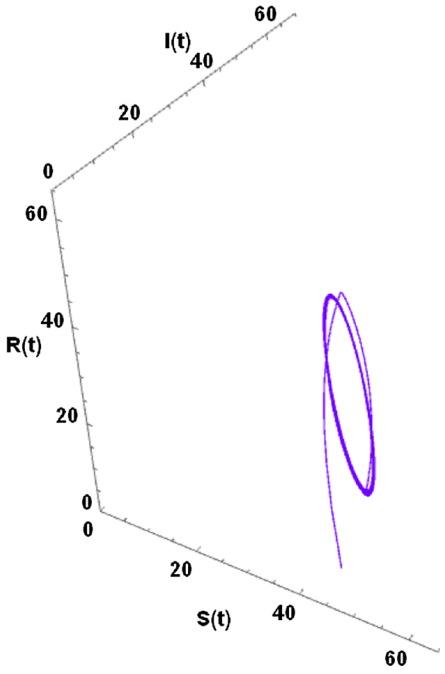
(e) Time series solution of the model (1) for $\varrho = 25$.



(f) Phase plot of susceptible, infected and removed population when $\varrho = 25$.



(g) Time series solution of the model (1) for $\varrho = 27$.



(h) Phase plot of susceptible, infected and removed population when $\varrho = 27$.

Fig. 11 continued

different social, educational, and limited information resources, some people may have incomplete information. Therefore, these partially aware individuals adopt insufficient preventive methods, through which infection reduces but at a low level. Public health initiatives can put extra effort into making individuals fully aware by enhancing health literacy and providing sufficient information resources. It can contribute to early case detection and help in reducing the transmission of infection. Thus, for the eradication of the disease, programs related to regular knowledge, full information, education, and communication, concerning the spread of infectious diseases and its importance to the public and health-care workers, will help them in improving their general attitude toward it. Timely disease pieces of information updates are highly needed. Together with awareness, appropriate treatment to infectives and the availability of health resources will help in diminishing the infection from society, effectively.

Acknowledgements The authors are grateful to the Delhi Technological University, Delhi, India, for providing financial support to carry out this research work. They are also indebted to the editor and the anonymous reviewers for their constructive comments and suggestions, which led to an improvement in the original manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interest regarding the publication of this article.

Appendix

Using Mathematica 11, we obtain the coefficients of the characteristic equation (32) as follows:

$$K_1 = \frac{a + (1+bI^*)^2(d+\theta+4\vartheta) + \delta_1(1+bI^*)^2 + \delta_2(1+bI^*)^2}{(1+bI^*)^2},$$

$$K_2 = \frac{(\delta_1+\delta_2)\left(a+(1+bI^*)^2(\theta+d+3\vartheta)\right) + 3\vartheta\left(a+(d+\theta+2\vartheta)(bI^*+1)^2\right)}{(1+bI^*)^2},$$

$$K_3 = \vartheta\left((\delta_1+\delta_2)\left(2a+(1+bI^*)^2(2(d+\theta)+3\vartheta)\right) + \vartheta\left(3a+(bI^*+1)^2(3(d+\theta)+4\vartheta)\right)\right),$$

$$K_4 = \frac{\vartheta^2(\delta_1+\delta_2+\vartheta)\left(a+(d+\theta+\vartheta)(1+bI^*)^2\right)}{(1+bI^*)^2},$$

$$K_5 = \frac{\beta_3\left(-A_P^* + \varepsilon I^{*2} + I^*\right) + \beta_2\left(-A_F^* + \varepsilon I^{*2} + I^*\right) + \beta_1\left(-S^* + \varepsilon I^{*2} + I\right)}{(\varepsilon I^*+1)^2},$$

$$\begin{aligned}
K_6 = {} & \frac{1}{(1+bI^*)^2(1+\varepsilon I^*)^2}(\beta_3(I^*(\varepsilon I^*+1)\\
& \left(a+(d+\theta+3\vartheta)(1+bI^*)^2\right)\\
& +(1+bI^*)^2((\delta_1+\delta_2)\,I^*(1+\varepsilon I^*)\\
& -A_P^*(\delta_1+\delta_2+3\vartheta))) + \beta_2(I^*(\varepsilon I^*+1)\\
& \left(a+(1+bI^*)^2(\theta+d+3\vartheta)\right)\\
& +(1+bI^*)^2((\delta_1+\delta_2)\\
& \times I^*(\varepsilon I^*+1) - A_F^*(\delta_1+\delta_2+3\vartheta)))\\
& +\beta_1(aI^*(\varepsilon I^*+1) + (1+bI^*)^2\\
& (I^*(1+\varepsilon I^*)(\theta+d+3\vartheta)\\
& -3S^*\vartheta) - (\delta_1+\delta_2)\,S^*(bI^*+1)^2)),
\end{aligned}$$

$$\begin{aligned}
K_7 = {} & \frac{1}{(1+bI^*)^2(1+\varepsilon I^*)^2}(\beta_3(I^*(1+\varepsilon I^*)((\delta_1+\delta_2)\\
& \left(a+(1+bI^*)^2(\theta+d+2\vartheta)\right)\\
& +\vartheta(2a+(1+bI^*)^2\\
& \times(2(d+\theta)+3\vartheta))) - \vartheta A_P^*(1+bI^*)^2\\
& (2\delta_1+2\delta_2+3\vartheta)) + \beta_2(I^*(\varepsilon I^*+1)((\delta_1+\delta_2)\\
& \times\left(a+(1+bI^*)^2(\theta+d+2\vartheta)\right)\\
& +\vartheta\left(2a+(1+bI^*)^2(2(d+\theta)+3\vartheta)\right))\\
& -\vartheta A_F^*(bI^*+1)^2\\
& (2\delta_1+2\delta_2+3\vartheta)) + \beta_1\vartheta(2aI^*(\varepsilon I^*+1)\\
& +(bI^*+1)^2(I^*(\varepsilon I^*+1)(2(d+\theta)+3\vartheta)\\
& -3S^*\vartheta) - 2\,(\delta_1+\delta_2)\,S^*(bI^*+1)^2)),
\end{aligned}$$

$$\begin{aligned}
K_8 = {} & \frac{\vartheta}{(bI^*+1)^2(\varepsilon I^*+1)^2}((\delta_1+\vartheta+\delta_2)\\
& (\beta_3(I^*(\varepsilon I^*+1)\\
& \left(a+(1+bI^*)^2(\theta+d+\vartheta)\right)\\
& -\vartheta A_P^*(1+bI^*)^2)
\end{aligned}$$

$$
\begin{aligned}
&+\beta_2\left(I^*\left(a+(bI^*+1)^2(\theta+d+\vartheta)\right)(1+\varepsilon I^*)\right.\\
&\left.-\vartheta A_F^*(1+bI^*)^2\right)+\beta_1\vartheta(aI^*(1+\varepsilon I^*)\\
&+(1+bI^*)^2(I^*(\theta+d+\vartheta)(\varepsilon I^*+1)-S^*\vartheta)\\
&-(\delta_1+\delta_2)S^*(1+bI^*)^2)),
\end{aligned}
$$

$$
\begin{aligned}
K_9=&\frac{1}{(1+\varepsilon I^*)^3}(I^*(\beta_1\left(\beta_3\left(-A_P^*-S^*+\varepsilon I^{*2}+I^*\right)\right.\\
&\left.+\beta_2\left(-A_F^*-S^*+\varepsilon I^{*2}+I^*\right)\right)\\
&+\beta_2\beta_3\left(-A_P^*-A_F^*+\varepsilon I^{*2}+I^*\right))^3),
\end{aligned}
$$

$$
\begin{aligned}
K_{10}=&\frac{I^*}{(1+\varepsilon I^*)^3(1+bI^*)^2}(\beta_1(\beta_3((1+\varepsilon I^*)aI^*\\
&-2\vartheta A_P^*(1+bI^*)^2+(1+bI^*)^2\\
&(I^*(\theta+d+2\vartheta)(1+\varepsilon I^*)\\
&-2S^*\vartheta)-\delta_1S^*(1+bI^*)^2)+\beta_2((1+\varepsilon I^*)aI^*\\
&-2\vartheta A_F^*(1+bI^*)^2+(1+bI^*)^2(I^*(1+\varepsilon I^*)\\
&\times(\theta+d+2\vartheta)-2S^*\vartheta)-\delta_2S^*(1+bI^*)^2))\\
&+\beta_2\beta_3((1+\varepsilon I^*)I^*\left(a+(1+bI^*)^2(\theta+d+2\vartheta)\right)\\
&+(1+bI^*)^2((\delta_1+\delta_2)\times I^*(\varepsilon I^*+1)\\
&-(\delta_1+\delta_2+2\vartheta)\left(A_P^*+A_F^*\right)))),
\end{aligned}
$$

$$
\begin{aligned}
K_{11}=&\frac{I^*}{(1+\varepsilon I^*)^3(1+bI^*)^2}(\beta_1\vartheta(\beta_3((1+\varepsilon I^*)aI^*\\
&-\vartheta A_P^*(1+bI^*)^2\\
&+(1+bI^*)^2(I^*(1+\varepsilon I^*)(d+\theta+\vartheta)\\
&-S^*\vartheta)-\delta_1S^*(bI^*+1)^2)+\beta_2(aI^*(\varepsilon I^*+1)\\
&-\vartheta A_F^*(1+bI^*)^2\\
&+(1+bI^*)^2(I^*(1+\varepsilon I^*)(\theta+d+\vartheta)\\
&-S^*\vartheta)-\delta_2S^*(bI^*+1)^2))\\
&+\beta_2\beta_3\left(\delta_1+\delta_2+\vartheta\right)(I^*(1+\varepsilon I^*)\\
&\left(a+(1+bI^*)^2(\theta+d+\vartheta)\right)\\
&-\vartheta(1+bI^*)^2\left(A_P^*+A_F^*\right))),
\end{aligned}
$$

$$
\begin{aligned}
&\frac{1}{(\varepsilon I^*+1)^3}\left(I^*(\beta_2\beta_3(-\vartheta)\right.\\
&\times(\delta_1+\delta_2+\vartheta)\left(A_P^*+A_F^*\right)\\
&-\beta_1(\beta_2\left(\vartheta^2A_F^*+S^*\left(\vartheta\left(\delta_2+\vartheta\right)-\beta_3I^{*2}(d+\theta+\vartheta)\right)\right)\\
&\left.+\beta_3\vartheta\left(\vartheta A_P^*+S^*\left(\delta_1+\vartheta\right)\right)\right)))\\
&\frac{1}{(\varepsilon I^*+1)^2}\left(\left(\delta_1+\delta_2+\vartheta\right)(\beta_2\vartheta^2\left(-A_F^*\right)\right.\\
&-\beta_3\left(\vartheta^2A_P^*-\beta_2I^{*2}(d+\theta+\vartheta)\right))\\
&+\beta_1\vartheta\left(\beta_2I^{*2}(d+\theta+\vartheta)+\beta_3I^{*2}(d+\theta+\vartheta)\right.\\
&\left.\left.-S^*\vartheta\left(\delta_1+\delta_2+\vartheta\right)\right)\right)\\
&-\frac{\beta_1\beta_2\beta_3S^*I^{*2}\vartheta\left(A_P^*+A_F^*+S^*\right)}{(\varepsilon I^*+1)^4}\\
&+\frac{1}{(bI^*+1)^2(\varepsilon I^*+1)^3}\left(a(\beta_1I^*\left(\beta_2I^*\left(\beta_3S^*I^*\right.\right.\right.\\
&\left.\left.+\varepsilon I^*\vartheta+\vartheta\right)+\vartheta(\varepsilon I^*+1)\left(\varepsilon I^*\vartheta+\beta_3I^*+\vartheta\right)\right)\\
&+(\delta_1+\delta_2+\vartheta)(\varepsilon I^*+1)\left(I^*\vartheta\varepsilon+\vartheta+\beta_2I^*\right)\\
&\left.\left(\varepsilon\vartheta I^*+\vartheta+\beta_3I^*\right)\right)),
\end{aligned}
$$

$$
\begin{aligned}
Q_1=&\frac{1}{(bI^*+1)^2(1+\varepsilon I^*)^4}\left(\beta_1(\beta_2I^*((1+\varepsilon I^*)(aI^*(1+\varepsilon I^*)\right.\\
&+(1+bI^*)^2(I^*(1+\varepsilon I^*)(d+\theta+2\vartheta)-2\vartheta S^*))\\
&-(bI^*+1)^2(\beta_3S^*I^*(A_P^*+S^*-I^*(\varepsilon I^*+1))\\
&+A_F^*\left(\beta_3S^*I^*+2(\varepsilon I^*\vartheta+\vartheta)\right)+\delta_2S^*(\varepsilon I^*+1)))\\
&+(\varepsilon I^*+1)(\beta_3I^*((1+\varepsilon I^*)aI^*\\
&-2\vartheta A_P^*(1+bI^*)^2+(1+bI^*)^2(I^*(1+\varepsilon I^*)\\
&(\theta+d+2\vartheta)-2S^*\vartheta)\\
&-(1+bI^*)^2\delta_1S^*)+\vartheta(1+\varepsilon I^*)(2aI^*(1+\varepsilon I^*)\\
&+(1+bI^*)^2(I^*(1+\varepsilon I^*)(2(d+\theta)+3\vartheta)-3\vartheta S^*)\\
&-2\left(\delta_1+\delta_2\right)S^*(1+bI^*)^2)))\\
&+(1+\varepsilon I^*)(\beta_2(I^*(\beta_3(I^*(\varepsilon I^*+1)\\
&\left(a+(1+bI^*)^2(\theta+d+2\vartheta)\right)\\
&+(1+bI^*)^2(1+\varepsilon I^*)((\delta_1+\delta_2)I^*\\
&-A_P^*\left(\delta_1+\delta_2+2\vartheta\right)))+(\varepsilon I^*+1)^2((\delta_1+\delta_2)\\
&\times\left(a+(1+bI^*)^2(\theta+d+2\vartheta)\right)
\end{aligned}
$$

$$
K_{12}=-\frac{\beta_1\beta_2\beta_3S^*I^{*2}\left(A_P^*+A_F^*+S^*-I^*(1+\varepsilon I^*)\right)}{(1+\varepsilon I^*)^4},
$$

$$
K_{13}=\frac{\beta_1\beta_2\beta_3S^*I^{*2}\left((1+\varepsilon I^*)aI^*-\vartheta(1+bI^*)^2\left(A_P^*+A_F^*\right)+(1+bI^*)^2(I^*(\theta+d+\vartheta)(1+\varepsilon I^*)-S^*\vartheta)\right)}{(bI^*+1)^2(\varepsilon I^*+1)^4}.
$$

The coefficients of the characteristic equation (36) are obtained as follows:

$$
\begin{aligned}
Q_0=&\vartheta^2\left(\delta_1+\delta_2+\vartheta\right)(d+\theta+\vartheta)\\
&+\frac{I^*\vartheta\left(\left(\beta_2+\beta_3\right)\left(\delta_1+\delta_2+\vartheta\right)+\beta_1\vartheta\right)(d+\theta+\vartheta)}{\varepsilon I^*+1}\\
&+\vartheta(2a+(bI^*+1)^2(2(d+\theta)+3\vartheta))))\\
&-A_F^*(bI^*+1)^2(\vartheta\left(2\delta_1+2\delta_2+3\vartheta\right)(\varepsilon I^*+1)\\
&+\beta_3I^*(\delta_1+\delta_2+2\vartheta)))\\
&+(\varepsilon I^*+1)(\beta_3(I^*(\varepsilon I^*+1)\\
&\times((\delta_1+\delta_2)(a+(1+bI^*)^2(d+\theta+2\vartheta))
\end{aligned}
$$

$$
\begin{aligned}
&+\vartheta\left(2a+(1+bI^*)^2(2(d+\theta)+3\vartheta)\right))\\
&-\vartheta A_P^*(bI^*+1)^2\\
&\times(2\delta_1+2\delta_2+3\vartheta))\\
&+\vartheta(\varepsilon I^*+1)^2\times((\delta_1+\delta_2)\\
&\left(2a+(1+bI^*)^2(2(d+\theta)+3\vartheta)\right)\\
&+\vartheta\left(3a+(1+bI^*)^2(4\vartheta+3(\theta+d))\right)))))\Big),
\end{aligned}
$$

$$
\begin{aligned}
Q_2=&\frac{1}{(1+\varepsilon I^*)^3(1+bI^*)^2}\Big(\beta_1((1+\varepsilon I^*)\\
&\left((1+\varepsilon I^*)aI^*+(1+bI^*)^2(I^*(1+\varepsilon I^*)\right.\\
&\left.(\theta+d+3\vartheta)-3S^*\vartheta\right)\\
&-(bI^*+1)^2\left(\beta_3 I^*\left(A_P^*+S^*-I^*(\varepsilon I^*+1)\right)\right.\\
&\left.+\beta_2 I^*\left(A_F^*+S^*-I^*(\varepsilon I^*+1)\right)\right.\\
&\left.+(\delta_1+\delta_2)S^*(\varepsilon I^*+1)\right))\\
&+\beta_2(I^*(1+\varepsilon I^*)^2\left(a+(1+bI^*)^2(d+\theta+3\vartheta)\right)\\
&+(1+bI^*)^2(I^*(\beta_3\left(-A_P^*+\varepsilon I^{*2}+I^*\right)\\
&+(\delta_1+\delta_2)(\varepsilon I^*+1)^2)\\
&-A_F^*\left((\delta_1+\delta_2+3\vartheta)(\varepsilon I^*+1)+\beta_3 I^*\right)))\\
&+(\varepsilon I^*+1)\\
&\times(\beta_3(I^*(1+\varepsilon I^*)\left(a+(1+bI^*)^2(\theta+3\vartheta+d)\right)\\
&+(1+bI^*)^2((\delta_1+\delta_2)I^*(\varepsilon I^*+1)\\
&-A_P^*(\delta_1+\delta_2+3\vartheta)))+(\delta_1+\delta_2)(\varepsilon I^*+1)^2\\
&\left(a+(1+bI^*)^2(\theta+d+3\vartheta)\right)))\\
&+3\vartheta\left(\frac{a}{(1+bI^*)^2}+d+\theta+2\vartheta\right),
\end{aligned}
$$

$$
\begin{aligned}
Q_3=&\frac{a}{(bI^*+1)^2}\\
&+\frac{\beta_3\left(-A_P^*+\varepsilon I^{*2}+I^*\right)+\beta_2\left(-A_F^*+\varepsilon I^{*2}+I^*\right)+\beta_1\left(-S+\varepsilon I^{*2}+I^*\right)}{(\varepsilon I^*+1)^2}\\
&+d+\delta_1+\delta_2+\theta+4\vartheta.
\end{aligned}
$$

References

1. Kermack, W.O., Mckendrick, A.G.: A contribution to the mathematical theory of epidemic. Proc. R. Soc. Lond. **115**(772), 700–721 (1927)
2. Capasso, V., Serio, G.: A generalization of the Kermack Mckendrick deterministic epidemic model. Math. Biosci. **42**(1–2), 43–61 (1978)
3. Anderson, R.M., May, R.M.: Regulation and stability of host-parasite population interactions: I. regulatory processes. J. Anim. Ecol. **47**(1), 219–267 (1978)
4. Cooke, K.L.: Stability analysis for a vector disease model. Rock. Mount. J. Math. **9**, 31–42 (1979)
5. Guckenheimer, J., Holmes, P.: Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields. Springer, Berlin (1983)
6. Hale, J.K., Lunel, S.M.V.: Introduction to Functional Differential Equations. Springer, New York (1993)
7. Hethcote, H.W., van den Driessche, P.: An SIS epidemic model with variable population size and a delay. J. Math. Biol. **34**, 177–194 (1995)
8. van den Driessche, P., Watmough, J.: Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math. Biosci. **180**(1–2), 29–48 (2002)
9. Murray, J.D.: Introduction to Mathematical Biology: I. An Introduction. Springer, Berlin (2002)
10. Wang, W., Ruan, S.: Bifurcation in an epidemic model with constant removal rates of the infectives. J. Math. Anal. Appl. **21**, 775–793 (2004)
11. Castillo-Chavez, C., Song, B.: Dynamical models of tuberculosis and their applications. Math. Biosci. Eng. **1**(2), 361–404 (2004)
12. Wang, X.: A simple proof of Descartes's rule of signs. Am. Math. Mon. **111**(6), 525–526 (2004)
13. Song, X., Cheng, S.: A delay-differential equation model of HIV infection of CD4+ T-cells. J. Kor. Math. Soc. **42**(5), 1071–1086 (2005)
14. Liu, Y., Cui, J.: The impact of media convergence on the dynamics of infectious diseases. Int. J. Biomath. **1**(1), 65–74 (2008)
15. Zhang, J.Z., Jin, Z., Liu, Q.X., Zhang, Z.Y.: Analysis of a delayed SIR model with nonlinear incidence rate. Discret. Dyn. Nat. Soc. **2008**, 636153 (2008)
16. Li, X.Z., Li, W.S., Ghosh, M.: Stability and bifurcation of an SIR epidemic model with nonlinear incidence and treatment. Appl. Math. Comput. **210**(1), 141–150 (2009)
17. Jones, J.H., Salathé, M.: Early assessment of anxiety and behavioral response to novel swine-origin influenza A(H1N1). PLoS One **4**(12), e8032 (2009)
18. Pappas, G., Kiriaze, I.J., Giannakis, P., Falagas, M.E.: Psychosocial consequences of infectious diseases. Clin. Microbiol. Infect. **15**(8), 743–747 (2009)
19. Funk, S., Gilad, E., Jansen, V.A.A.: Endemic disease, awareness, and local behavioral response. J. Theor. Biol. **264**(2), 501–509 (2010)
20. Kiss, I.Z., Cassell, J., Recker, M., Simon, P.L.: The impact of information transmission on epidemic outbreaks. Math. Biosci. **225**(1), 1–10 (2010)
21. Misra, A.K., Sharma, A., Singh, V.: Effect of awareness programs in controlling the prevalence of an epidemic with time delay. J. Biol. Syst. **19**(2), 389–402 (2011)
22. Cai, L., Li, X., Ghosh, M.: Global dynamics of a mathematical model for HTLV-I infection of CD4+ T-cells. Appl. Math. Model. **35**(7), 3587–3595 (2011)
23. Hatzopoulos, V., Taylor, M., Simon, P.L., Kiss, I.Z.: Multiple sources and routes of information transmission: implications for epidemic dynamics. Math. Biosci. **231**, 197–209 (2011)
24. Zhou, L., Fan, M.: Dynamics of an SIR epidemic model with limited medical resources revisited. Nonlinear Anal.: Real World Appl. **13**(1), 312–324 (2012)
25. Hattaf, K., Lashari, A.A., Louartassi, Y., Yousfi, N.: A delayed SIR epidemic model with general incidence rate. Electron. J. Qual. Theory Differ. Equ. **3**, 1–9 (2013)
26. Dubey, B., Patra, A., Srivastava, P.K., Dubey, U.S.: Modeling and analysis of an SEIR model with different types of nonlinear treatment rates. J. Biol. Syst. **21**(03), 1350023 (2013)

27. Zhao, H., Lin, Y., Dai, Y.: An SIRS epidemic model incorporating media coverage with time delay. Comput. Math. Methods Med. **2014**, 680743 (2014)
28. Zuo, L., Liu, M.: Effect of awareness programs on the epidemic outbreaks with time delay. Abstr. Appl. Anal. **2014**, 940841 (2014)
29. Zuo, L., Liu, M., Wang, J.: The impact of awareness programs with recruitment and delay on the spread of an epidemic. Math. Problems Eng. **2015**, 235935 (2015)
30. Buonomo, B., Cerasuolo, M.: The effect of time delay in plant-pathogen interactions with host demography. Math. Biosci. Eng. **12**(3), 473–490 (2015)
31. Pruyt, E., Auping, W.L., Kwakkel, J.H.: Ebola in West Africa: model-based explo- ration of social psychological effects and interventions. Syst. Res. Behav. Sci. **32**, 2–14 (2015)
32. Yang, M., Sun, F.: Global stability of SIR models with nonlinear Incidence and discontinuous treatment. Electron. J. Differ. Equ. **2015**(304), 1–8 (2015)
33. Greenhalgh, D., Rana, S., Samanta, S., Sardar, T., Bhattacharya, S., Chattopadhyay, J.: Awareness programs control infectious disease - multiple delay induced mathematical model. Appl. Math. Comput. **251**, 539–563 (2015)
34. Dubey, B., Dubey, P., Dubey, U.S.: Role of media and treatment on an SIR model. Nonlinear Anal.: Modell. Control **21**(2), 185–200 (2015)
35. Kumar, A.N.: Stability of a time delayed SIR epidemic model along with nonlinear incidence rate and Holling Type-II treatmemt rate. Int. J. Comput. Methods **15**(6), 1850055 (2018)
36. Goel, K., Nilam, : A mathematical and numerical study of a SIR epidemic model with time delay, nonlinear incidence and treatment rates. Theory Biosci. **138**(2), 203–213 (2019)
37. Kumar, A., Nilam, : Mathematical analysis of a delayed epidemic model with nonlinear incidence and treatment rates. J. Eng. Math. **115**, 1–20 (2019)
38. Upadhyay, R.K., Pal, A.K., Kumari, S., Roy, P.: Dynamics of an SEIR epidemic model with nonlinear incidence and treatment rates. Nonlinear Dyn. **96**(4), 2351–2368 (2019)
39. Goel, K., Nilam, : Stability behavior of a nonlinear mathematical epidemic transmission model with time delay. Nonlinear Dyn. **98**, 1501–1518 (2019)
40. Goel, K., Kumar, A., Nilam, : A deterministic time-delayed SVIRS epidemic model with incidences and saturated treatment. J. Eng. Math. **121**, 19–38 (2020)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/342676330

Numerical heat transfer analysis & predicting thermal performance of fins for a novel heat exchanger using machine learning

Article *in* Case Studies in Thermal Engineering · July 2020

DOI: 10.1016/j.csite.2020.100706

CITATIONS
0

READS
7

3 authors, including:



Gaurav Krishnayatra
Delhi Technological University
3 PUBLICATIONS **0** CITATIONS
SEE PROFILE



Sulekh Tokas
Delhi Technological University
2 PUBLICATIONS **0** CITATIONS
SEE PROFILE

All content following this page was uploaded by Sulekh Tokas on 14 July 2020.

The user has requested enhancement of the downloaded file.

Case Studies in Thermal Engineering 21 (2020) 100706

Contents lists available at ScienceDirect

Case Studies in Thermal Engineering

journal homepage: http://www.elsevier.com/locate/csite

Numerical heat transfer analysis & predicting thermal performance of fins for a novel heat exchanger using machine learning

Gaurav Krishnayatra *, Sulekh Tokas, Rajesh Kumar

Department of Mechanical Engineering, Delhi Technological University, Delhi, India

ARTICLE INFO

Keywords:
Convection heat transfer
Fin effectiveness
Fin efficiency
Novel fins
Machine learning
K-nearest neighbor

ABSTRACT

In the present case study, the thermal performance of fins for a novel axial finned-tube heat exchanger is investigated and predicted using machine learning regression technique. The effects of variation in the fin spacing, fin thickness, material, and the convective heat transfer coefficient on the overall efficiency and total effectiveness have been analyzed and commented upon. The *k*-Nearest Neighbor (*k*-NN), a machine learning algorithm, is used for regression analysis to predict the thermal performance outputs and the results showed high prediction accuracies. The *k*-NN algorithm is robust and precise which can be used by thermal system design engineers for predicting output variables. The temperature profiles of various geometries have been depicted and compared in the results. It was concluded that the efficiency is increasing with fin thickness & decreasing with fin spacing and the maximum efficiency $\eta_{max} = 0.99975$ is achieved at $\delta^* = 0.1$ & $t^* = 0.0133$ having $h = 5\ W/m.K$ for copper material. The effectiveness is increasing with fin spacing & fin thickness and the maximum effectiveness $\varepsilon_{max} = 122.766$ is for $\delta^* = 8$ & $t^* = 0.4$ having $h = 5\ W/m.K$.

1. Introduction

Extended surfaces (fins) are widely employed in various heat transfer components used in a wide range of applications like heat sinks, microprocessor cooling & latent thermal energy storage devices [1,2] and thus the fins have always been a notable domain for research. Increasing the exposed surface area by employing the fins is a prominent method to enhance the rate of heat transfer with the advantage of being economically feasible. Extensive studies are available which analyze the thermal performance and costing of different types of fins. Harper and Brown [3] presented two-dimensional analytical solutions for circumferential and longitudinal fins having rectangular and wedge type shape. It was deduced that the one-dimensional model is sufficient to evaluate the performance of the fin. Schmidt [4] examined the performance of various profiles for longitudinal fins and concluded that the parabolic profile is the optimum type of profile for longitudinal fins. Heggs & Ooi [5] evaluated the performance of radial rectangular fins and presented variations in thermal performance for different designs. Alam & Ghoshdastidar [6] performed numerical study using FDM on a circular tube with internal longitudinal fins having a tapered lateral profile, it was observed that there is a significant improvement in the net heat transfer rate with the installation of fins. Nagarani et al. [7] performed optimization using genetic algorithm on the geometrical parameters of the elliptical annular fin (EAF). The result showed that the EAF is more effective than the conventional circular fin for a

* Corresponding author.
E-mail address: gauravkrishnayatra_mt2k18@dtu.ac.in (G. Krishnayatra).

https://doi.org/10.1016/j.csite.2020.100706
Received 3 June 2020; Received in revised form 25 June 2020; Accepted 27 June 2020
Available online 4 July 2020
2214-157X/© 2020 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

fixed cross-sectional area irrespective of the efficiency when the shape factor (SF) is less than 0.5. Cuce & Cuce [8] used CFD to investigate the performance of novel rectangular fin designs with single-step change (RFSSC) and double-step change (RFDSC). The RFSSC fins were found to be more efficient than the conventional rectangular fin (CRF). Aziz and Khani [9] developed analytical solutions to examine the performance of rectangular and various convex parabolic fins installed on a rotating shaft using homotopy analysis method (HAM). The highest rate of heat transfer was recorded for uniform fin thickness and the effectiveness increased with the rotational speed of the shaft. Okiy [10] investigated the effect of two-dimensional heat transfer in radial and planar fins in comparison to the assumption of one-dimensional heat transfer in fin analysis. It was concluded that one-dimensional heat transfer assumption provided inaccurate solutions with respect to two-dimensional heat transfer for short planar fins. Samana et al. [11] experimentally examined the fin efficiency of a solid wire fin used in a wire-on-tube heat exchanger and the fin efficiency increased with the use of oscillating heat pipe in comparison to solid wire fin. Li et al. [12] performed experiments to study the effects of various thermal boundary conditions and thermal conductivity on the thermal performance of a pin fin array using liquid crystal sheet, and it was concluded that thermal conductivity is directly influencing the rate of heat transfer.

After a thorough literature review, it was deduced that some designs of finned-tube heat exchangers are particularly overlooked and many studies have utilized the artificial neural network [13,14] however, very limited studies have utilized the *k*-Nearest Neighbor for regression analysis. A novel type of fins is designed and its thermal performance has been analyzed in this computational investigation. The finned-tube heat exchanger consists of a hollow tube with an isothermal boundary condition at the inner wall (in order to simulate some of the industrial applications) having the novel rectangular fins attached axially at the outer wall. The novel design of fins consists of a fin-on-fin approach, where in Fig. 1 it is visible that the tube has primary fins (colored in black), this is a simple case of longitudinal fins attached to the outer surface of the tube. But consequently another set of fins, the secondary fins (colored in blue), are attached on to the primary fins in order to increase the total finned area that proportionately enhance the total rate of heat transfer. In this study, the variations of the overall Efficiency and the Effectiveness of the novel finned-tube system has been analyzed with respect to the varying fin thickness, fin spacing, material of the system, and the convective heat transfer coefficient using Ansys Fluent. Moreover, Machine learning technique called *k*-Nearest Neighbor (*k*-NN) has been employed for regression analysis to predict the thermal performance of the novel finned-tube system.

2. Physical model & governing equations

The design consists of a tube with the inner diameter (D_i) and the outer diameter (D_o) of 5 & 6 mm respectively with the fin thickness (t) being varied from 0.2 to 2.4 mm and having the fin spacing (δ) from 0.6 to 48 mm as shown in Fig. 1. The two materials of the finned-tube system chosen are Copper $(K_{cu} = 387.6\ W/m.K)$ and Steel $(K_{ss} = 16.7\ W/m.K)$. The conjugated heat transfer is taking place from the inner isothermal wall $(T_w = 400\ K)$ to the ambient air $(T_\infty = 300\ K)$. To reduce computational cost and time, the simulated flow of air on the outer surface of the finned-tube system has not been considered. However, a range of convective heat transfer coefficient has been taken which covers almost all the practical applicated values found in heat exchangers in order to visualize the impact of its variation on the efficiency and the effectiveness of the system. The convective heat transfer coefficient (h) at the outer surface is being varied from 5 to 200 W/m.K.

The governing partial differential equations used to solve the conjugated heat transfer problem are linearized and computed in the Ansys Fluent using with the following assumptions.

1. The material is homogenous and isotropic.
2. The fin tips are adiabatic.
3. The convective heat transfer coefficient is uniform over the entire outer surface area.
4. The radiation phenomenon is neglected.

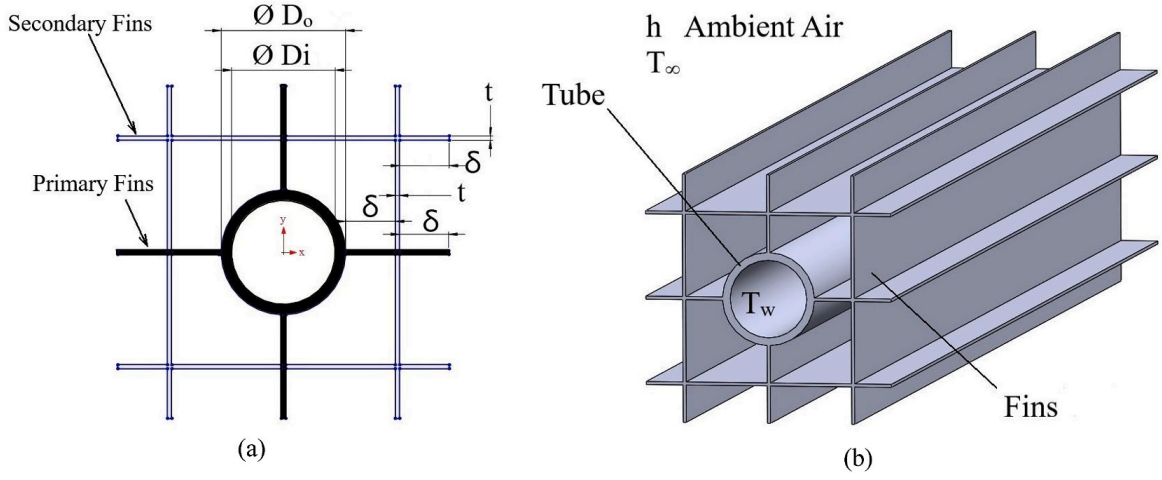


Fig. 1. a) Schematic model of the novel finned-tube heat exchanger b) 3-Dimensional Representation of the model.

The energy equation for the finned-tube system [1]:

$$\left\{\frac{d^2T}{dx^2}+\frac{d^2T}{dy^2}\right\}=0 \tag{1}$$

The heat transfer equation for the outer surface [1]:

$$Q=\frac{(T_w-T_\infty)}{\frac{\ln\left(D_o/D_i\right)}{2\pi kL}+\frac{1}{\eta_o A_t h}} \tag{2}$$

Where, k is the thermal conductivity of the material, h is the convective heat transfer coefficient and η_o is the overall efficiency of the finned-tube [1].

$$\eta_o=1-\frac{NA_f}{A_t}\left(1-\eta_f\right) \tag{3}$$

Where, η_f is the efficiency of a single fin, A_f the surface area of single fin and A_t the total surface area of the finned-tube system. The total surface area of the novel finned-tube system is calculate by:

$$A_t=L_{exp}\times L \tag{4}$$

$$L_{exp}=48\delta+8(D_o-t) \tag{5}$$

Where, axial length (L) of the system is unity and L_{exp} is the total perimetric length of the outer surface of the novel finned-tube system.

Equations (2) and (3) are used to validate the computational results for a simple finned-tube system having 4 equidistant axial fins attached at the outer surface of a hollow tube having same diameters with fin length $L_f = 10$ *mm* and fin thickness $t = 0.5$ *mm* as shown in Fig. 2 along with the heat flux variation with convective heat transfer coefficient which is showing remarkable agreement.

The overall Efficiency for the system is given as:

$$\eta=\frac{q}{h(T_w-T_\infty)} \tag{6}$$

Where, q is the area weighted average of the heat flux at the outer surface of the novel finned-tube system computed from the simulations. And the total effectiveness of the system is:

$$\varepsilon=\frac{Q}{hA_o(T_w-T_\infty)} \tag{7}$$

Where, Q is the area surface integral of the heat flux at the outer surface computed by simulation.

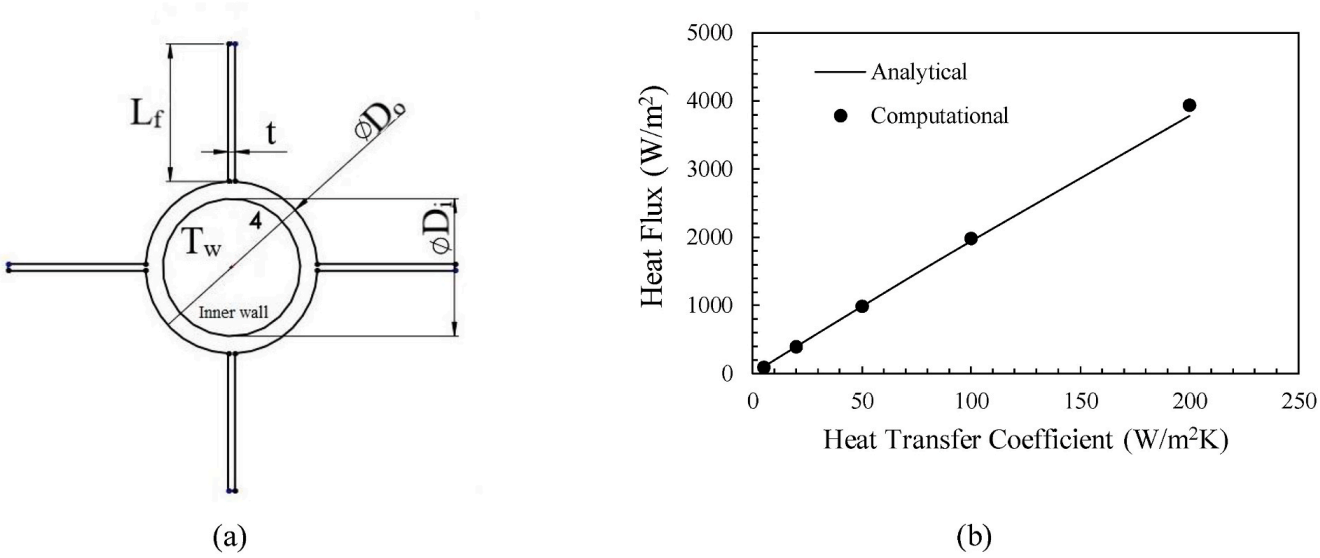


Fig. 2. a) Schematic finned-tube system used for convective heat transfer validation with 4 axial fins at the outer surface of the tube b) Variation of heat flux with convective heat transfer coefficient representing the comparing between analytical and computational heat flux values.

3. Computational & machine learning modelling

3.1. Computational modelling

A 2-Dimensional analysis is performed in this study in order to reduce the computational time. Moreover, the 2D model of the finned-tube system is symmetrical about 4 axes, therefore the $1/8^{th}$ 2D model is selected for the study as shown in Fig. 3. The 2D models were developed using Solidworks software and then were modeled & solved in Ansys Fluent. The models were discretized into a finite number of control volumes on which the governing equations were solved using appropriate approximations. Second-order upwind was used for energy equation and the residual value was set to 10^{-9}. Tetrahedron meshing was done on the solid domain as shown in Fig. 3 and the mesh convergence test was also performed to make the result independent of the size of the elements. Table 1 is showing the variation of heat flux with the element size and it can be inferred that the heat flux is getting independent when the size of the element is at 0.0125 mm.

3.2. Machine learning modelling

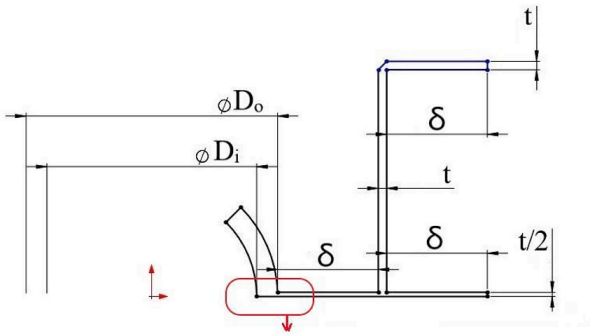
This section explains the utilization of supervised machine learning technique called *k*-Nearest Neighbor (*k*-NN) which has been applied as a regression tool [15]. The *k*-NN algorithm is robust, precise and computationally fast when used as a regression model, moreover it works well with a fewer number of input variables and relatively small data set which is used in designing of thermal systems. The dataset for the regression analysis is procured from the simulations result for all the geometries considered in the study, the dataset includes the output variables namely the efficiency (η) & the effectiveness (ε) of the finned-tube heat exchanger and the outer diameter of tube (D_o), fin spacing (δ), fin thickness (t), thermal conductivity of the material (k), and the convective heat transfer coefficient (h) are the input variables. Therefore, the variables are represented in the forms of following functions.

$$\eta = f_1(D_o,\ \delta,\ t,\ k,\ h) \tag{8}$$

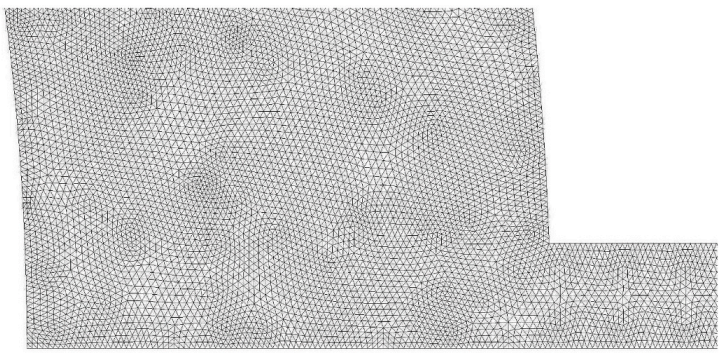
$$\varepsilon = f_2(D_o,\ \delta,\ t,\ k,\ h) \tag{9}$$

In order to dimensionally balance the equation, the input variables are converted into non-dimensional forms subsequently decreasing the number of input variables.

$$\eta = f_3(\delta^*,\ t^*,\ M) \tag{10}$$



(a)



(b)

Fig. 3. a) $1/8^{th}$ model designed for the simulation b) The red encircled portion of the $1/8^{th}$ model is shown after descritization

Table 1
Variation of heat flux with number of elements and element size.

Max. Element Size (mm)	Number of Elements	Heat Flux $q(W/m^2)$	Difference in q (%)
0.05	31392	498.187	0.212
0.04	51936	497.129	1.913
0.03	96128	487.796	2.154
0.02	191072	498.536	-0.266
0.015	344592	497.210	0.057
0.0125	461472	497.494	-0.012
0.01	601844	497.555	–

$$\varepsilon = f_4(\delta^*,\ t^*,\ M) \tag{11}$$

Where, δ^* is the non-dimensional fin spacing, t^* is the non-dimensional fin thickness and M is a new non-dimensional parameter used for this study, relations given below:

$$\delta^* = \frac{\delta}{D_o} \tag{12}$$

$$t^* = \frac{t}{D_o} \tag{13}$$

$$M = \frac{h}{8k} L_{exp} \tag{14}$$

The k-NN algorithm is a "black-box" regression tool where the actual functions are not derived instead they are predicted from the dataset. The dataset is divided into two parts: 65% Testing data & 35% Training data, the training dataset is used for "learning" and the testing dataset is used for validating the predictions with actual output values. The prediction accuracy of k-NN regression is calibrated by two factors namely, the "k" number of close neighbors and the type of metric distance between the outputs [16]. In Fig. 4, the output y_i is scattered with respect to the inputs x_i & z_i for $k = 3$ number of neighbors. The p_i is the prediction computed on the basis of the median of Euclidean distances $(d_{12},\ d_{13}\ \&\ d_{14})$ of y_1 from $y_2,\ y_3\ \&\ y_4$, which are the 3 closest neighboring outputs. The Euclidean distance is calculated by

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (z_i - z_j)^2} \tag{15}$$

Where, d_{ij} calculates the Euclidean distance between the ith and jth elements in the input data.

The predicted values of the outputs are compared to the computed outputs from the simulation results in terms of R^2 values to check the accuracy of the regression model and the selection of "k" is based on the accuracy of the regression model. The R^2 value is the

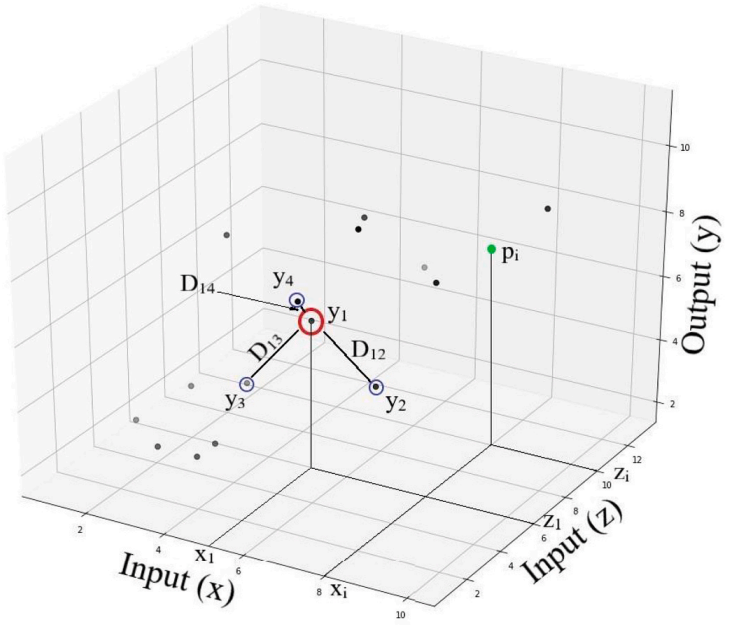


Fig. 4. Three-Dimensional scattering of output variable shown with variation of inputs 1 & 2 for $k = 3$ number of neighbors.

coefficient of determination which is given by the following relation.

$$R^2 = 1 - \frac{\sum_i (y_i - p_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{16}$$

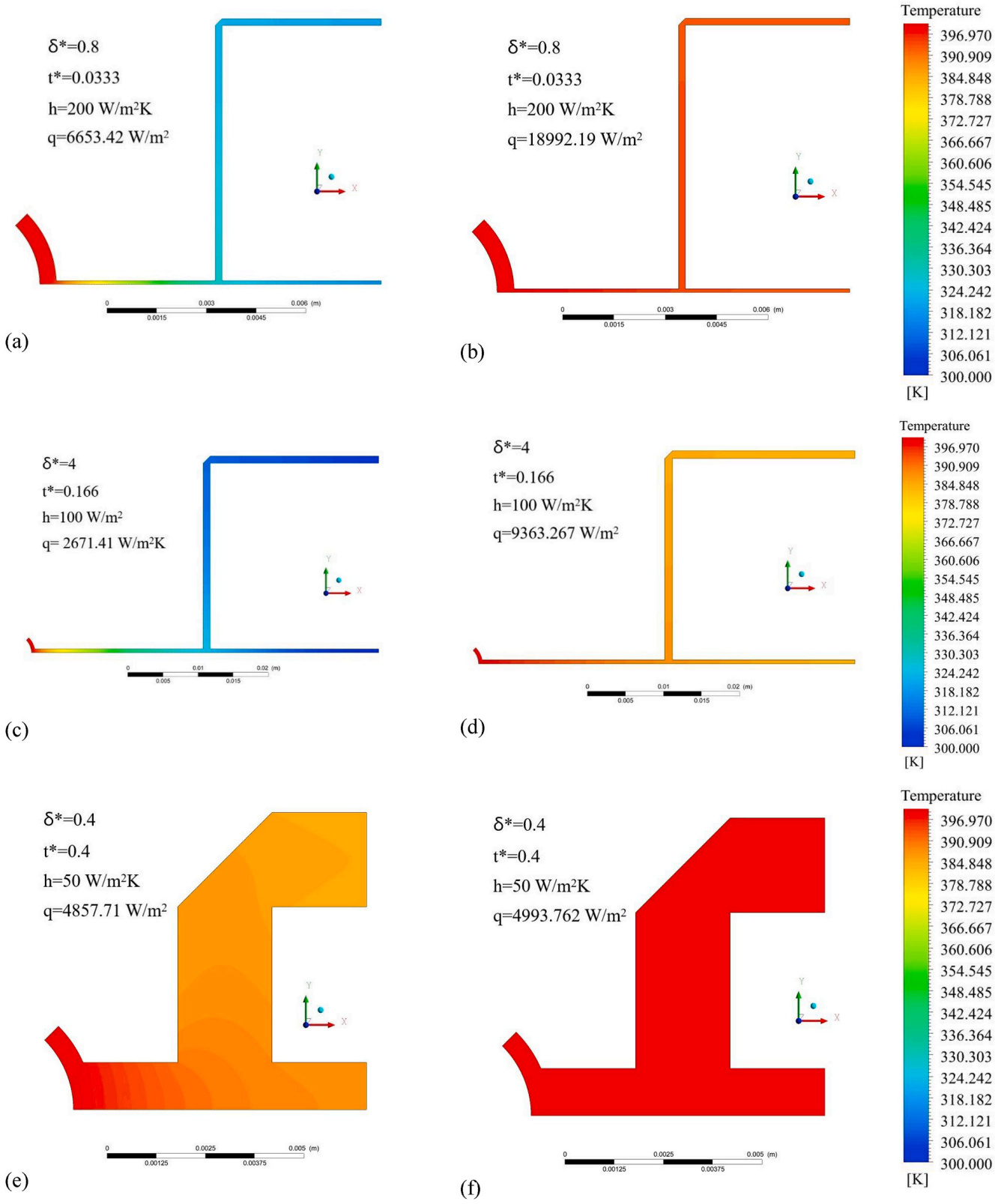
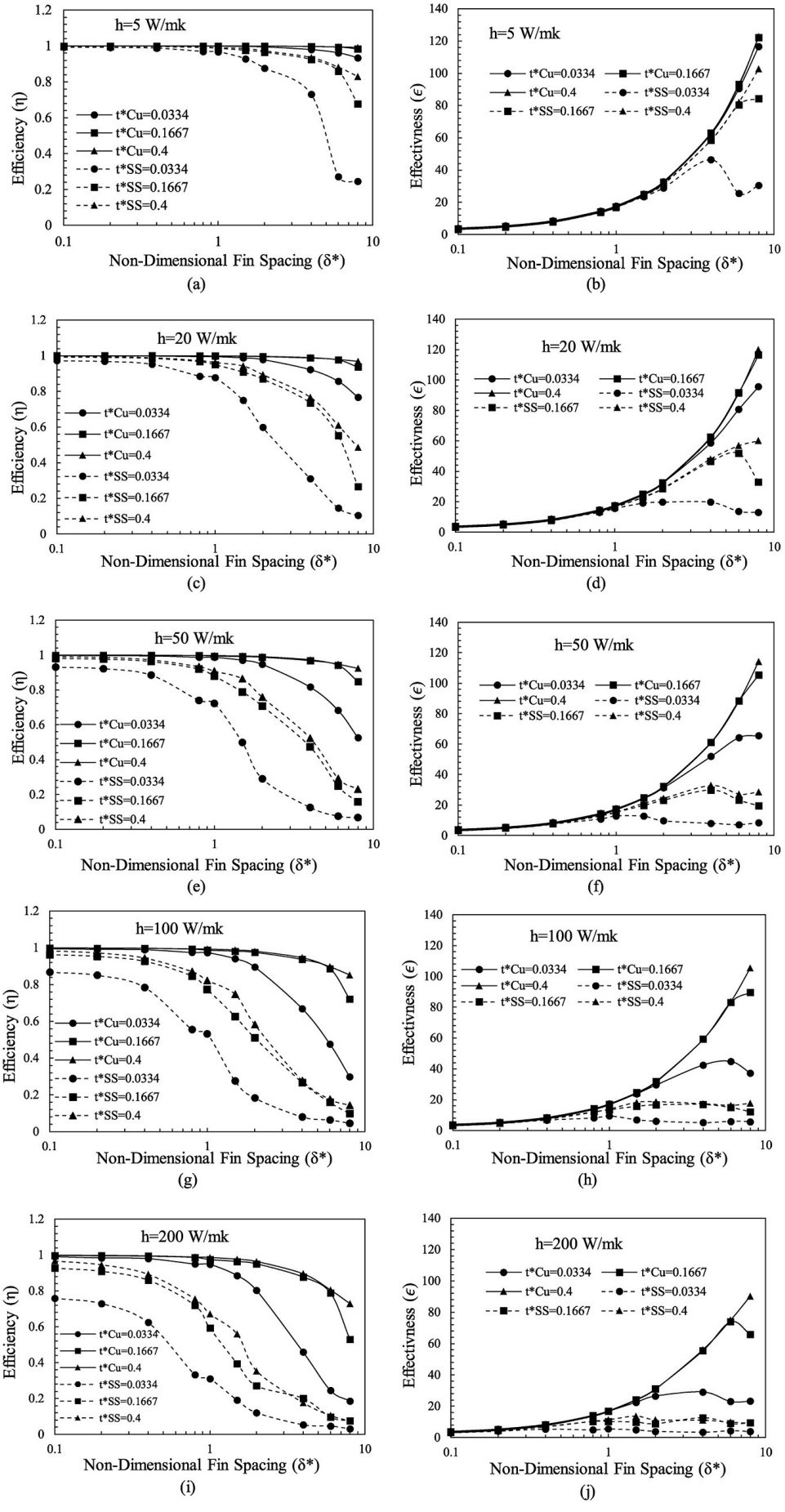


Fig. 5. Temperature plumes of cross-section for different geometries and different convective heat transfer coefficients.



(caption on next page)

Fig. 6. Thermal performance of the novel finned-tube heat exchangers with respect to non-dimensional fin spacing for various non-dimensional fin thickness and convective heat transfer coefficients.

Where, y_i & p_i are the output variable and the corresponding predicted value for i^{th} element in the data and $\bar{y}$ is the mean.

4. Results & discussion

This study provides insights which explain the variations of the overall efficiency (η) and the total effectiveness (ε) of a novel finned-tube system with respect to the fin spacing, the fin thickness, thermal conductivity of the material and the convective heat transfer coefficient. The number of parameters is reduced and converted to non-dimensional forms. Therefore, the non-dimensional fin spacing $\left(\delta^* = \frac{\delta}{D_o}\right)$, non-dimensional fin thickness $\left(t^* = \frac{t}{D_o}\right)$ & the convective heat transfer coefficient is being varied from 0.1 to 8, from 0.0333 to 0.4 and from 5 to 200 W/m^{2}K respectively for copper and steel fins. Fig. 5 shows the temperature profiles having same temperature bound for various geometrical models along with the heat flux values. It can be inferred that the system having copper material is showing more rate of heat transfer due to the fact that copper has higher thermal conductivity. The temperature gradients developed along the fin length for copper fins is lower than that for steel fins, consequently increasing the efficiency of copper fins. The drop in temperature along the fins length is increasing with the increase in fin thickness. In Fig. 6, the overall efficiency and total effectiveness of the novel finned-tube system is represented with respect to the δ^* for different values of t^*. For steel fins the decrease in fin efficiency is gradual till $\delta^* = 1$, the it is decreasing rather quickly with δ^* in contrast to copper fins. Moreover, the rate of decrease of efficiency is also directly proportional to the convective heat transfer coefficient. It can be observed that for same values of δ^* & t^*, higher values of h are showing larger decrease in the efficiency. It can be asserted that overall efficiency of finned-tube system decreases with increase in δ^* & h, and it increases with increase in t^*. The maximum efficiency $\eta_{max} = 0.99975$ is achieved at $\delta^* = 0.1$ & $t^* = 0.0133$ for $h = 5\ W/m.K$ with copper material. The total effectiveness of the system for copper fins is seen to be insensitive to the

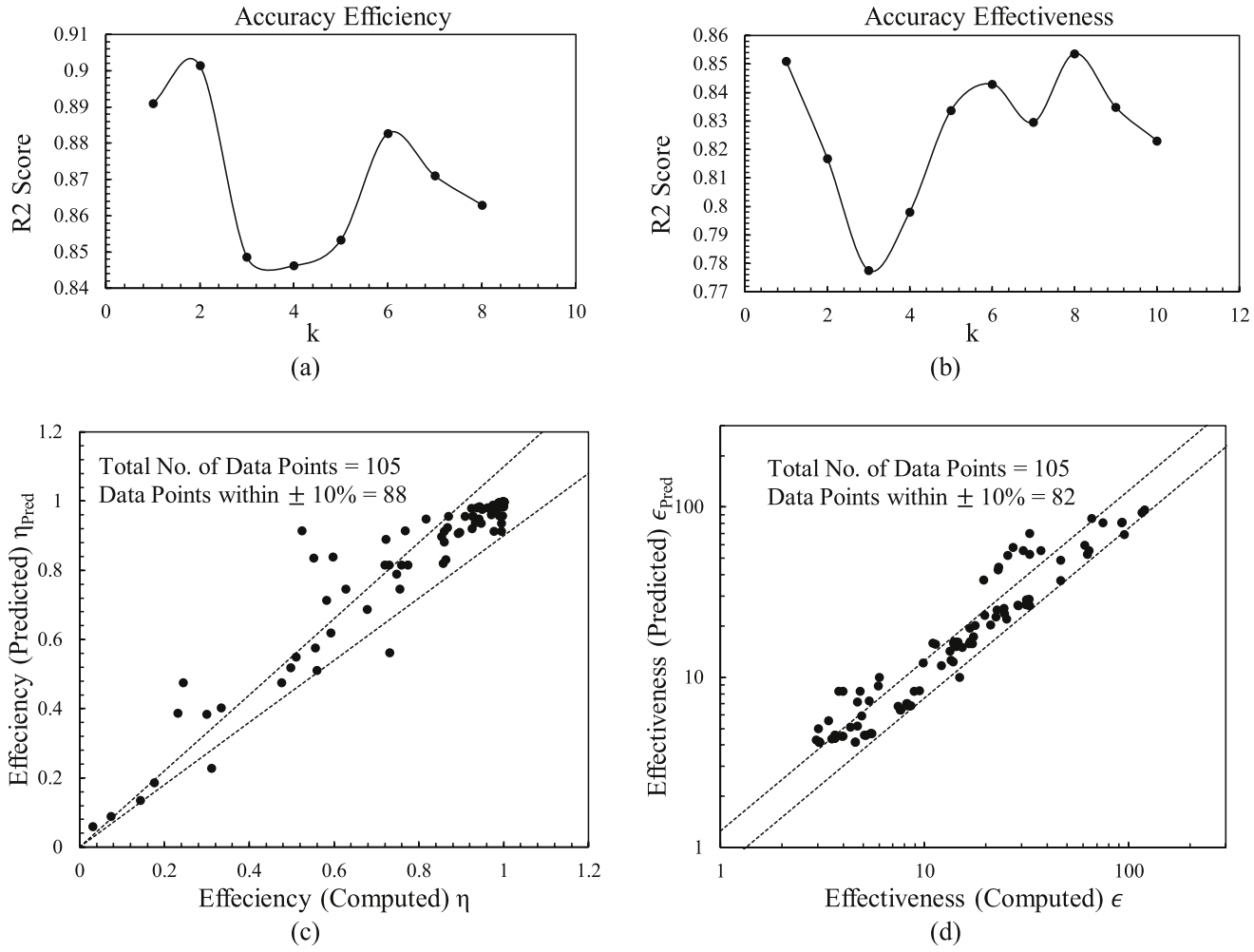


Fig. 7. Results of k-NN regression showing variation of R^2 score with k value (a & b) along with predicted and computed outputs for efficiency and effectiveness (c & d).

fin thickness for $h = 5$ & 20 W/m^2K, however, it is decreasing with the increase in δ^* for $h = 50$, 100 & 200 W/m^2K. The effectiveness of steel fins is observed to be decreasing with δ^* for all values of h. It can be concluded that the total effectiveness is increasing with increase in δ^* & t^* and is decreasing with h. The maximum effectiveness $\varepsilon_{max} = 122.766$ is for $\delta^* = 8$ & $t^* = 0.4$ for $h = 5\ W/\ m.K$.

The variation of R^2 values with "k" has been recorded and presented in Fig. 7 and it is clear that $k = 2$ is predicting the highest accuracy for Efficiency and $k = 8$ is giving the highest accuracy for Effectiveness predictions. The R^2 value for efficiency at $k = 2$ is 0.9014 and for effectiveness at $k = 8$ is 0.8537, which implies that the regression model predictions are in good agreement with the actual computed outputs as shown in Fig. 7 c & d. It can be observed that for Efficiency, 88 out of 105 data points are lying within the 10% error range and for effectiveness, 85 out of 105 data points are lying with the 10% error range. The k-NN regression algorithm is showing remarkable accuracies therefore, it is an exceptional tool for thermal system designer for predicting outputs without computationally solving the thermal system at its entirety.

5. Conclusion

The convective heat transfer coefficient (h) is being varied along with variation in fin geometry $(\delta$ & $t)$ for two materials copper and steel to analyze the effects on the thermal performance of this novel finned-tube heat exchanger. It is concluded that copper is the obvious choice of material for the finned-tube system. The efficiency has direct proportionality with fin thickness and inverse proportionality with fin spacing & heat transfer coefficient. The maximum efficiency is achieved with the shortest & thickest fin for the lowest heat transfer coefficient. The effectiveness is showing direct relation with fin spacing & fin thickness and inverse relation with the heat transfer coefficient. The maximum overall effectiveness of the system is achieved for the longest and thickest fin with the lowest heat transfer coefficient. The k-NN regression model resulted in high prediction accuracies and because of its advantages, k-NN can be an extremely useful tool for thermal system design engineers.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Gaurav Krishnayatra: Data curation, Conceptualization, Writing - original draft. **Sulekh Tokas:** Conceptualization, Writing - review & editing. **Rajesh Kumar:** Supervision, Writing - review & editing.

Nomenclature

D_o	Outer diameter of the tube, m
D_i	Inner diameter of the tube, m
δ	Fin spacing, m
L_f	Length of fin, m
t	Thickness of fin, m
T	Temperature (general), $^\circ C$
T_w	Temperature of inner wall, $^\circ C$
T_∞	Temperature of the free stream, $^\circ C$
L	Axial length of the hollow cylinder, m
t^*	Non-dimensional fin thickness, $t^* = t/D_o$
δ^*	Non-dimensional fin spacing, $\delta^* = \delta/D_o$
A_t	Total surface area, m^2
h	Convective heat transfer coefficient, $W/m^2.K$
q	Heat transfer per unit area, simulations result, W/m^2
Q	Total rate of heat transfer, W, $Q = \frac{(T_w - T_\infty)}{\frac{\ln\left(D_o/D_i\right)}{2\pi kL} + \frac{1}{\eta_o A_t h}}$
k_{ss}	Thermal conductivity of Steel, $W/m.K$
k_{cu}	Thermal conductivity of Copper, $W/m.K$
k	Thermal conductivity (general), $W/m.K$
η_o	Overall Efficiency of finned-tube system, $\eta_o = 1 - \frac{NA_f}{A_t}(1 - \eta_f)$
η_f	Efficiency of single fin
N	Number of Fins
A_f	Area of single fins, m^2
L_{exp}	Outer surface perimetric length of the finned-tube system, m, $L_{exp} = 112\delta + 16(D_o - t)$

η	Efficiency computed by computational results, $\eta = \frac{q}{h(T_w - T_\infty)}$
ε	Effectiveness computed by computational results, $\varepsilon = \frac{Q}{hA_o(T_w - T_\infty)}$
M	Non-dimensional parameter "M", $M = \frac{h}{8k}L_{exp}$
η_{pred}	Predicted Efficiency
ε_{pred}	Predicted Effectiveness
η_{max}	Maximum value of the Efficiency
ε_{max}	Maximum value of the Effectiveness
d_{ij}	Euclidean distance between ith and jth elements of the predictors, $d_{ij} = \sqrt{(x_i - x_j)^2 + (z_i - z_j)^2}$
x_i, z_i	Predictor variables for ith element
k	Number of neighbors used in k-NN algorithm
y_i	ith Output variable used for regression
p_i	Predicted Output for ith element
R^2	R-squared values, $R^2 = 1 - \frac{\sum_i (y_i - p_i)^2}{\sum_i (y_i - \bar{y})^2}$

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csite.2020.100706.

References

[1] A.D. Kraus, A. Aziz, J. Welty, Extended Surface Heat Transfer, John Wiley & sons, 2001.
[2] N. Nagarani, K. Mayilsamy, A. Murugesan, G.S. Kumar, Review of utilization of extended surfaces in heat transfer problems, Renew. Sustain. Energy Rev. 29 (2014) 604–613, https://doi.org/10.1016/j.rser.2013.08.068.
[3] D.R. Harper, W.B. Brown, Mathematical Equations for Heat Conduction in the Fins of Air Cooled Engines, vol. 158, National Advisory Committee on Aeronautics, Washington, DC, 1922. NACA Rep. https://ntrs.nasa.gov/search.jsp?R=19930091223.
[4] E. Schmidt, Die warmeuebertragung durch rippen, Z. Des. Vereines Dtsch. Ingenieure 70 (1926) 885–951.
[5] P.J. Heggs, T.H. Ooi, Design charts for radial rectangular fins in terms of performance ratio and maximum effectiveness, Appl. Therm. Eng. 24 (2004) 1341–1351, https://doi.org/10.1016/j.applthermaleng.2003.12.021.
[6] A. Alam, P.S. Ghoshdastidar, A study of heat transfer effectiveness of circular tubes with internal longitudinal fins having tapered profiles, Int. J. Heat Mass Tran. 45 (2002) 1371–1376, https://doi.org/10.1016/S0017-9310(01)00240-X.
[7] N. Nagarani, K. Mayilsamy, A. Murugesan, G.S. Kumar, Review of utilization of extended surfaces in heat transfer problems, Renew. Sustain. Energy Rev. 29 (2014) 604–613, https://doi.org/10.1016/j.proeng.2012.06.343.
[8] P.M. Cuce, E. Cuce, Optimization of configurations to enhance heat transfer from a longitudinal fin exposed to natural convection and radiation, Int. J. Low Carbon Technol. 9 (2014) 305–310, https://doi.org/10.1093/ijlct/ctt005.
[9] A. Aziz, F. Khani, Analytical solutions for a rotating radial fin of rectangular and various convex parabolic profiles, Commun. Nonlinear Sci. Numer. Simulat. 15 (2010) 1565–1574, https://doi.org/10.1016/j.cnsns.2009.07.008.
[10] R.E.K.V. Okiy, An assessment of extended surfaces-two dimensional effects, Int. J. Eng. Res. Afr. 15 (2015) 71–85. https://doi.org/10.4028/www.scientific.net/JERA.15.71.
[11] T. Samana, T. Kiatsiriroat, A. Nuntaphan, Enhancement of fins efficiency of a solid wire fin by oscillating heat pipe under forced convection, Case Studies Thermal Engg 2 (2014) 36–41, https://doi.org/10.1016/j.csite.2013.10.003.
[12] W. Li, L. Yang, J. Ren, H. Jiang, Effect of thermal boundary conditions and thermal conductivity on conjugate heat transfer performance in pin fin arrays, Int. J. Heat Mass Tran. 95 (2016) 579–592, https://doi.org/10.1016/j.ijheatmasstransfer.2015.12.010.
[13] M. Ramezanizadeh, M.A. Nazari, Modeling thermal conductivity of Ag/water nanofluid by applying a mathematical correlation and artificial neural network, Int. J. Low Carbon Technol. 14 (2019) 468–474, https://doi.org/10.1093/ijlct/ctz030.
[14] A. Maleki, A. Haghighi, M.I. Shahrestani, Z. Abdelmalek, Applying different type of artificial neural network for modelling thermal conductivity of nanofluids containing silica particles, J. Therm. Anal. Calorim. (2020), https://doi.org/10.1007/s10973-020-09541-x.
[15] P. Harrington, Machine Learning in Action, Manning Publications Co., New York, USA, 2012.
[16] N.S. Altman, An introduction to kernel and nearest-neighbor nonparametric regression, Am. Statistician 46 (1992) 175–185, https://doi.org/10.1080/00031305.1992.10475879.

View publication stats

arXiv:2007.06069v1 [math.CV] 12 Jul 2020

ON CERTAIN GENERALIZATIONS OF $\mathcal{S}^*(\psi)$

S. SIVAPRASAD KUMAR AND KAMALJEET GANGANIA

Abstract. We deal with different kinds of generalizations of $\mathcal{S}^*(\psi)$, the class of Ma-Minda starlike functions, in addition to a majorization result of $\mathcal{C}(\psi)$, the class of Ma-Minda convex functions, which are enlisted as follows: 1. Let h be an analytic function, f be in $\mathcal{C}(\psi)$ and h be majorized by f in the unit disk $\mathbb{D}$, then for a given ψ, we derive a general equation, which yields the radius constant r_ψ such that $|h'(z)| \leq |f'(z)|$ in $|z| \leq r_\psi$. Consequently, obtain results associating $\mathcal{S}^*(\psi)$ and others. 2. We find the largest radius r_0 so that the product function $g(z)h(z)/z$ belongs to a desired class for $|z| < r_0$ whenever $g \in \mathcal{S}^*(\psi_1)$ and $h \in \mathcal{S}^*(\psi_2)$. Also we obtain a condition for the functions to be in $\mathcal{S}^*(\psi)$ 3. We obtain the modified distortion theorem for $\mathcal{S}^*(\psi)$ with a general perspective. 4. For a fixed $f \in \mathcal{S}^*(\psi)$, the class of subordinants $S_f(\psi) := \{g : g \prec f\}$ is introduced and studied for the Bohr-phenomenon and a couple of conjectures are also proposed.

2010 AMS Subject Classification. Primary 30C80, Secondary 30C45.
Keywords and Phrases. Subordination, Bohr-Radius, Majorization, Distortion theorem.

1. Introduction

Let $\mathcal{A}$ denote the class of analytic functions of the form $f(z) = z + \sum_{k=2}^{\infty} a_k z^k$ in the open unit disk $\mathbb{D} := \{z : |z| < 1\}$. Using subordination, Ma-Minda [13] introduced the unified class of starlike and convex functions defined as follows:

$$\mathcal{S}^*(\psi) := \left\{ f \in \mathcal{A} : \frac{zf'(z)}{f(z)} \prec \psi(z) \right\} \text{ and } \mathcal{C}(\psi) := \left\{ f \in \mathcal{A} : 1 + \frac{zf''(z)}{f'(z)} \prec \psi(z) \right\}, \tag{1.1}$$

where ψ is a Ma-Minda function, which is analytic and univalent with $\Re\psi(z) > 0$, $\psi'(0) > 0$, $\psi(0) = 1$ and $\psi(\mathbb{D})$ is symmetric about real axis. Note that $\psi \in \mathcal{P}$, the class of normalized Carathéodory functions. The class $\mathcal{S}^*(\psi)$ unifies various subclasses of starlike functions, which are obtained for an appropriate choice of ψ. Ma-Minda discussed many properties of the class $\mathcal{S}^*(\psi)$, in particular, they proved the distortion theorem [13, Theorem 2, p.162] with some restriction on ψ, namely

$$\min_{|z|=r} |\psi(z)| = \psi(-r) \quad \text{and} \quad \max_{|z|=r} |\psi(z)| = \psi(r). \tag{1.2}$$

In section 4, we modify the distortion theorem by relaxing this restriction on ψ to obtain a more general result. In 1914, Harald Bohr [5] proved the following remarkable result related to the power series:

Theorem 1.1 ([5]). *Let $g(z) = \sum_{k=0}^{\infty} a_k z^k$ be an analytic function in $\mathbb{D}$ and $|g(z)| < 1$ for all $z \in \mathbb{D}$, then $\sum_{k=0}^{\infty} |a_k| r^k \leq 1$ for all $z \in \mathbb{D}$ with $|z| = r \leq 1/3$.*

The work of the second author is supported by University Grant Commission, New-Delhi, India under UGC-Ref. No.:1051/(CSIR-UGC NET JUNE 2017).

Bohr actually proved the above result for $r \leq 1/6$. Further Wiener, Riesz and Shur independently sharpened the result for $r \leq 1/3$. Presently the Bohr inequality for functions mapping unit disk onto different domains, other than unit disk is an active area of research. For the recent development on Bohr-phenomenon, see the articles [1, 2, 3, 4, 10, 11, 17, 18, 19, 20, 22] and references therein. The concept of Bohr-phenomenon in terms of subordination can be described as: Let $f(z) = \sum_{k=0}^{\infty} a_k z^k$ and $g(z) = \sum_{k=0}^{\infty} b_k z^k$ are analytic in $\mathbb{D}$ and $f(\mathbb{D}) = \Omega$. For a fixed f, consider a class of analytic functions $S(f) := \{g : g \prec f\}$ or equivalently $S(\Omega) := \{g : g(z) \in \Omega\}$. Then the class $S(f)$ is said to satisfy Bohr-phenomenon, if there exists a constant $r_0 \in (0, 1]$ satisfying the inequality $\sum_{k=1}^{\infty} |b_k| r^k \leq d(f(0), \partial\Omega)$ for all $|z| = r \leq r_0$ and $g(z) \in S(f)$, where $d(f(0), \partial\Omega)$ denotes the Euclidean distance between $f(0)$ and the boundary of $\Omega = f(\mathbb{D})$. The largest such r_0 for which the inequality holds, is called the Bohr-radius.

In 2014, Muhanna et al. [19] proved the Bohr-phenomenon for $S(W_\alpha)$, where $W_\alpha := \{w \in \mathbb{C} : |\arg w| < \alpha\pi/2, 1 \leq \alpha \leq 2\}$, which is a Concave-wedge domain (or exterior of a compact convex set) and the class $R(\alpha, \beta, h)$ defined by $R(\alpha, \beta, h) := \{g \in \mathcal{A} : g(z) + \alpha z g'(z) + \beta z^2 g''(z) \prec h(z)\}$, where h is a convex function (or starlike) and $R(\alpha, \beta, h) \subset S(h)$. In 2018, Bhowmik and Das [3] proved the Bohr-phenomenon for the classes given by $S(f) = \{g \in \mathcal{A} : g \prec f \text{ and } f \in \mu(\lambda)\}$, where $\mu(\lambda) = \{f \in \mathcal{A} : |(z/f(z))^2 f'(z) - 1| < \lambda, 0 < \lambda \leq 1\}$ and $S(f) = \{g \in \mathcal{A} : g \prec f \text{ and } f \in \mathcal{S}^*(\alpha), 0 \leq \alpha \leq 1/2\}$, where $\mathcal{S}^*(\alpha)$ is the well-known class of starlike functions of order α. In Section 5, for any fixed $f \in \mathcal{S}^*(\psi)$, we introduce and study the Bohr-phenomenon inside the disk $|z| \leq 1/3$ for the following class of analytic subordinants:

$$S_f(\psi) := \left\{ g(z) = \sum_{k=1}^{\infty} b_k z^k : g \prec f \right\}. \tag{1.3}$$

Note that $\mathcal{S}^*(\psi) \subset \bigcup_{f \in \mathcal{S}^*(\psi)} S_f(\psi)$. As an application, we obtain the Bohr-radius for the class $S(f)$, where $f \in \mathcal{S}^*((1 + Dz)/(1 + Ez))$, the class of Janowski starlike functions, with some additional restriction on D and E apart from $-1 \leq E < D \leq 1$. Now recall the following definition and a result due to T. H. MacGregor [14]:

Definition 1.2. ([14]) Let f and g be analytic in $\mathbb{D}$. A function $g(z)$ is said to be majorized by $f(z)$, denoted by $g << f$, if there exists an analytic function $\Phi(z)$ in $\mathbb{D}$ satisfying $|\Phi(z)| \leq 1$ and $g(z) = \Phi(z) f(z)$ for all $z \in \mathbb{D}$.

Theorem 1.3 ([14])**.** *Let g be majorized by f in $\mathbb{D}$ and $g(0) = 0$. If $f(z)$ is univalent in $\mathbb{D}$, then $|g'(z)| \leq |f'(z)|$ in $|z| \leq 2 - \sqrt{3}$. The constant $2 - \sqrt{3}$ is sharp.*

Recently Tang and Deng [25] obtained the majorization results for $\mathcal{S}^*(\psi)$ for some specific choices of ψ, motivated by this in section 2, we devise a general approach to handle the same for $\mathcal{C}(\psi)$, which is precisely stated as: if $g \in \mathcal{A}$, $f \in \mathcal{C}(\psi)$ and g is majorized by f in $\mathbb{D}$, then we find the largest radius $r_\psi \leq 1$ such that $|g'(z)| \leq |f'(z)|$ in $|z| \leq r_\psi$. Several other results in this direction are also obtained. In section 3, we consider the radius problem posed by Obradović and Ponnusamy [21] namely: Let $g \in \mathcal{S}^*(\psi_1)$ and $h \in \mathcal{S}^*(\psi_2)$, then find the largest radius $r_0 \leq 1$ such that the function $F(z) = (g(z)h(z))/z$ belongs to certain well-known class of starlike functions in $|z| < r_0$. As a special case, we also obtain a result of Obradović and Ponnusamy [21]. Further we obtain the condition for functions to be in $\mathcal{S}^*(\psi)$ which is an extention of the Bulboacă and Tuneski [6]. Throughout this paper we shall assume that the function ϕ has real coefficients in it's power series expansion.

2. Majorization

Let us consider the analytic function $\psi(z) := 1+B_1z+B_2z^2+\cdots$. Here $B_1 = \psi'(0)$, the coefficient of z, plays a major role in deciding the orientation of the function ψ. Thus ψ is positively or negatively oriented depends on whether B_1 is positive or negative. Ma-Minda only considered the case $\psi'(0) > 0$, as it may be possible that for the case when $\psi'(0) < 0$, many postulates for the class $\mathcal{S}^*(\psi)$ need not remain same. With this perspective, we begin with the following:

Theorem 2.1. *Let $\Re\phi(z) > 0$ and ϕ be convex in $\mathbb{D}$ with $\phi(0) = 1$. Suppose ψ be the function such that $m_r := \min\limits_{|z|=r}|\psi(z)|$ and also satisfies the differential equation*

$$\psi(z) + \frac{z\psi'(z)}{\psi(z)} = \phi(z). \tag{2.1}$$

Let $g \in \mathcal{A}$ and $f \in \mathcal{C}(\phi)$. If g is majorized by f in $\mathbb{D}$, then

$$|g'(z)| \leq |f'(z)| \quad \text{in } |z| \leq r_\psi, \tag{2.2}$$

where r_ψ is the least positive root of the equation

$$(1-r^2)m_r - 2r = 0. \tag{2.3}$$

Proof. Let us define $p(z) := zf'(z)/f(z)$. Since $f \in \mathcal{C}(\phi)$, therefore we have $1 + zf''(z)/f'(z) \prec \phi(z)$, which can be equivalently written as

$$p(z) + \frac{zp'(z)}{p(z)} = 1 + \frac{zf''(z)}{f'(z)} \prec \phi(z). \tag{2.4}$$

Since $\Re\phi(z) > 0$ and ϕ is convex in $\mathbb{D}$, therefore the solution ψ of the differential equation (2.1) is analytic in $\mathbb{D}$ with $\Re\psi(z) > 0$ and has the following integral form given by

$$\psi(z) := h(z)\left(\int_0^z \frac{h(t)}{t}dt\right)^{-1},$$

where

$$h(z) = z\exp\int_0^z \frac{\phi(t)-1}{t}dt.$$

Since $\Re\psi(z) > 0$ and p satisfies the subordination (2.4), therefore we conclude that ψ is univalent and $p \prec \psi$, where ψ is the best dominant. Thus we have obtained that $f \in \mathcal{C}(\phi)$ implies $zf'(z)/f(z) \prec \psi(z)$ and ψ is the best dominant, which is a univalent Caratheódory function. Now as $g \in \mathcal{A}$ and $f \in \mathcal{C}(\phi)$, therefore we obtain the following well defined equality

$$\frac{f(z)}{f'(z)} = \frac{z}{\psi(\omega(z))}, \quad (z \in \mathbb{D})$$

where ω is a Schwarz function. Hence, using $\min_{|z|=r}|\psi(\omega(z))| \geq \min_{|z|=r}|\psi(z)|$ and the hypothesis $\min\limits_{|z|=r}|\psi(z)| = m_r$, we obtain that

$$\left|\frac{f(z)}{f'(z)}\right| \leq \frac{r}{m_r}, \quad (0 < r < 1). \tag{2.5}$$

Now if g is majorized by f, then by definition, we have $g(z) = \Psi(z)f(z)$, where Ψ is analytic and satisfies $|\Psi(z)| \leq 1$ in $\mathbb{D}$ such that $g'(z) = \Psi(z)f'(z) + \Psi'(z)f(z)$.

Thus using (2.5) together with the following Schwarz-Pick inequality

$$|\Psi'(z)| \leq \frac{1-|\Psi(z)|^2}{1-|z|^2},$$

we obtain

$$|g'(z)| \leq |f'(z)| \left(|\Psi(z)| + \frac{1-|\Psi(z)|^2}{1-r^2} \frac{r}{m_r} \right) = |f'(z)| h(\beta, r), \tag{2.6}$$

where $|\Psi(z)| := \beta$ and

$$h(\beta, r) = \beta + \frac{1-\beta^2}{1-r^2} \frac{r}{m_r}.$$

Thus to arrive at (2.2), it suffices to show that $h(\beta, r) \leq 1$, which is equivalent to show that

$$k(\beta, r) := (1-r^2) m_r - (\beta+1) r \geq 0. \tag{2.7}$$

Since $\frac{\partial}{\partial \beta} k(\beta, r) = -r < 0$, Therefore, (2.7) holds whenever

$$k(r) := \min_{\beta} k(\beta, r) = k(1, r) \geq 0.$$

Note that $k(r)$ is a continuous function of r and further $k(0) = m_0 = \psi(0) = 1 > 0$ and $k(1) < 0$. Thus there exists a point $r_\psi \in (0,1)$ such that $k(r) \geq 0$ for all $r \in [0, r_\psi]$, where r_ψ is the least positive root of (2.3). □

The following result was proved by MacGregor [14]:

Remark 2.1. Let $g \in \mathcal{A}$ and $f \in \mathcal{C}$. If g is majorized by f in $\mathbb{D}$, then $|g'(z)| \leq |f'(z)|$ in $|z| \leq 1/3$. The result is sharp.

In our next result, we show the application to the Janowski class, which covers many well-known classes:

Corollary 2.2. *Let f belongs to $\mathcal{C}[D, E]$, where $-1 \leq E < D \leq 1$ along with $1 + D/E \geq 0$ and $-1 \leq E < 0$. If g is majorized by f, then*

$$|g'(z)| \leq |f'(z)| \quad \textit{in} \quad |z| \leq r_0,$$

where r_0 is the smallest positive root of the equation

$$(1-r^2) \left({}_2F_1 \left(1 - \frac{D}{E}, 1, 2; \frac{-Er}{1-Er} \right) \right)^{-1} - 2r = 0.$$

Proof. In Theorem 2.1, put $\phi(z) = (1+Dz)/(1+Ez)$. Then we have $\psi(z) := 1/q(z)$, where

$$q(z) = \begin{cases} \int_0^1 \left(\frac{1+Etz}{1+Ez} \right)^{\frac{D-E}{E}} dt, & \text{if } E \neq 0; \\ \\ \int_0^1 e^{D(t-1)z} dt, & \text{if } E = 0, \end{cases}$$

which further can be represented in terms of confluent and Gaussian hypergeometric functions, respectively as follows:

$$q(z) = \begin{cases} {}_2F_1 \left(1 - \frac{D}{E}, 1, 2; \frac{Ez}{1+Ez} \right), & \text{if } E \neq 0; \\ {}_1F_1 \left(1, 2; -Dz \right), & \text{if } E = 0. \end{cases}$$

Since $1 + D/E \geq 0$ and $-1 \leq E < 0$, therefore we have

$$\min_{|z|=r} \Re \psi(z) = \psi(-r) = \frac{1}{q(-r)} = \left({}_2F_1 \left(1 - \frac{D}{E}, 1, 2; \frac{-Er}{1-Er} \right) \right)^{-1}.$$

Since $\Re\psi(z) > 0$ and $\min_{|z|=r} \Re\psi(z) = \psi(-r)$, therefore we conclude that $\min_{|z|=r} |\psi(z)| = \psi(-r)$ and hence, the result follows from Theorem 2.1. □

Now we have the result for the class of convex functions of order α using Corollary 2.2:

Corollary 2.3. *Let f belongs to $\mathcal{C}[1-2\alpha, -1]$, where $0 \leq \alpha < 1$. If g is majorized by f, then $|g'(z)| \leq |f'(z)|$ in $|z| \leq r_0$, where r_0 is the smallest positive root of the equation*

$$(1-r^2)\left({}_2F_1\left(2(1-\alpha), 1, 2; \frac{r}{1+r}\right)\right)^{-1} - 2r = 0.$$

Corollary 2.4. *Let f belongs to $\mathcal{C}[D, 0]$. If g is majorized by f, then $|g'(z)| \leq |f'(z)|$ in $|z| \leq r_0$, where r_0 is the smallest positive root of the equation*

$$(1-r^2)(Dre^{-Dr}/(e^{-Dr}-1)) + 2r = 0.$$

Proof. From the proof of Corollary 2.2, we obtain that $\psi(z) = Dze^{Dz}/(e^{Dz}-1)$, when $\phi(z) = 1 + Dz$. Now with a little computation, we find that the function $l(z) = ze^z/(e^z-1)$ is convex univalent in $\mathbb{D}$. Therefore, the function $\psi(z) = l(Dz)$ is also convex in $\mathbb{D}$ for each fixed $0 < D \leq 1$. Since ψ is also symmetric about the real axis, we conclude that $\min_{|z|=r} |\psi(z)| = \psi(-r)$. Hence the result. □

Theorem 2.5. *Let ϕ be convex in $\mathbb{D}$, with $\Re\phi(z) > 0$, $\phi(0) = 1$ and suppose $f \in \mathcal{A}$ satisfies the differential subordination*

$$\frac{zf'(z)}{f(z)} + z\left(\frac{zf'(z)}{f(z)}\right)' \prec \phi(z). \tag{2.8}$$

If g is majorized by f, then $|g'(z)| \leq |f'(z)|$ in $|z| \leq r_0$, where r_0 is the least positive root of the equation

$$(1-r^2)\min_{|z|=r} \Re\psi(z) - 2r = 0,$$

where

$$\psi(z) := \frac{1}{z}\int_0^z \phi(t)dt.$$

Proof. Let $p(z) = zf'(z)/f(z)$. Then the subordination (2.8) can equivalently be written as:

$$p(z) + zp'(z) \prec \phi(z).$$

A simple calculation show that the analytic function $\psi(z) := (1/z)\int_0^z \phi(t)dt$ satisfies

$$\psi(z) + z\psi'(z) = \phi(z).$$

Now from the Hallenbeck and Ruscheweyh result [16, Theorem 3.1b, p. 71], we have $p \prec \psi$, where ψ is the best dominant and also convex. Further, since $\Re\phi(z) > 0$, using the integral operator [16, Theorem 4.2a, p. 202] preserving functions with positive real part, we see that ψ is a Caratheódory function. Thus we have

$$\frac{f(z)}{zf'(z)} \prec \frac{1}{\psi(z)} \quad \text{which implies} \quad \left|\frac{f(z)}{f'(z)}\right| \leq \frac{r}{\min_{|z|=r}|\psi(z)|} = \frac{r}{\min_{|z|=r}\Re\psi(z)}.$$

Now proceeding same as in the Theorem 2.1 result follows. □

Corollary 2.6. *Suppose $f \in \mathcal{A}$ satisfies the differential subordination*

$$\frac{zf'(z)}{f(z)} + z\left(\frac{zf'(z)}{f(z)}\right)' \prec \frac{1+z}{1-z}.$$

If g is majorized by f, then $|g'(z)| \leq |f'(z)|$ in $|z| \leq r_0$, where r_0 is the least positive root of the equation

$$(1-r^2)(2\log(1+r) - r) - 2r^2 = 0.$$

Corollary 2.7. *Suppose $f \in \mathcal{A}$ satisfies the differential subordination*

$$\frac{zf'(z)}{f(z)} + z\left(\frac{zf'(z)}{f(z)}\right)' \prec e^z.$$

If g is majorized by f, then $|g'(z)| \leq |f'(z)|$ in $|z| \leq r_0$, where r_0 is the least positive root of the equation

$$(1-r^2)(1-e^{-r}) - 2r^2 = 0.$$

Theorem 2.8. *Let ϕ be convex in $\mathbb{D}$, with $\Re\phi(z) > 0$, $\phi(0) = 1$ and suppose $f \in \mathcal{A}$ satisfies the differential subordination*

$$\frac{zf'(z)}{f(z)}\left(\frac{zf'(z)}{f(z)} + 2z\left(\frac{zf'(z)}{f(z)}\right)'\right) \prec \phi(z), \quad \alpha \in [0,1). \tag{2.9}$$

If g is majorized by f, then $|g'(z)| \leq |f'(z)|$ in $|z| \leq r_0$, where r_0 is the least positive root of the equation

$$(1-r^2)\min_{|z|=r}|\sqrt{\psi(z)}| - 2r = 0,$$

where

$$\psi(z) := \frac{1}{z}\int_0^z \phi(t)dt.$$

Proof. Let $p(z) = zf'(z)/f(z)$. Then the subordination (2.9) can be equivalently written as:

$$p^2(z) + 2zp(z)p'(z) \prec \phi(z),$$

which using the change of variable $P(z) = p^2(z)$ becomes

$$P(z) + zP'(z) \prec \phi(z).$$

Now proceeding as in Theorem 2.5, we see that $p(z) \prec \sqrt{\psi(z)}$ and $\sqrt{\psi(z)}$ is the best dominant. Further, since $\Re\phi(z) > 0$, using [16, Theorem 4.2a, p. 202], we see that ψ is a Caratheódory function. Therefore,

$$|\arg\sqrt{\psi(z)}| = \frac{1}{2}|\arg\psi(z)| \leq \frac{\pi}{4},$$

which implies $\Re\sqrt{\psi(z)} > 0$. Thus we have

$$\frac{f(z)}{zf'(z)} \prec \frac{1}{\sqrt{\psi(z)}} \quad \text{which implies} \quad \left|\frac{f(z)}{f'(z)}\right| \leq \frac{r}{\min_{|z|=r}|\sqrt{\psi(z)}|}.$$

Now proceeding same as in the Theorem 2.1 result follows. □

Corollary 2.9. *Suppose $f \in \mathcal{A}$ satisfies the differential subordination*

$$\frac{zf'(z)}{f(z)}\left(\frac{zf'(z)}{f(z)}+2z\left(\frac{zf'(z)}{f(z)}\right)'\right) \prec \frac{1+(2\alpha-1)z}{1+z}.$$

If g is majorized by f, then $|g'(z)| \leq |f'(z)|$ in $|z| \leq r_0$, where r_0 is the least positive root of the equation

$$(1-r^2)\min_{|z|=r} \Re\sqrt{\psi(z)} - 2r = 0,$$

where

$$\psi(z) := \frac{1}{z}\int_0^z \frac{1+(2\alpha-1)t}{1+t}dt.$$

Corollary 2.10. *Suppose $f \in \mathcal{A}$ satisfies the differential subordination*

$$\frac{zf'(z)}{f(z)}\left(\frac{zf'(z)}{f(z)}+2z\left(\frac{zf'(z)}{f(z)}\right)'\right) \prec 1+\alpha z, \quad (\alpha \in (0,1]).$$

If g is majorized by f, then $|g'(z)| \leq |f'(z)|$ in $|z| \leq r_0$, where r_0 is the least positive root of the equation

$$(1-r^2)\sqrt{1-\beta r} - 2r = 0, \quad \text{where} \quad \beta = \alpha/2.$$

Now we state the following result without proof as it follows from Theorem 2.1:

Theorem 2.11. *Let $\psi \in \mathcal{P}$ be a univalent function such that*

$$m_r := \min_{|z|=r}|\psi(z)| = \begin{cases} \psi(-r), & \text{if } \psi'(0) > 0; \\ \psi(r), & \text{if } \psi'(0) < 0. \end{cases}$$

Let $g \in \mathcal{A}$ and $f \in \mathcal{S}^*(\psi)$. If g is majorized by f in $\mathbb{D}$, then

$$|g'(z)| \leq |f'(z)| \quad \text{in } |z| \leq r_\psi,$$

where r_ψ is the least positive root of the equation

$$(1-r^2)m_r - 2r = 0.$$

Example 2.12. *Let us consider the analytic functions $\psi_1(z) = \sqrt{1-z}$ and $\psi_2(z) = \sqrt{1+z}$. Note that $\psi_1'(0) < 0$, $\psi_2'(0) > 0$ and for $|z| = r$,*

$$m_{r_1} = \min_{|z|=r}|\psi_1(z)| = \psi_1(r) = \sqrt{1-r} = \psi_2(-r) = \min_{|z|=r}|\psi_2(z)| = m_{r_2}.$$

Now from Theorem 2.11, we obtain the following result:
If $g \in \mathcal{A}$, $f \in \mathcal{S}^*(\psi_i)$, where $i = 1, 2$ and g is majorized by f, then $|g'(z)| \leq |f'(z)|$ in $|z| \leq r_0$, where r_0 is the least positive root of the equation

$$(1-r^2)\sqrt{1-r} - 2r = 0.$$

Interestingly, the desired radius in both the cases remain same as $\psi_1(\mathbb{D}) = \psi_2(\mathbb{D})$, though ψ_1 and ψ_2 are oppositely oriented.

Remark 2.2. Taking $\alpha = 0$ or $\eta = 1$ in corollary 2.13, case (ii) and (iii), respectively, we obtain the result proved by T. H. MacGregor [14], namely: *Let $g \in \mathcal{A}$ and $f \in \mathcal{S}^*$. If $g << f$ in $\mathbb{D}$, then $|g'(z)| \leq |f'(z)|$ in $|z| \leq 2-\sqrt{3}$.* The result is sharp.

Now we obtain the following majorization results for some known classes as well those introduced and studied in [7, 8, 15, 23].

Corollary 2.13. *Let $g \in \mathcal{A}$ and $f \in \mathcal{S}^*(\psi)$. If $g << f$ in $\mathbb{D}$, then $|g'(z)| \leq |f'(z)|$ in $|z| \leq r_\psi$, where r_ψ is the least positive root of the equation $P(r) = 0$ and the result follows for each one of the following cases:*

(*i*) $P(r) = (1-r^2)((1-Dr)/(1-Er)) - 2r$ *when* $\psi(z) = \frac{1+Dz}{1+Ez}$, *where* $-1 \leq E < D \leq 1$.
(*ii*) $P(r) = (1-r)(1-(1-2\alpha)r) - 2r$ *when* $\psi(z) = \frac{1+(1-2\alpha)z}{1+z}$, *where* $0 \leq \alpha < 1$.
(*iii*) $P(r) = (1-r^2)((1-r)/(1+r))^\eta - 2r$ *when* $\psi(z) = \left(\frac{1+z}{1-z}\right)^\eta$, *where* $0 < \eta \leq 1$.
(*iv*) $P(r) = (1-r^2)\left(\sqrt{2} - (\sqrt{2}-1)\sqrt{\frac{1+r}{1-2(\sqrt{2}-1)r}}\right) - 2r$ *when* $\psi(z) = \sqrt{2} - (\sqrt{2}-1)\sqrt{\frac{1-z}{1+2(\sqrt{2}-1)z}}$.
(*v*) $P(r) = (1-r^2)(b(1-r))^{1/a} - 2r$ *when* $\psi(z) = (b(1+z))^{1/a}$, *where* $a \geq 1$ *and* $b \geq 1/2$.
(*vi*) $P(r) = (1-r^2) - 2re^r$ *when* $\psi(z) = e^z$.
(*vii*) $P(r) = (1-r^2)(\sqrt{1+r^2} - r) - 2r$ *when* $\psi(z) = z + \sqrt{1+z^2}$.
(*viii*) $P(r) = (1-r^2) - r(1+e^r)$ *when* $\psi(z) = \frac{2}{1+e^{-z}}$.
(*ix*) $P(r) = (1-r^2)(1-\sin r) - 2r$ *when* $\psi(z) = 1 + \sin z$.

Remark 2.3. In Corollary 2.13, case (*ix*), we obtained the radius $r_\psi \approx 0.312478$ which improves the majorization-radius $r_s \approx 0.309757$ obtained in [26].

Let $\psi(z) = 1+z/(1-\alpha z^2)$, $0 \leq \alpha < 1$, introduced and studied by Kargar et al. [9]. Clearly $\psi \in \mathcal{P}$ only when $\alpha = 0$ and hence Theorem 2.11 holds when $\psi(z) = 1+z$. Moreover, for some $r > 0$, the quantity $z/\psi(z)$ does not exist for all $|z| = r$. In view of the same, the result proved by Tang and Deng [25], needs correction and the corrected version is stated in the following corollaries:

Corollary 2.14. *Let $g \in \mathcal{A}$ and $f \in \mathcal{S}^*(1+\beta z)$, $0 < \beta \leq 1$. If $g << f$ in $\mathbb{D}$, then*

$$|g'(z)| \leq |f'(z)| \quad in \quad |z| \leq r_\beta,$$

where r_β is the least positive root of the equation

$$(1-r^2)(1-\beta r) - 2r = 0.$$

Now we obtain the result related to $\mathcal{BS}(\alpha)$, the class of Booth lemniscate starlike functions, when $\alpha \neq 0$.

Corollary 2.15. *Let $0 < \alpha < 1$ and r_α be the unique root of the equation*

$$\alpha r^2 + r - 1 = 0. \tag{2.10}$$

Let $g \in \mathcal{A}$ and $g << f$ in $\mathbb{D}$, where $f \in \mathcal{BS}(\alpha)$. Then

$$|g'(z)| \leq |f'(z)| \quad in \quad |z| \leq r_{B(\alpha)} := \min\{r_\alpha, r_0\},$$

where r_0 is the least positive root of the equation

$$(1-r^2)\left(1 - \frac{r}{1-\alpha r^2}\right) - 2r = 0.$$

Proof. Observe that $\Re\left(1 + \frac{z}{1-\alpha z^2}\right) > 0$ for $|z| < r_\alpha$, where r_α is the unique root of (2.10). Thus the inequality in (2.5) holds for $|z| = r < r_\alpha$ and the result follows at once. □

3. **Product of starlike functions and a sufficient condition**

Assume that ψ_1 and ψ_2 belong to $\mathcal{P}$ and satisfy the following conditions for $|z| = r$ and $i = 1, 2$

$$\max_{|z|=r} \Re\psi_i(z) = \psi_i(r) \quad \text{and} \quad \min_{|z|=r} \Re\psi_i(z) = \psi_i(-r). \tag{3.1}$$

Motivated by Obradović and Ponnusamy [21], in this section, we consider the radius problem to generalize their result and also establish a similar result for the Urlagaddi class $\mathcal{M}(\beta) := \{f \in \mathcal{A} : \Re(zf'(z)/f(z)) < \beta,\ \beta > 1\}$. Also we extend a result of the Bulboacă and Tuneski [6] for the class $\mathcal{S}^*(\psi)$.

Theorem 3.1. *Let $g \in \mathcal{S}^*(\psi_1)$ and $h \in \mathcal{S}^*(\psi_2)$, where ψ_i satisfy the first condition in (3.4). Then the function F defined by*

$$F(z) = \frac{g(z)h(z)}{z} \tag{3.2}$$

belongs to $\mathcal{M}(\beta)$ in the disk $|z| < r_\beta = \min\{1, r_0(\beta)\}$, where $r_0(\beta)$ is the least positive root of the equation

$$\psi_1(r) + \psi_2(r) - 1 - \beta = 0. \tag{3.3}$$

The radius r_β is sharp.

Proof. Let $g \in \mathcal{S}^*(\psi_1)$ and $h \in \mathcal{S}^*(\psi_2)$. Then in view of (3.4) and subordination principle, it follows that

$$\Re\frac{zg'(z)}{g(z)} \le \psi_1(r) \quad \text{and} \quad \Re\frac{zh'(z)}{h(z)} \le \psi_2(r)$$

in $|z| \le r < 1$. Since

$$\frac{zF'(z)}{F(z)} = \frac{zg'(z)}{g(z)} + \frac{zh'(z)}{h(z)} - 1,$$

we have for $|z| = r$,

$$\Re\frac{zF'(z)}{F(z)} \le \psi_1(r) + \psi_2(r) - 1 \le \beta,$$

whenever $r \le \min\{1, r_0(\beta)\}$, where $r_0(\beta)$ is the least positive root of the equation (3.3). The sharpness follows by considering the functions

$$g(z) = z\exp\int_0^z \frac{\psi_1(t) - 1}{t}dt \quad \text{and} \quad h(z) = z\exp\int_0^z \frac{\psi_2(t) - 1}{t}dt.$$

Corollary 3.2. *Let $g \in \mathcal{S}^*(\gamma)$ and $h \in \mathcal{S}^*(\tau)$. Then the function F defined in (3.2) belongs to $\mathcal{M}(\beta)$ in the disk $|z| < \min\{1, r_0(\beta)\}$, where*

$$r_0(\beta) = \frac{\beta - 1}{3 + \beta - 2(\gamma + \tau)}.$$

The proof of the following result is much akin to Theorem 3.1, so omitted here.

Theorem 3.3. *Let $g \in \mathcal{S}^*(\psi_1)$ and $h \in \mathcal{S}^*(\psi_2)$, where ψ_i satisfy the second condition in (3.4). Then the function F defined in (3.2) is starlike of order γ in the disk $|z| < r_\gamma$, where r_γ is the least positive root of the equation*

$$\psi_1(-r) + \psi_2(-r) - 1 - \gamma = 0.$$

The radius r_γ is sharp.

We obtain the following result proved by Obradović and Ponnusamy [21]:

Remark 3.1. Let $g \in \mathcal{S}^*(\gamma)$ and $h \in \mathcal{S}^*(\tau)$. Then the function F defined in (3.2) is starlike of order γ_0 in the disk

$$|z| < \frac{1-\gamma_0}{\gamma_0 + 3 - 2(\gamma+\tau)}.$$

Remark 3.2. Note that the identity function $z \in \mathcal{S}^*(\psi)$. Thus if we choose $g(z) = z$ (or $h(z) = z$) in (3.2), then the problem reduces to obtaining the $\mathcal{M}(\beta)$-radius (or $\mathcal{S}^*(\gamma)$-radius) of the class $\mathcal{S}^*(\psi_2)$ (or $\mathcal{S}^*(\psi_1)$). It is also evident that the conditions given in (3.4) establish the inclusion relations $\mathcal{S}^*(\psi) \subseteq \mathcal{M}(\psi(1))$ and $\mathcal{S}^*(\psi) \subseteq \mathcal{S}^*(\psi(-1))$, respectively.

In the following, we extend the results of the Bulboacă and Tuneski [6]:

Theorem 3.4. *Let h be analytic with $h(0) = 0$, $h'(0) \neq 0$. Suppose that h satisfies*

$$\Re\left(1 + \frac{zh''(z)}{h(z)}\right) \geq -\frac{1}{2}$$

and

$$\frac{1}{z}\int_0^z h(t)dt \prec \frac{\psi(z)-1}{\psi(z)}. \tag{3.4}$$

If $f \in \mathcal{A}$, then

$$\frac{f(z)f''(z)}{(f'(z))^2} \prec h(z) \quad \textit{implies} \quad f \in \mathcal{S}^*(\psi).$$

Proof. Using the result [6, Theorem 3.1, p. 3], we see that $f(z)f''(z)/(f'(z))^2 \prec h(z)$ implies

$$\frac{1}{z}\int_0^z \left(1 - \left(\frac{f(t)}{f'(t)}\right)'\right) dt = 1 - \frac{f(z)}{zf'(z)} \prec \frac{1}{z}\int_0^z h(t)dt.$$

From the above subordination, we have

$$\frac{f(z)}{zf'(z)} \prec 1 - \frac{1}{z}\int_0^z h(t)dt.$$

Now to prove that $f \in \mathcal{S}^*(\psi)$, it suffices to consider

$$1 - \frac{1}{z}\int_0^z h(t)dt \prec \frac{1}{\psi(z)},$$

which is equivalent to (3.4). This completes the proof. □

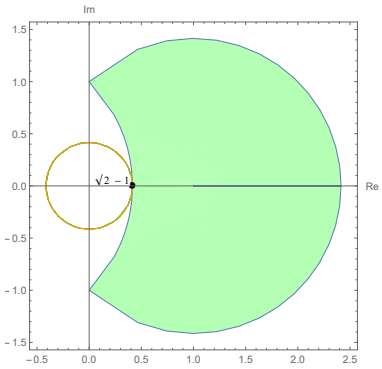
4. **Distortion theorem**

Ma-Minda [13] proved the distortion theorem for the class $\mathcal{S}^*(\psi)$ with some restriction on ψ, namely $|\psi(z)|$ attains its maximum and minimum value respectively at $z = r$ and $z = -r$, see eq. (1.2). Now what if ψ does not satisfy the condition (1.2) and why the condition (1.2) is so important? To answer this, we first need to recall the following result:

Lemma 4.1. ([13]) *Let $f \in \mathcal{S}^*(\psi)$ and $|z_0| = r < 1$. Then $-f_0(-r) \leq |f(z_0)| \leq f_0(r)$. Equality holds for some $z_0 \neq 0$ if and only if f is a rotation of f_0, where $zf_0(z)/f_0(z) = \psi(z)$ such that*

$$f_0(z) = z\exp\int_0^z \frac{\psi(t)-1}{t}dt. \tag{4.1}$$

We see that a Ma-Minda starlike function, in general, need not satisfy the condition (1.2). To examine the same, let us consider two different Ma-Minda starlike functions, namely $\psi_1(z) := z + \sqrt{1+z^2}$ and $\psi_2(z) := 1 + ze^z$. The unit disk images of ψ_1 and ψ_2 are displayed in figure 1 and figure 2.



Legend for figure 3 -

Let $\min_{|z|=r}|\psi_2(z)| =: \gamma_i(r)$, where
$z = r_i e^{i\theta}$, $0 \le \theta \le \pi$,
then from table 1, we have
$\gamma_1(1) = 0.372412$,
$\gamma_2(4/5) = 0.527912$,
$\gamma_3(2/3) = 0.611553$,
$\gamma_4(1/2) = 0.693287$,
$\gamma_5(r) = 1 - re^{-r}$, where $r \le (3-\sqrt{5})/2$.

FIGURE 1. $\psi_1(z) := z + \sqrt{1+z^2}$

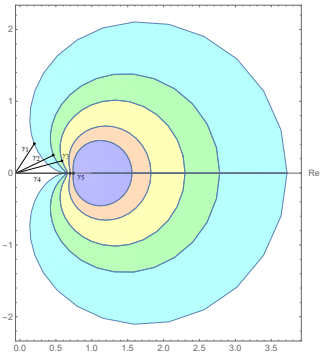


FIGURE 2. $\psi_2(z) := 1 + ze^z$

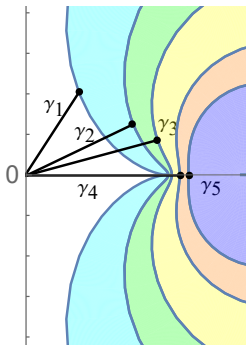


FIGURE 3. A zoom of figure 2

We know that the radius of a circle, centered at origin and touching only the boundary points of a image domain of a complex function, yields the optimal values of the modulus of the function. For example, see figure 1 to locate the lower bound of the modulus for a crescent function. Therefore it is evident from figure 1 that both the bounds $\psi_1(-r)$ and $\psi_1(r)$ of $|\psi_1|$ are attained on the real line and we have $\psi_1(-r) \le |\psi_1(z)| \le \psi_1(r)$ for each $|z| = r$. Whereas, from figure 2, we see

that only the upper bound $\psi_2(r)$ of $|\psi_2|$ is attained on the real line and $|\psi_2(z)| \leq \psi_2(r)$ for each $|z| = r$. Although both ψ_1 and ψ_2 are Ma-Minda functions but the distortion theorem of Ma-Minda [13, theorem 2, p. 162] does not accommodate the function ψ_2, as the lower bound for $|\psi_2(z)|$ is not attained on the real line for all $|z| = r > (3 - \sqrt{5})/2$, see figure 3. To overcome this limitation, we modify the distortion theorem, wherein we theoretically assume the modulus bounds of the function and obtain a more general result. Thus the Ma-Minda functions, for which modulus bounds are not attained on the real line but could be estimated, can now be entertained for distortion theorem using the following result:

Theorem 4.1 (Modified Distortion Theorem)**.** *Let ψ be a Ma-Minda function. Assume that $\min_{|z|=r}|\psi(z)| = |\psi(z_1)|$ and $\max_{|z|=r}|\psi(z)| = |\psi(z_2)|$, where $z_1 = re^{i\theta_1}$ and $z_2 = re^{i\theta_2}$ for some $\theta_1, \theta_2 \in [0, \pi]$. Let $f \in \mathcal{S}^*(\psi)$ and $|z_0| = r < 1$. Then*

$$|\psi(z_1)| \left(\frac{-f_0(-r)}{r}\right) \leq |f'(z_0)| \leq \left(\frac{f_0(r)}{r}\right) |\psi(z_2)|. \tag{4.2}$$

Proof. Let $p(z) = zf'(z)/f(z)$. Then $f \in \mathcal{S}^*(\psi)$ if and only if $p(z) \prec \psi(z)$. Using a result [13, Theorem 1, p.161], we have

$$\frac{f(z)}{z} \prec \frac{f_0(z)}{z}, \tag{4.3}$$

where f_0 is given by (4.1). Now using Maximum-Minimum principle of modulus, (4.3) and by Lemma 4.1, $-f_0(-r)/r \leq |f(z_0)/z| \leq f_0(r)/r$, we easily obtain for $|z_0| = r$

$$\begin{aligned}
|\psi(z_1)| \left(\frac{-f_0(-r)}{r}\right) &= \min_{|z|=r} |\psi(z)| \min_{|z|=r} \left|\frac{f_0(z)}{z}\right| \\
&\leq \left|p(z_0)\frac{f(z_0)}{z_0}\right| \\
&= |f'(z_0)| \leq \max_{|z|=r} |\psi(z)| \max_{|z|=r} \left|\frac{f_0(z)}{z}\right| \\
&= \left(\frac{f_0(r)}{r}\right) |\psi(z_2)|,
\end{aligned}$$

that is,

$$|\psi(z_1)| \left(\frac{-f_0(-r)}{r}\right) \leq |f'(z_0)| \leq \left(\frac{f_0(r)}{r}\right) |\psi(z_2)|,$$

where z_1 and z_2 are as defined in the hypothesis. Hence the result. □

To illustrate Theorem 4.1, we consider the function $\psi(z) = 1 + ze^z$. Then we have the following expression for its modulus:

$$|\psi(z)| = \sqrt{1 + re^{r\cos\theta}(re^{r\cos\theta} + 2\cos(\theta + r\sin\theta))}. \tag{4.4}$$

Using equation (4.4) and Theorem 4.1, we obtain the following table 1, providing the minimum for various choices of r.

Now using Theorem 4.1, we obtain the following distortion theorem for the class $\mathcal{S}^*(1 + ze^z)$:

TABLE 1. The lower bounds for $|1+ze^z|$ for different choices of $r=|z|$.

r	$0\leq\theta_1\leq\pi$	$\lvert\psi(re^{i\theta_1})\rvert$	$m(r,\theta_1)=\lvert\psi(re^{i\theta_1})\rvert(-f_0(-r)/r)$
1	1.88438	0.372412	0.197923
4/5	2.01859	0.527912	0.304374
2/3	2.17677	0.611553	0.375966
1/2	2.58169	0.693287	0.467769
$r\leq(3-\sqrt{5})/2$	π	$\psi_2(-r)$	$f_0'(-r)$

Example 4.2. *Let $\psi(z)=1+ze^z$ and $\min\limits_{|z|=r}|\psi(z)|=|\psi(z_1)|$, where $z_1=re^{i\theta_1}$ for some $\theta_1\in[0,\pi]$. Let $f\in\mathcal{S}^*(\psi)$ and $|z_0|=r<1$. Then*

$$m(r,\theta_1)\leq|f'(z_0)|\leq f_0'(r),\quad\left(r>\tfrac{3-\sqrt{5}}{2}\right)$$

and

$$f_0'(-r)\leq|f'(z_0)|\leq f_0'(r),\quad\left(r\leq\tfrac{3-\sqrt{5}}{2}\right),$$

where $f_0(z)=z\exp(e^z-1)$ and $m(r,\theta_1)$ is provided in table 1 for some specific values of r.

Remark 4.1. In Theorem 4.1, if we assume that $\theta_1=\pi$ and $\theta_2=0$, then extremes in equation (4.2) simplifies to $f_0'(-r)$ and $f_0'(r)$, respectively since $zf_0'(z)/f_0(z)=\psi(z)$. Thus the extremes in the equation (4.2) are in terms of r alone and also lead to the sharp bounds. Consequently, we obtain the following distortion theorem of Ma-Minda [13] as a special case of Theorem 4.1:

Let $\min_{|z|=r}|\psi(z)|=\psi(-r)$ and $\max_{|z|=r}|\psi(z)|=\psi(r)$. If $f\in\mathcal{S}^*(\psi)$ and $|z_0|=r<1$. Then

$$f_0'(-r)=\psi(-r)\frac{f_0(-r)}{-r}\leq|f'(z_0)|\leq\frac{f_0(r)}{r}\psi(r)=f_0'(r).$$

Equality holds for some $z_0\neq0$ if and only if f is a rotation of f_0.

5. Bohr-Phenomenon for functions in $S_f(\psi)$

Note that "the Bohr radius of the class $\mathcal{X}$ is at least r_x", this holds for every result in this section. In general, Bohr radius is estimated for a specific class provided the sharp coefficients bounds of the functions in that class are known. For instance, consider the class of starlike univalent functions, where we have the sharp coefficient bounds: $|a_n|\leq n$. However, for most of the Ma-Minda subclasses, the better coefficients bounds are yet not known. Hence, we encounter the following problem, especially in context of Ma-minda classes, which we deal here to a certain extent:
Problem: *If coefficients bounds are not known, how one can find a good lower estimate for the Bohr radius of a given class?*

To readily understand the above problem, consider the class $\mathcal{S}^*(1+ze^z)$, where the sharp coefficients bounds for functions in this class are unknown. In this situation, how one can find the Bohr radius for this class or is there any way out with the lower bounds all alone? Here below we state Theorem 5.1, where we find a solution for this problem. Note that the Bohr radius $3-2\sqrt{2}\approx0.1713$ for the class $\mathcal{S}^*$ serves as a lower bound for the class $S_f(\psi)$ and is also a special case of Theorem 5.1.

Let $\mathbb{B}(0,r) := \{z \in \mathbb{C} : |z| < r\}$, $g(z) = \sum_{k=1}^{\infty} b_k z^k$, $\mathcal{S}^*(\psi)$ and $S_f(\psi)$ as defined in (1.1) and (1.3) respectively. For any $g \in S_f(\psi)$, we find the radius r_b so that $S_f(\psi)$ obey the following Bohr-phenomenon:

$$\sum_{k=1}^{\infty} |b_k| r^k \leq d(f(0), \partial\Omega) \quad \text{for} \quad |z| = r \leq r_b, \tag{5.1}$$

where $d(f(0), \partial\Omega)$ denotes the Euclidean distance between $f(0)$ and the boundary of $\Omega = f(\mathbb{D})$. Now we prove our main result:

Theorem 5.1. *Let r_* be the Koebe-radius for the class $\mathcal{S}^*(\psi)$, $f_0(z)$ be given by the equation (4.1) and $g(z) = \sum_{k=1}^{\infty} b_k z^k \in S_f(\psi)$. Assume $f_0(z) = z + \sum_{n=2}^{\infty} t_n z^n$ and $\hat{f}_0(r) = r + \sum_{n=2}^{\infty} |t_n| r^n$. Then $S_f(\psi)$ satisfies the Bohr-phenomenon*

$$\sum_{k=1}^{\infty} |b_k| r^k \leq d(f(0), \partial\Omega), \quad \textit{for } |z| = r \leq r_b, \tag{5.2}$$

where $r_b = \min\{r_0, 1/3\}$, $\Omega = f(\mathbb{D})$ and r_0 is the least positive root of the equation

$$\hat{f}_0(r) = r_*.$$

Proof. Since $g \in S_f(\psi)$, we have $g \prec f$ for a fixed $f \in \mathcal{S}^*(\psi)$. By letting r tends to 1 in Lemma 4.1, we obtain the Koebe-radius $r_* = -f_0(-1)$. Therefore $\mathbb{B}(0, r_*) \subset f(\mathbb{D})$, which implies $r_* \leq d(0, \partial\Omega) = |f(z)|$ for $|z| = 1$. Also using [13, Theorem 1, p.161], we have

$$\frac{f(z)}{z} \prec \frac{f_0(z)}{z}. \tag{5.3}$$

Recall the result [3, Lemma 1, p.1090], which reads as: let f and g be analytic in $\mathbb{D}$ with $g \prec f$, where $f(z) = \sum_{n=0}^{\infty} a_n z^n$ and $g(z) = \sum_{k=0}^{\infty} b_k z^k$. Then $\sum_{k=0}^{\infty} |b_k| r^k \leq \sum_{n=0}^{\infty} |a_n| r^n$ for $|z| = r \leq 1/3$. Now using the result for $g \prec f$ and (5.3), we have

$$\sum_{k=1}^{\infty} |b_k| r^k \leq r + \sum_{n=2}^{\infty} |a_n| r^n \leq \hat{f}_0(r) \quad \text{for } |z| = r \leq 1/3.$$

Finally, to establish the inequality (5.2), it is enough to show $\hat{f}_0(r) \leq r_*$. But this holds whenever $r \leq r_0$, where r_0 is the least positive root of the equation $\hat{f}_0(r) = r_*$. The existence of the root r_0 is ensured by the relations $\hat{f}_0(1) \geq |f_0(1)| \geq r_*$ and $\hat{f}_0(0) < r_*$. Thus, if $r_b = \min\{r_0, 1/3\}$ then $\sum_{k=1}^{\infty} |b_k| r^k \leq d(0, \partial\Omega)$ holds. Hence the result. □

Remark 5.1. Let us further assume that the coefficients B_n of ψ are positive. Then the function $f_0(z) = z + \sum_{n=2}^{\infty} t_n z^n$ defined by integral representation (4.1) can be written as

$$f_0(z) = z \exp\left(\sum_{n=1}^{\infty} \frac{B_n}{n} z^n \right),$$

which implies $f_0(r) = \hat{f}_0(r)$ for $|z| = r$.

From the proof of Theorem 5.1, we have the following:

Theorem 5.2. *Let r_* be the Koebe-radius for the class $\mathcal{S}^*(\psi)$, $f_0(z)$ be given by the equation (4.1) and $f(z) = z + \sum_{n=2}^{\infty} a_n z^n \in \mathcal{S}^*(\psi)$. Assume $f_0(z) = z + \sum_{n=2}^{\infty} t_n z^n$ and $\hat{f}_0(r) = r + \sum_{n=2}^{\infty} |t_n| r^n$. Then $\mathcal{S}^*(\psi)$ satisfies the Bohr-phenomenon*

$$r + \sum_{n=2}^{\infty} |a_n| r^n \leq d(f(0), \partial\Omega), \quad \text{for } |z| = r \leq r_b,$$

where $r_b = \min\{r_0, 1/3\}$, $\Omega = f(\mathbb{D})$ and r_0 is the least positive root of the equation

$$\hat{f}_0(r) = r_*.$$

Some Applications:

(a). If we choose $\psi(z) = (1 + Dz)/(1 + Ez)$, $-1 \leq E < D \leq 1$, then $\mathcal{S}^*(\psi)$ denotes the class of Janowski starlike functions and we have

$$r_* = \begin{cases} (1-E)^{\frac{D-E}{E}}, & E \neq 0; \\ e^{-D}, & E = 0. \end{cases} \tag{5.4}$$

and

$$f_0(z) = \begin{cases} z(1+Ez)^{\frac{D-E}{E}}, & E \neq 0; \\ z\exp(Dz), & E = 0. \end{cases} \tag{5.5}$$

Observe that if $E \neq 0$, the n-th ($n \geq 2$) coefficients of $f_0(z)$ is given by

$$t_n = \prod_{k=2}^{\infty} \frac{D - (k-1)E}{(k-1)!}. \tag{5.6}$$

Thus from Theorem 5.1, we have the following result:

Corollary 5.3. *Let $\psi(z) = (1 + Dz)/(1 + Ez)$, $-1 \leq E < D \leq 1$. Then $S_f(\psi)$ (and $\mathcal{S}^*(\psi)$) satisfies the Bohr-phenomenon (5.1) for $|z| = r \leq r_b$, where $r_b = \min\{r_0, 1/3\}$ and r_0 is the least positive root of the equation*

$$r + \sum_{n=2}^{\infty} |t_n| r^n - (1-E)^{\frac{D-E}{E}} = 0,$$

where t_n is as defined in (5.6).

Note that for the Janowski class, sharp coefficients bounds in general are not known. Now as an application of Corollary 5.3, we obtain the following result when $t_n > 0$:

Corollary 5.4. (Bohr-radius with Janowski class) *Let $\psi(z) = (1 + Dz)/(1 + Ez)$, $-1 \leq E < D \leq 1$.*

(*i*) *If $E = 0$ and $D \geq \frac{3}{4}\log 3$. Then $S_f(\psi)$ (and $\mathcal{S}^*(\psi)$) satisfies the Bohr-phenomenon (5.1) for $|z| = r \leq r_0$, where r_0 is the only real root of the equation*

$$1 - re^{D(1+r)} = 0. \tag{5.7}$$

(*ii*) *If $E \neq 0$ and further satisfies*

$$3(1-E)^{\frac{D-E}{E}} \leq (1 + E/3)^{\frac{D-E}{E}}. \tag{5.8}$$

Then $S_f(\psi)$ (and $\mathcal{S}^*(\psi)$) satisfies the Bohr-phenomenon (5.1) for $|z| = r \leq r_0$, where r_0 is the only real root of the equation

$$(1-E)^{\frac{D-E}{E}} - r(1+Er)^{\frac{D-E}{E}} = 0. \tag{5.9}$$

The result is sharp for the function f_0 defined in (5.5).

Proof. (i): Since $E = 0$, we have $r_* = e^{-D}$. Moreover $\hat{f}_0(r) = f_0(r) = r\exp(Dr)$. Now we need to show

$$r\exp(Dr) \leq e^{-D} \tag{5.10}$$

or equivalently $T(r) := 1 - re^{D(1+r)} \geq 0$ holds for $r \leq r_0$. Which obviously holds for $\frac{3}{4}\log 3 \leq D \leq 1$. Since $d(f_0(0), \partial f_0(\mathbb{D})) = r_*$, therefore we see from inequality (5.10) that Bohr-radius is sharp for the function f_0 given by (5.5).

(ii): Proceeding as in case (i), it is sufficient to show the inequality

$$r(1+Er)^{\frac{D-E}{E}} \leq (1-E)^{\frac{D-E}{E}} \tag{5.11}$$

or equivalently $g(r) := (1-E)^{\frac{D-E}{E}} - r(1+Er)^{\frac{D-E}{E}} \geq 0$ holds for $r \leq r_0$. Which obviously follows whenever D and E satisfies (5.8). In view of the inequality (5.11), the sharp Bohr-radius is achieved for the function f_0 given by (5.5). □

Remark 5.2. **(Bohr-radius with starlike functions of order α)** Let $\psi(z) := (1+(1-2\alpha)z)/(1-z)$, where $0 \leq \alpha < 1$. We see $\mathcal{S}^*(\psi) := \mathcal{S}^*(\alpha)$ and for this class, we have

$$r_* = \frac{1}{2^{2(1-\alpha)}} \quad \text{and} \quad f_0(z) = \frac{z}{(1-z)^{2(1-\alpha)}}.$$

Observe that here $\hat{f}_0(r) = f_0(r)$. Now as an application of Corrollary 5.4, we obtain the result due to Bhowmik et al. [3], namely: *If* $0 \leq \alpha \leq 1/2$. *Then* $S_f(\psi)$ *satisfies the Bohr-phenomenon* $\sum_{k=1}^{\infty}|b_k|r^k \leq d(f(0), \partial\Omega)$, *for* $|z| = r \leq r_b$, *where* $r_b = \min\{r_0, 1/3\} = r_0$ *and* r_0 *is the only real root of the equation* $(1-r)^{2(1-\alpha)}/r = 2^{2(1-\alpha)}$. *The result is sharp.*

Now form the above remark, in particular, we have:

Corollary 5.5. *If* $0 \leq \alpha \leq 1/2$. *Then the class* $\mathcal{S}^*((1+(1-2\alpha)z)/(1-z))$ *satisfies the Bohr-phenomenon* (5.1) *for* $|z| = r \leq r_0$, *where* r_0 *is the only real root of the equation*

$$(1-r)^{2(1-\alpha)}/r = 2^{2(1-\alpha)}.$$

The result is sharp. In particular, the Bohr radius for the class $\mathcal{S}^*$ *is* $3 - 2\sqrt{2} \approx 0.1713$.

(b). If we choose $\psi(z) = \sqrt{1+z}$, then $\mathcal{S}^*(\psi) := \mathcal{SL}^*$, the class of lemniscate starlike functions and for this class we have:

$$f_0(z) = \frac{4z\exp(2\sqrt{1+z}-2)}{(1+\sqrt{1+z})^2} \quad \text{and} \quad r_* = -f_0(-1) \approx 0.541341. \tag{5.12}$$

Also in this case $\hat{f}_0(r) = f_0(r)$ and therefore, we obtain the following corollary:

Corollary 5.6. *The class* $S_f(\psi)$ *(and* $\mathcal{SL}^*$*), where* $\psi(z) = \sqrt{1+z}$ *satisfies the Bohr-phenomenon* (5.1) *for* $|z| = r \leq 1/3$.

(c). If we consider $\psi(z) = 1 + ze^z$, then $\mathcal{S}^*(\psi) := \mathcal{S}^*_{\wp}$, the class of cardioid starlike functions introduced in [12] and for this class, we have:

$$f_0(z) = z\exp(e^z - 1) \quad \text{and} \quad r_* = -f_0(-1) \approx 0.531464. \tag{5.13}$$

Here we can also see that $\hat{f}_0(r) = f_0(r)$ and we obtain the following corollary:

Corollary 5.7. *The class $S_f(\psi)$ (and $\mathcal{S}^*_\wp$), where $\psi(z) = 1 + ze^z$ satisfies the Bohr-phenomenon* (5.1) *for $|z| = r \leq 1/3$.*

Ali et al. [2] also showed that the coefficient bound of a function in a class have a role in the estimation of the Bohr-radius. Observed that for each $f \in \mathcal{S}^*(\psi)$, the class $S_f(\psi)$ satisfies the Bohr-phenomenon for $r \leq \min(1/3, r_0)$, where r_0 is the least positive root of $\hat{f}_0(r) - r_* = 0$. Since $\mathcal{S}^*(\psi) \subset \bigcup_{f\in\mathcal{S}^*(\psi)} S_f(\psi)$, therefore the Bohr-radius for the class $\mathcal{S}^*(\psi)$ is $r \geq \min(1/3, r_0)$. In Corollary 5.7, we find $r_0 \approx 0.349681$ (an upper bound for Bohr radius), which is almost close to $1/3 \approx 0.33333$ and is the unique root of $f_0(r) - r_* = 0$. Moreover, the bound for the coefficients of the functions belonging to $\mathcal{S}^*_\wp$ and $\mathcal{SL}^*$ have been conjectured [12, 24] with the extremals given in (5.13) and (5.12) respectively. Thus by using Theorem 5.1 and the approach dealt in [2] (assuming that conjectures are true), we propose the following conjectures:

Conjecture. The Bohr-radius for the class $\mathcal{S}^*_\wp$ is $r_0 \approx 0.349681$ which is the unique root in $(0, 1)$ of the equation

$$re^{e^r} = e^{1/e}.$$

Conjecture. The Bohr-radius for the class $\mathcal{SL}^*$ is $r_0 \approx 0.439229$, which the unique root in $(0, 1)$ of the equation

$$e^2 r \exp(2\sqrt{1+r} - 2) = (1 + \sqrt{1+r})^2.$$

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] Ali, R. M., Barnard, R. W., Solynin, A. Yu.: A note on the Bohrs phenomenon for power series. J. Math. Anal. Appl. **449**, 154–167 (2017)

[2] Ali, R. M., Jain, N. K., Ravichandran, V.: Bohr radius for classes of analytic functions. Results Math. **74**, Art. 179, 13 pp. (2019)

[3] Bhowmik, B., Das, N.: Bohr phenomenon for subordinating families of certain univalent functions. J. Math. Anal. Appl. **462**, 1087–1098 (2018)

[4] Boas, H. P., Khavinson, D.: Bohr's power series theorem in several variables. Proc. Amer. Math. Soc. **125**, 2975–2979 (1997)

[5] Bohr, H.: A theorem concerning power series. Proc. London Math. Soc. **13**, 1–5 (1914)

[6] Bulboacă, T., Tuneski, N.: New criteria for starlikeness and strongly starlikeness. Mathematica. **43**, 11–22 (2001)

[7] Cho, N. E., Kumar, V., Kumar, S. S., Ravichandran, V.: Radius problems for starlike functions associated with the sine function. Bull. Iranian Math. Soc. **45**, 213–232 (2019)

[8] Goel, P., Kumar, S. S.: Certain Class of Starlike Functions Associated with Modified Sigmoid Function. Bull. Malays. Math. Sci. Soc. (2019) https://doi.org/10.1007/s40840-019-00784-y.

[9] Kargar, R., Ebadian, A., Sokół, J.: On Booth lemniscate and starlike functions. Anal. Math. Phys. **9**(1), 143–154 (2019)

[10] Kayumov, I. R., Ponnusamy, S.: Bohr inequality for odd analytic functions. Comput. Methods Funct. Theory. **17**, 679–688 (2017)

[11] Kayumov, I. R., Ponnusamy, S.: Improved version of Bohr's inequality. C. R. Math. Acad. Sci. Paris. **356**, 272–277 (2018)

[12] Kumar, S. S., Kamaljeet, G.: A Cardioid Domain and Starlike Functions. Analysis and Mathematical Physics. (2020), (Communicated)

[13] Ma, W.C., Minda, D.: A unified treatment of some special classes of univalent functions. *Proceedings of the Conference on Complex Analysis (Tianjin, 1992)*. 157–169, Conf. Proc. Lecture Notes Anal., I Int. Press, Cambridge, MA.

[14] MacGregor, T. H.: Majorization by univalent functions. Duke Math. J. **34**, 95–102 (1967)

[15] Mendiratta, R., Nagpal, S., Ravichandran, V.: On a subclass of strongly starlike functions associated with exponential function. Bull. Malays. Math. Sci. Soc. **38**, 365–386 (2015)

[16] Miller S. S., Mocanu, P. T.: *Differential subordinations*, Monographs and Textbooks in Pure and Applied Mathematics. Marcel Dekker, Inc, New York (2000)

[17] Muhanna, Y. A.: Bohr's phenomenon in subordination and bounded harmonic classes. Complex Var. Elliptic Equ. **55**, 1071–1078 (2010)

[18] Muhanna, Y. A., Ali, R. M.: Bohr's phenomenon for analytic functions into the exterior of a compact convex body. J. Math. Anal. Appl. **379**, 512–517 (2011)

[19] Muhanna, Y. A., Ali, R. M., Ng, Z. C., Hansi, S. F. M.: Bohr radius for subordinating families of analytic functions and bounded harmonic mappings. J. Math. Anal. Appl. **420**, 124–136 (2014)

[20] Muhanna, Y. A., Ali R. M., Ponnusamy, S.: On the Bohr inequality. In Progress in Approximation Theory and Applicable Complex Analysis (Edited by N.K. Govil et al.), Springer Optimization and Its Applications. **117**, 265295 (2016)

[21] Obradović, M., Ponnusamy, S.: Product of univalent functions. Math. Comput. Modelling. **57**, 793–799 (2013)

[22] Paulsen, V. I., Popescu, G., Singh, D.: On Bohr's inequality. Proc. London Math. Soc. **85**, 493–512 (2002)

[23] Raina, R. K., Sokół, J.: Some properties related to a certain class of starlike functions. C. R. Math. Acad. Sci. Paris. **353**, 973–978 (2015)

[24] Sokół, J.: Coefficient estimates in a class of strongly starlike functions. Kyungpook Math. J. **49**, 349–353 (2009)

[25] Tang H., Deng, G.: Majorization problems for some subclasses of starlike functions. J. Math. Res. Appl. **39**, 153–159 (2019)

[26] Tang, H., Srivastava, H. M., Li, Sh., Deng, G. T.: Majorization results for subclasses of starlike functions based on the sine and cosine functions. Bull. Iran. Math. Soc. (2019). https://doi.org/10.1007/s41980-019-00262-y.

Department of Applied Mathematics, Delhi Technological University, Delhi–110042, India

E-mail address: spkumar@dce.ac.in

Department of Applied Mathematics, Delhi Technological University, Delhi–110042, India

E-mail address: gangania.m1991@gmail.com

Journal Pre-proofs



Regular paper

On the Outage Probability and BER of a DF-Based Multi-Hop System Over - - and - - -Extreme Fading Channels

Tau Raphael Rasethuntsa, Manpreet Kaur, Sandeep Kumar

PII: S1434-8411(20)30238-7
DOI: https://doi.org/10.1016/j.aeue.2020.153324
Reference: AEUE 153324

To appear in: *International Journal of Electronics and Communications*

Received Date: 30 January 2020
Accepted Date: 20 June 2020

Please cite this article as: T.R. Rasethuntsa, M. Kaur, S. Kumar, On the Outage Probability and BER of a DF-Based Multi-Hop System Over - - and - - -Extreme Fading Channels, *International Journal of Electronics and Communications* (2020), doi: https://doi.org/10.1016/j.aeue.2020.153324

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier GmbH.

Journal Pre-proofs

On the Outage Probability and BER of a DF-Based Multi-Hop System Over $\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme Fading Channels

Tau Raphael Rasethuntsa, Manpreet Kaur*, *Sandeep Kumar†

Abstract

This paper analyzes the performance of decode-and-forward (DF)-based multi-hop relaying communication systems over $\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme fading channels. Analytical formulas for the probability density function (PDF) and the cumulative distribution function (CDF) of the received signal-to-noise ratio (SNR) as well as its moment generating function (MGF) are derived. Based on the derived PDFs, novel closed-form expressions for the outage probability (OP), bit error rate (BER) under coherent and non-coherent modulation schemes are derived. Additionally, asymptotic analyses of BER based on Poincare series expansions of SNR PDFs are carried out and results show good approximations for low SNR regimes. Simulations are performed to verify the accuracy and to show the tightness of the theoretical analysis.

Index Terms

Multi-hop system, decode-and-forward relay, non-linear generalized fading, outage Probability, Bit error rate

Tau Raphael Rasethuntsa is based in, Lenasia Extension 9, Gauteng Province, Johannesburg, South Africa and Ha Abia/Mahlabatheng, Moreneng, Maseru, Lesotho, e-mail: tauzand@gmail.com

Manpreet Kaur is with Department of Computer Science & Engineering, Delhi Technological University, Delhi, India, e-mail: manpreettiet@gmail.com*

Corresponding Author: *Sandeep Kumar is with the Central Research Laboratory, Bharat Electronics Limited, Ghaziadbad, Uttar Pradesh, India, e-mail: sann.kaushik@gmail.com†

Journal Pre-proofs

Journal Pre-proofs

I. Introduction

The use of multi-hop communication systems as a way of increasing network coverage area and energy-efficiency with high data rates is gaining importance in next-generation wireless networks. In a multi-hop wireless communication system, information is transmitted from a source device to a destination device through several intermediate nodes, referred to as relays [1]. The relaying systems are classified in three categories 1) decode and forward (DF); 2) amplify and forward (AF); and 3) compress and forward (CF). In the DF systems, the relay device decodes the received data and then re-sends the decoded data to the next relay, therefore preventing the propagation of noise [2]. In AF, the received signal is amplified without decoding and forwarded to the next node. In CF, the received data is compressed, re-encoded and forwarded to the next-hop device [3].

The performance analysis of multi-hop communication systems over numerous fading distributions has been well documented in the literature [4], [5], [6], [7], [8], [9], [10]. The ergodic capacity of a multi-hop relaying system over Rayleigh fading channels has been studied in [4]. The end-to-end performance of a DF relaying system based on Ricean and Nakagami-m fading channels was explored in [5]. Yang et al. [6] studied the unified performance analysis of a DF-based multi-hop system based on second order statistics in average outage duration (AOD) and level crossing rate (LCR) over Nakagami, Rayleigh and Rician fading channels. The performance study of a multi-hop communication system with regenerative relays has been investigated over generalized $\eta-\mu$, $\alpha-\mu$ and $\kappa-\mu$ fading channels in [7], [8] and [9], respectively. Using traditional performance metrics as well as LCR and amount of fade duration (AFD), Cao et al. [10] presented the performance analysis of end-to-end wireless links over the Generalized-K fading channel.

In wireless communication, multi-path and shadowing highly deteriorate the strength of transmitting signals. A wide range of arithmetical distributions are available to characterize the fading effect in various signal propagation scenarios [11]. For instance, motivated by the pioneering works in [12] and [13] the generalized $\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme fading models were proposed in [14] and [15]. The $\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme are the generalized distributions which accounts for small-scale variations of mobile radio signals in line-of-sight (LOS) communications. The

$\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme models are very flexible models that can account well for non-linearity of multi-path propagation. The $\alpha-\kappa-\mu$ distribution is a popular model as it includes some realistic fading channels such as $\kappa-\mu$, Rayleigh, Nakagami-m, Weibull, and Rice as its special cases. The $\alpha-\kappa-\mu$-Extreme fading distribution is derived from the $\alpha-\kappa-\mu$ distribution by following the same methodology as in the derivation of the $\kappa-\mu$-Extreme model from the $\kappa-\mu$ distribution [15].

The $\alpha-\kappa-\mu$-Extreme fading distribution is used to characterize the mobile radio propagation under nonlinear severe multi-path fading effects usually encountered in enclosed environments such as airplanes, trains, and shopping malls etc. [16]. Furthermore, there are only a few studies that deal with the performance of wireless communications over $\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme fading channels. In [17] and [18] analytical expressions for OP, average error rate, and channel capacity of a single-input–single-output system were derived in the presence of $\alpha-\kappa-\mu$ fading. The second-order statistics of $\alpha-\kappa-\mu$ distribution were considered in [19] whereas [20] studied the spectrum sensing over $\alpha-\kappa-\mu$ fading channels. A comprehensive performance evaluation of cooperative spectrum sensing over $\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme fading channels has been presented in [16]. Recently in [21] and [22], energy detector performance over $\alpha-\kappa-\mu$ based composite fading channels were studied and the effect of all the fading parameters on the system performance was discussed.

In summary, the performance analysis of DF cooperative systems over many types of fading channels has been extensively studied, however, the performance of the DF over that $\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme fading channels in multi-hop systems has not been considered. Moreover, given the several practical application scenarios for the multi-hop wireless communication systems and the fact that$\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme fading model fits empirical data of wireless channels much better than previously studied fading models, it is important to investigate the different performance metrics of the multi-hop wireless systems operating over these channels. Henceforth, this paper investigates the outage probability (OP) and (bit error rate) BER performance of the DF multi-hop communication systems over the generalized fading model. The end-to-end performance of DF multi-hop systems is studied through numerical evaluation, which validate the correctness of our derived expressions compared to well-known fading distributions, and

Journal Pre-proofs

study the impacts of fading parameters and path length on the system performance.

The remainder of this paper is organized as follows. The system and channel model are described in Section II. In Section III, we provide the closed-form expressions for the statistical parameters over the channels. Different performance metrics and their approximations for the DF-based multi-hop wireless systems over $\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme fading channels has been presented in Section IV and V respectively. Performance evaluation, results, and discussion are provided in Section VI. Finally, we conclude this paper in Section VI.

II. Model Formulation and Description

A. System and Channel Models

Consider an n-hop system shown in Figure 1 transmitting information from the source S to the destination D via $n-1$ intermediate relays $R_1, R_2, \ldots, R_{n-1}$ in time-sharing principles. The relay link of the present model has a property of regenerative transmission and the channels of the n hops are assumed independent. As a result, if the signal-to-noise ratio (SNR) of any of the $n-1$ hops falls below a predetermined threshold γ_{th}, the system in regarded as failed [10], [8].

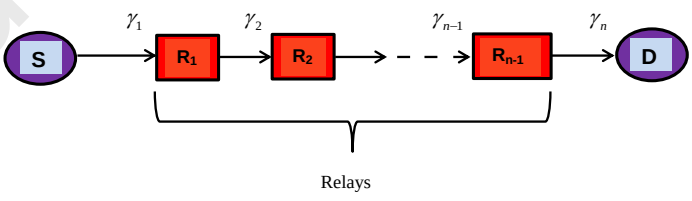


Fig. 1. A n-link DF-based multi-hop system.

Journal Pre-proofs

TABLE I
LIST OF SYMBOLS

Notation	Name of the Function	References
$\delta(\cdot)$	Dirac delta function	[23]
$Q(\cdot,\cdot)$	Generalized Marcum Q-function	[24, Eq. 2.23]
$\mathbb{E}(\cdot)$	Expectation operator	[16]
$\tilde{\Phi}(.;.;.)$	Regularized confluent hypergeometric function	[25, Eq. 9.210(1)]
$G(\cdot)$	Meijer's functions	[25, Eq. 9.301]
$H(\cdot)$	Fox-H function	[26, Eq. 2.57]
$I(\cdot)$	Modified Bessel functions of first kind	[27, Eq. 2.3.31)]
$\Gamma(.)$	Gamma function	[27, Eq. 2.3.22)]
$erfc(.)$	Complementary error function	[27, Eq. 2.3.17)]

If the SNR of the i-th hop is denoted by γ_i, with probability density function (PDF) f_{γ_i}, $i = 1, \dots n$, the statistical properties of the end-to-end system SNR are then determined by the SNR of the last hop γ_n. According to [23], the PDF of the end-to-end system SNR γ can be written as

$$p_\gamma(\gamma) = \mathcal{A}\delta(\gamma_n) + \bar{\mathcal{A}} p_{\gamma_n}(\gamma_n), \tag{1}$$

where $\mathcal{A}$ denotes the probability of the first outage occurring for the first $n-1$ hops and $\bar{\mathcal{A}} = 1 - \mathcal{A}$. The detail of all the symbols used in this work has been shown in Table I. The probability $\mathcal{A}$ is given by [23]

$$\begin{aligned} \mathcal{A} &= 1 - Pr\left(\gamma_1 > \gamma_{th}, \gamma_2 > \gamma_{th}, \cdots, \gamma_{m-1} > \gamma_{th}\right) \\ &= 1 - \prod_{k=1}^{m-1} Pr\left(\gamma_n > \gamma_{th}\right) = 1 - \prod_{k=1}^{m-1} \left[1 - P_{\gamma_n}(\gamma_{th})\right], \end{aligned} \tag{2}$$

where P_{γ_n} is the cumulative distribution function (CDF) of γ_n.

1) The $\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme fading models: The PDF of the instantaneous SNR of the n-th hop $\gamma_n > 0$ following the $\alpha_n - \kappa_n - \mu_n$ distribution is given by

$$p_{\gamma_n}(\gamma_n) = \mathcal{K}_n \gamma_n^{(\frac{\alpha_n}{2})\omega_n - 1} e^{-\sigma_n \gamma_n^{\frac{\alpha_n}{2}}} I_{\mu_n - 1}\left(\sqrt{\theta_n \gamma_n^{\frac{\alpha_n}{2}}}\right), \tag{3}$$

Journal Pre-proofs

where, in order to improve clarity in presentation, we have set

$$\mathcal{K}_n = \frac{\alpha_n \sigma_n^{\omega_n} e^{-\mu_n \kappa_n}}{2(\mu_n \kappa_n)^{\omega_n - 1}},\ \omega_n = \frac{\mu_n + 1}{2},\ \sigma_n = \frac{\mu_n(\kappa_n + 1)}{\bar{\gamma}_n^{\frac{\alpha_n}{2}}},$$
$$\theta_n = 4\mu_n \kappa_n \sigma_n \tag{4}$$

with $\alpha_n, \mu_n, \kappa_n > 0$. The average SNR per symbol $\bar{\gamma}_n$ is given by

$$\bar{\gamma}_n = \frac{e^{\mu_n \kappa_n} \left[\mu_n(\kappa_n + 1)\right]^{\frac{2}{\alpha_n}}}{\Gamma\left(\frac{2}{\alpha_n} + \mu_n\right) \tilde{\Phi}\left(\frac{2}{\alpha_n} + \mu_n; \mu_n; \mu_n \kappa_n\right)} \frac{E_b}{N_0}, \tag{5}$$

where E_b denotes the energy per symbol and N_0 is the single-sided power spectral density. Using the method of [28], one can easily show that the CDF of γ_n is written as

$$P_{\gamma_n}(\gamma_n) = 1 - Q_{\mu_n}\left(\sqrt{2\kappa_n \mu_n}, \sqrt{2\sigma_n \gamma_n^{\frac{\alpha_n}{2}}}\right),\ \gamma_n > 0, \tag{6}$$

The PDF of the instantaneous SNR of the n-th hop $\gamma_n > 0$ following the $\alpha_n - \kappa_n - \mu_n$-Extreme distribution is given by

$$p_{\gamma_n}^e(\gamma) = \frac{a_n \gamma_n^{\alpha_n/4 - 1}}{e^{b_n \gamma_n^{\frac{\alpha_n}{2}}}} I_1\left(\sqrt{c_n \gamma_n^{\frac{\alpha_n}{2}}}\right) + \frac{e^{-2m_n}}{2\sqrt{\gamma_n \bar{\gamma}_n}} \delta\left(\sqrt{\frac{\gamma_n}{\bar{\gamma}}}\right), \tag{7}$$

where

$$a_n = \frac{\alpha_n m_n e^{-2m_n}}{\bar{\gamma}_n^{\alpha_n/4}}, b = \frac{2m_n}{\bar{\gamma}_n^{\frac{\alpha_n}{2}}}, c_n = \frac{(16 m_n^2)}{\bar{\gamma}_n^{\frac{\alpha_n}{2}}}, \tag{8}$$

with $\alpha_n, \mu_n, \kappa_n > 0$ and $m_n = \mu_n(\kappa_n + 1)^2/(2\kappa_n + 1)$ as the Nakagami parameter which is inversely proportional to the fading severity. The average SNR per symbol is given by

$$\bar{\gamma}_n^e = \frac{e^{2m_n} (2m_n)^{\frac{2}{\alpha_n} - 1}}{\Gamma\left(\frac{2}{\alpha_n} + \mu_n\right) \tilde{\Phi}\left(\frac{2}{\alpha_n} + \mu_n; \mu_n; \mu_n \kappa_n\right)} \frac{E_b}{N_0}. \tag{9}$$

The CDF of the $\alpha - \kappa - \mu$-Extreme fading model can also be derived using the method of [28] resulting in

$$P_{\gamma_n}^e(\gamma) = 1 - Q_0\left(2\sqrt{m_n}, 2\sqrt{m_n(\gamma/\bar{\gamma})^{\frac{\alpha}{2}}}\right). \tag{10}$$

The symbols in (4) and (8) will be employed throughout the paper unless the original notation produces shorter expressions.

Journal Pre-proofs

Remark 1. It is also worth noting that the proposed PDFs representations renders, purely on a symbolic level, the $\alpha-\kappa-\mu$-Extreme fading model PDF without the extra $\delta(\cdot)$ term as a special case of the PDF of the $\alpha-\kappa-\mu$ model in a sense that $\mathcal{K}_n = a_n$, $\omega_n = 1/2$, $\sigma_n = b_n$, $\theta_n = c_n$ and $\mu_n = 2$. This observation will be particularly useful for reducing redundancy in the forthcoming derivations.

III. Statistical Properties

Two fading channels namely $\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme are considered in this study. For the convenience of readers, the analysis of $\alpha-\kappa-\mu$ channel is performed under **Case I** and that of $\alpha-\kappa-\mu$-Extreme channel under **Case II** in each section.

A. PDF and CDF of End-to-End SNR

Case I : Assume that the SNRs of the n hops are independent and not necessarily identically distributed (INID) random variables following the $\alpha-\kappa-\mu$ distribution. Then we have from (3) and (6) that the PDF and CDF of the end-to-end SNR γ must be given respectively by

$$p_\gamma(\gamma) = \mathcal{A}\delta(\gamma) + \mathcal{K}_n\bar{\mathcal{A}}\frac{\gamma^{(\frac{\alpha_n}{2})\omega_n-1}}{e^{\sigma_n\gamma^{\frac{\alpha_n}{2}}}} I_{\mu_n-1}\left(\sqrt{\theta_n\gamma^{\frac{\alpha_n}{2}}}\right) \tag{11}$$

where $\mathcal{A} = 1-\prod_{k=1}^{n-1} Q_{\mu_k}\left(\sqrt{2\kappa_k\mu_k}, \sqrt{2\sigma_k\gamma_{th}^{\frac{\alpha_k}{2}}}\right)$ and

$$P_\gamma(\gamma) = 1-\bar{\mathcal{A}}Q_{\mu_n}\left(\sqrt{2\kappa_n\mu_n}, \sqrt{2\sigma_n\gamma^{\frac{\alpha_n}{2}}}\right). \tag{12}$$

Case II : On the other hand if the SNRs of the n hops are INID random variables following the $\alpha-\kappa-\mu$-Extreme distribution, then we have from (7) and (10) that the PDF of the end-to-end SNR γ must be given by

$$p_\gamma^e(\gamma) = \mathcal{A}^e\delta(\gamma) + \bar{\mathcal{A}}^e\left[\frac{a_n\gamma^{\frac{\alpha_n}{4}-1}}{e^{b_n\gamma^{\frac{\alpha_n}{2}}}} I_1\left(\sqrt{c_n\gamma^{\frac{\alpha_n}{2}}}\right) + \delta^*(\gamma)\right] \tag{13}$$

Journal Pre-proofs

where $\delta^*(\gamma) = \dfrac{e^{-2m_n}}{2\sqrt{\gamma\bar{\gamma}_n}}\delta\left(\sqrt{\dfrac{\gamma_n}{\bar{\gamma}}}\right)$
and $\mathcal{A}^e = 1 - \prod_{k=1}^{n-1} Q_0\left(2\sqrt{m_k}, 2\sqrt{m_k(\gamma_k/\bar{\gamma}_k)^{\frac{\alpha_k}{2}}}\right)$. The CDF of γ can be written as

$$P_\gamma^e(\gamma) = 1 - \bar{\mathcal{A}}^e Q_0\left(2\sqrt{m_n}, 2\sqrt{m_n(\gamma/\bar{\gamma})^{\frac{\alpha_n}{2}}}\right). \tag{14}$$

B. Moment Generating Function

Another useful statistical characteristic of random variables is the moment generating function (MGF) used to obtain BER under non-coherent modulation, see Section IV. The MGF of γ, denoted by $\mathcal{M}_\gamma$, is defined as

$$\mathcal{M}_\gamma(s) \triangleq \int_0^\infty e^{-s\gamma} f_\gamma(\gamma) d\gamma. \tag{15}$$

For **Case I**, we substitute (11) into (15) to get the integral

$$\mathcal{M}_\gamma(s) = \mathcal{A} + \mathcal{K}_n \bar{\mathcal{A}} \overbrace{\int_0^\infty \frac{\gamma^{(\frac{\alpha_n}{2})\omega_n - 1}}{e^{s\gamma} e^{\sigma_n \gamma^{\frac{\alpha_n}{2}}}} I_{\mu_n - 1}\left(\sqrt{\theta_n \gamma^{\frac{\alpha_n}{2}}}\right) d\gamma}^{\mathcal{I}_1} \tag{16}$$

In order to evaluate $\mathcal{I}_1$, we set $y = \gamma^{\alpha_n/2}$ and employ [26, Eq. 1.25] and [29, Eq. 8.4.22(1)] as well as the relationship between the Meijer-G and the Fox-H functions as

$$\begin{aligned} \mathcal{I}_1 =& \pi \int_0^\infty y^{\omega_n - 1} e^{-\sigma_n y} H_{0,1}^{1,0}\left(s^{\frac{\alpha_n}{2}} y \,\middle|\, \begin{matrix} - \\ \left(0, \frac{\alpha_n}{2}\right) \end{matrix}\right) \\ & \times H_{1,3}^{1,0}\left(\frac{\theta_n}{4} y \,\middle|\, \begin{matrix} \left(\frac{\mu_n}{2}, 1\right) \\ (\omega_n - 1, 1), (1 - \omega_n, 1), \left(\frac{\mu_n}{2}, 1\right) \end{matrix}\right) dy \end{aligned} \tag{17}$$

Now, expanding the two Fox H-functions in terms of their definition as given in [30, Eq. 1.1.1] and reversing the order of integration, we have that

$$\begin{aligned} \mathcal{I}_1 =& \pi \frac{1}{(2\pi i)^2} \int_{\mathcal{L}_r} \int_{\mathcal{L}_t} \overbrace{\int_0^\infty y^{\omega_n + r + t - 1} e^{-\sigma_n y} dy}^{\mathcal{G}} \\ & \times \frac{\Gamma\left[-\left(\frac{\alpha_n}{2}\right) r\right] \Gamma(\omega_n - 1 - t) \left(s^{\frac{\alpha_n}{2}}\right)^r \left(\frac{\theta_n}{4}\right)^t dr dt}{\Gamma(\omega_n + t) \Gamma\left(1 - \frac{\mu_n}{2} + t\right) \Gamma\left(\frac{\mu_n}{2} - t\right)} \end{aligned} \tag{18}$$

Journal Pre-proofs

where $\mathcal{L}_r$ and $\mathcal{L}_t$ are two suitable contours in the complex r and t planes respectively. The integral $\mathcal{G}$ can be evaluated with the aid of [25, Eq. 3.326(2)] to yield

$$\begin{aligned}\mathcal{I}_1 =& \frac{\pi}{\sigma_n^{\omega_n}(2\pi i)^2}\int_{\mathcal{L}_r}\int_{\mathcal{L}_t}\frac{\Gamma\left[-\left(\frac{\alpha_n}{2}\right)r\right]\Gamma(\omega_n-1-t)}{\Gamma(\omega_n+t)} \\ &\times\frac{\Gamma\left(\omega_n+r+t\right)\left(s^{\frac{\alpha_n}{2}}/\sigma_n\right)^r\left(\frac{\theta_n}{4\sigma_n}\right)^t drdt}{\Gamma\left(1-\frac{\mu_n}{2}+t\right)\Gamma\left(\frac{\mu_n}{2}-t\right)}\end{aligned} \tag{19}$$

where $\Re(\omega_n+r+t)>0$. Now comparing (19) with the definition of the bivariate Fox H-function given in [26, Eq. 2.57] yields a closed-form expression for the MGF as shown in (22) of Table II. Remark 1 suggests that the MGF for **Case II** can be obtained from (22) by inspection resulting in the MGF for **Case II** being as shown in (23) of Table II. It is also worth pointing out that (15) can be written with the aid of integration by parts as

$$\mathcal{M}_\gamma(s) \triangleq s\int_0^\infty e^{-s\gamma}F_\gamma(\gamma)d\gamma. \tag{20}$$

Substituting (12) into (20), a more restricted alternative form for the MGF of γ for **Case I** valid for $2/\alpha_2, \mu_n \in \mathbb{N}$, can be obtained from [31, Theorem 2] for **Case I** as

$$\begin{aligned}\mathcal{M}_\gamma(s) =& 1-\frac{\alpha_n}{2}\bar{\mathcal{A}}+\frac{\bar{\mathcal{A}}\exp(-\kappa_n\mu_n)\left(\frac{2}{\alpha_n\sigma_n}\right)^{\frac{2}{\alpha_n}}s}{\left(\sqrt{\frac{2}{\alpha_n}}\right)(2\pi)^{\frac{1}{\alpha_n}-\frac{1}{2}}} \\ &\times\sum_{r=\mu_n}^{\infty}\sum_{n=0}^{\infty}\frac{(2\kappa_n\mu_n/\alpha_n)^n}{n!\Gamma(n+r+1)}\left(\frac{2}{\alpha_n}\right)^r \\ &\times G_{\frac{2}{\alpha_n},1}^{1,\frac{2}{\alpha_n}}\left[s\left(\frac{2}{\alpha_n\sigma_n}\right)^{\frac{2}{\alpha_n}}\middle|\begin{matrix}\Delta(\frac{2}{\alpha_n},1-n-r-\frac{2}{\alpha_n})\\ 0\end{matrix}\right].\end{aligned} \tag{21}$$

where $\Delta(k,\vartheta)=\frac{\vartheta}{k},\ldots,\frac{\vartheta+k-1}{k}$ comprises k elements. A similar expression can be obtained for **Case II**.

Journal Pre-proofs

TABLE II
MOMENT GENERATING FUNCTOINS

$$\mathcal{M}_{\gamma}(s) = \mathcal{A} + \bar{\mathcal{A}}\mathcal{K}_n \frac{\pi}{\sigma_n^{\omega_n}} H_{1,0:1,0;1,3}^{0,1:1,0;1,0} \left(\begin{array}{c|c} \frac{s^{\frac{\alpha_n}{2}}}{\sigma_n} & \mathbf{\Delta_1}\left(\alpha_n, \mu_n, \omega_n\right) \\ \frac{\theta_n}{4\sigma_n} & \mathbf{\Delta_2}\left(\alpha_n, \mu_n, \omega_n\right) \end{array} \right) \tag{22}$$

$$\mathcal{M}_{\gamma}^{e}(s) = \mathcal{A}^{e} + \bar{\mathcal{A}}^{e} \frac{a_n \pi}{\sqrt{b_n}} H_{1,0:1,0;1,3}^{0,1:1,0;1,0} \left(\begin{array}{c|c} \frac{s^{\frac{\alpha_n}{2}}}{b_n} & \mathbf{\Delta_1}\left(\alpha_n, 2, \frac{1}{2}\right) \\ \frac{c_n}{4b_n} & \mathbf{\Delta_2}\left(\alpha_n, 2, \frac{1}{2}\right) \end{array} \right). \tag{23}$$

$$\mathbf{\Delta_1} \equiv (1-\omega_n; 1, 1) : \; - \; ; \left(\frac{\mu_n}{2}, 1\right)$$

$$\mathbf{\Delta_2} \equiv - \; : \left(0, \frac{\alpha_n}{2}\right); \; (\omega_n - 1, 1), (1-\omega_n, 1), \left(\frac{\mu_n}{2}, 1\right)$$

IV. Performance Analysis of Multi-Hop DF-Based Relay Systems

A. Outage probability

The OP is defined as the probability that the SNR per symbol falls below a certain threshold γ_{th}. By employing (12) and (14), this probability can be written for **Case I** as

$$\begin{aligned} P_{out} &= P\left(\gamma \leq \gamma_{th}\right) \\ &= 1 - \bar{\mathcal{A}} Q_{\mu_n}\left(\sqrt{2\kappa_n \mu_n}, \sqrt{2\sigma_n \gamma_{th}^{\alpha_n/2}}\right). \end{aligned} \tag{24}$$

Similarly, for **Case II**, we have that OP can be written as

$$P_{out}^{e} = 1 - \bar{\mathcal{A}}^{e} Q_0\left(2\sqrt{m_n}, 2\sqrt{m_n(\gamma_{th}/\bar{\gamma})^{\alpha_n/2}}\right). \tag{25}$$

B. Bit Error Rate

1) Coherent Modulation Schemes: For a given SNR γ, the BER for a variety of coherent modulation schemes can be obtained by averaging the PDF f_{γ} over the conditional additive white Gaussian noise (AWGN) BER given as [32, Eq. 17]

$$P_b(e|\gamma) \triangleq (\phi/2)\text{erfc}\left(\rho\sqrt{\gamma/2}\right) \tag{26}$$

Journal Pre-proofs

where erfc$(\cdot)$ is the complementary error function [25, Eq. 8.250(4)]. The constants ϕ and ρ vary depending on the type of modulation scheme. For BFSK, $\phi = 1$ and $\rho = 1$. For BPSK, $\phi = 1$ and $\rho = \sqrt{2}$. For 4-QAM and QPSK, $\phi = 2$ and $\rho = 1$. Finally, for M-PAM, $\phi = 2(1 - 1/M)$ and $\rho = \sqrt{6/(M^2 - 1)}$. Thus, for **Case I** we average (11) over (26) to obtain the integral

$$P_b = \frac{\phi}{2} \int_0^\infty p_\gamma(\gamma) \text{erfc}\left(\rho\sqrt{\gamma/2}\right) d\gamma. \tag{27}$$

Substituting (11) into (27), one has that

$$P_b = \frac{\phi\mathcal{A}}{2} \int_0^\infty \delta(\gamma) \text{erfc}\left(\rho\sqrt{\gamma/2}\right) d\gamma + \frac{\phi\mathcal{K}_n\bar{\mathcal{A}}}{2}\mathcal{I}_2 \tag{28}$$

where

$$\mathcal{I}_2 = \int_0^\infty \frac{\gamma^{(\alpha_n/2)\omega_n - 1}}{e^{\sigma_n\gamma^{\alpha_n/2}}} I_{\mu_n - 1}\left(\sqrt{\theta_n\gamma^{\frac{\alpha_n}{2}}}\right) \text{erfc}\left(\rho\sqrt{\frac{\gamma}{2}}\right) d\gamma. \tag{29}$$

Following the same method as in Section III, using [29, Eq. 8.4.14(2)] to express the complementary error function in terms of its Mellin-Barnes representation, we obtain after some simple mathematical manipulations the BER under coherent modulation schemes for **Case I** and **Case II** as shown in (30) and (31) of Table III, respectively.

2) Non-Coherent Modulation Schemes: Under a range of different non-coherent modulation schemes, the BER of the end-to-end SNR γ can be written as [32, Eq. 27]

$$P_{bn} \triangleq \phi \int_0^\infty e^{-\rho\gamma} f_\gamma(\gamma) d\gamma. \tag{32}$$

The values of ϕ and ρ also vary depending on the modulation scheme. Specifically, for BFSK, $\phi = \frac{1}{2}$ and $\rho = \frac{1}{2}$. For DBPSK, $\phi = \frac{1}{2}$ and $\rho = 1$ and for M-FSK $\phi = (M - 1)/2$ and $\rho = \frac{1}{2}$. Upon substituting (11) into (32), it is clear that the resulting integral is of the form of the MGF integral given in (16). It is therefore straightforward to show that

$$P_{bn} = \phi\mathcal{M}_\gamma(\rho). \tag{33}$$

Journal Pre-proofs

TABLE III
BER UNDER COHERENT MODULATION

$$P_b = \frac{\phi}{2}\mathcal{A} + \bar{\mathcal{A}}\mathcal{K}_n \frac{\phi\sqrt{\pi}}{2}\left(\frac{(\sqrt{2})^{\alpha_n}}{\rho^{\alpha_n}}\right)^{\omega_n} \times H^{0,2:1,0;1,0}_{2,1:0,1;1,3}\left(\begin{array}{c|c} \sigma_n\dfrac{(\sqrt{2})^{\alpha_n}}{\rho^{\alpha_n}} & \boldsymbol{\nabla_1}(\alpha_n,\mu_n,\omega_n) \\ \theta_n\dfrac{(\sqrt{2})^{\alpha_n}}{4\rho^{\alpha_n}} & \boldsymbol{\nabla_2}(\alpha_n,\mu_n,\omega_n) \end{array}\right) \tag{30}$$

$$P_b^e = \frac{\phi}{2}\mathcal{A}^e + a_n\bar{\mathcal{A}}^e \frac{\phi\sqrt{\pi}}{2}\left(\frac{(\sqrt{2})^{\alpha_n}}{\rho^{\alpha_n}}\right)^{\frac{1}{2}} \times H^{0,2:1,0;1,0}_{2,1:0,1;1,3}\left(\begin{array}{c|c} b_n\dfrac{(\sqrt{2})^{\alpha_n}}{\rho^{\alpha_n}} & \boldsymbol{\nabla_1}\left(\alpha_n,2,\frac{1}{2}\right) \\ c_n\dfrac{(\sqrt{2})^{\alpha_n}}{4\rho^{\alpha_n}} & \boldsymbol{\nabla_2}\left(\alpha_n,2,\frac{1}{2}\right) \end{array}\right). \tag{31}$$

$$\boldsymbol{\nabla_1} \equiv \left(1-\frac{\alpha_n}{2}\omega_n;\frac{\alpha_n}{2},\frac{\alpha_n}{2}\right),\left(\frac{1}{2}-\frac{\alpha_n}{2}\omega_n;\frac{\alpha_n}{2},\frac{\alpha_n}{2}\right):-;\left(\frac{\mu_n}{2},1\right)$$

$$\boldsymbol{\nabla_2} \equiv \left(-\frac{\alpha_n}{2}\omega_n;\frac{\alpha_n}{2},\frac{\alpha_n}{2}\right):(0,1);\,(\omega_n-1,1),(1-\omega_n,1),\left(\frac{\mu_n}{2},1\right)$$

Similarly, the BER for non-coherent modulation schemes under **Case II**, can be derived using (23) resulting in

$$P_{bn}^e = P_{bn} = \phi\mathcal{M}_\gamma^e(\rho). \tag{34}$$

V. BER AND APPROXIMATIONS FOR LOW SNR REGIMES

In this section, we explore the asymptotic behaviour of the BER by employing asymptotic Poincare series expansions for both the $\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme fading models in (11) and (13). Following the idea from [33] of using [25, Eq. 8.445] and [25, Eq. 0.316], we have that the PDFs of the SNRs for the n-th hop of **Case I** and **Case II** fading channels have Poincare asymptotic series expansion near the origin given respectively by

$$p_\gamma(\gamma) \dot{\sim} \mathcal{A}\delta(\gamma) + \bar{\mathcal{A}}\mathcal{K}_n\left(\frac{\theta_n}{4}\right)^{\omega_n-1}\sum_{k=0}^{N-1} d_k\gamma^{t_k} + \mathcal{O}\left(\gamma^{t+N-1}\right) \tag{35}$$

Journal Pre-proofs

where $\mathcal{O}$ denotes the order of a term in the asymptotic series and

$$d_k = \sum_{j=0}^{k} \frac{(-1)^j \sigma_n^j \left(\frac{\theta_n}{4}\right)^{k-j}}{j!(k-j)!\Gamma(k+\mu_n-j)},\ t_k = \alpha_n(k+\mu_n)/2 - 1. \tag{36}$$

$$p_\gamma^e(\gamma) \dot{\sim} \mathcal{A}^e \delta(\gamma) + \bar{\mathcal{A}}^e \frac{a_n\sqrt{c_n}}{2} \sum_{k=0}^{N-1} d_k^e \gamma^{t_k^e} + \mathcal{O}\left(\gamma^{t+N-1}\right) \tag{37}$$

where

$$d_k^e = \sum_{j=0}^{k} \frac{(-1)^j b_n^j \left(\frac{c_n}{4}\right)^{k-j}}{j!(k-j)!\Gamma(k+2-j)},\ t_k^e = \alpha_n(k+1)/2 - 1. \tag{38}$$

We will now attempt to derive asymptotic BER expressions based on the series expansions in (35) and (37). Due to the form of the series, the process of deriving BER equates to a straightforward evaluation of Mellin transforms of appropriate functions.

1) Coherent Modulation Schemes: Substituting (35) and (37) into (26) and employing [29, Eq. 8.4.15(2)], we have that the asymptotic BER under coherent modulation for **Case I** and **Case II** are given respectively by

$$\mathcal{P}_b \dot{\sim} \frac{\phi}{2}\mathcal{A} + \bar{\mathcal{A}}\mathcal{K}_n \left(\frac{\theta_n}{4}\right)^{\omega_n - 1} \frac{\phi}{\sqrt{\pi}} \sum_{k=0}^{N-1} d_k \frac{2^{t_k}\Gamma\left(\frac{3}{2}+t_k\right)}{(t_k+1)\rho^{2(t_k+1)}}, \tag{39}$$

$$\mathcal{P}_b^e \dot{\sim} \frac{\phi}{2}\mathcal{A}^e + \phi\bar{\mathcal{A}}^e \frac{a_n\sqrt{c_n}}{2\sqrt{\pi}} \sum_{k=0}^{N-1} d_k^e \frac{2^{t_k^e}\Gamma\left(\frac{3}{2}+t_k^e\right)}{(t_k^e+1)\rho^{2(t_k^e+1)}}. \tag{40}$$

2) Non-coherent Modulation Schemes: Similarly, by substituting (35) and (37) into (32) and employing [29, Eq. 8.4.3(1)], we have that the asymptotic BER under non-coherent modulation for **Case I** and **Case II** are given respectively by

$$\mathcal{P}_{bn} \dot{\sim} \phi\mathcal{A} + \phi\bar{\mathcal{A}}\mathcal{K}_n \left(\frac{\theta_n}{4}\right)^{\omega_n - 1} \sum_{k=0}^{N-1} d_k \frac{\Gamma(t_k+1)}{\rho^{t_k+1}} \tag{41}$$

$$\mathcal{P}_{bn}^e \dot{\sim} \phi\mathcal{A}^e + \phi\bar{\mathcal{A}}^e \frac{a_n\sqrt{c_n}}{2} \sum_{k=0}^{N-1} d_k^e \frac{\Gamma(t_k^e+1)}{\rho^{t_k^e+1}}. \tag{42}$$

VI. Numerical Experiments and Simulations

In this section, the comprehensive performance evaluation of our DF-based multi-hop system has been obtained over $\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme fading channel for **Case I** and **Case II**,

Journal Pre-proofs

respectively. Specifically, we consider equally spaced n hops between a source S and destination D. Although the results produced here are applicable for arbitrary values of the fading parameters α, κ, μ and m that are positive real numbers, and are present in terms of standard bivariate Fox H-function, still the choice of the parameters are made as described in [15], [16], [34], [35] and to give better visibility of the results. The Matlab code given in [36] has been used for the computation of bivariate Fox H-function in this work. To further ensure fairness in comparisons, $\bar{\gamma}_n$ in (5) and (9) have to be multiplied by $(n/n_{max})^{\beta}$, where β is the propagation loss exponent which is taken as 3 for the current setup. Monte Carlo simulations are included to verify and investigate the correctness of the performance metrics derived in the preceding sections.

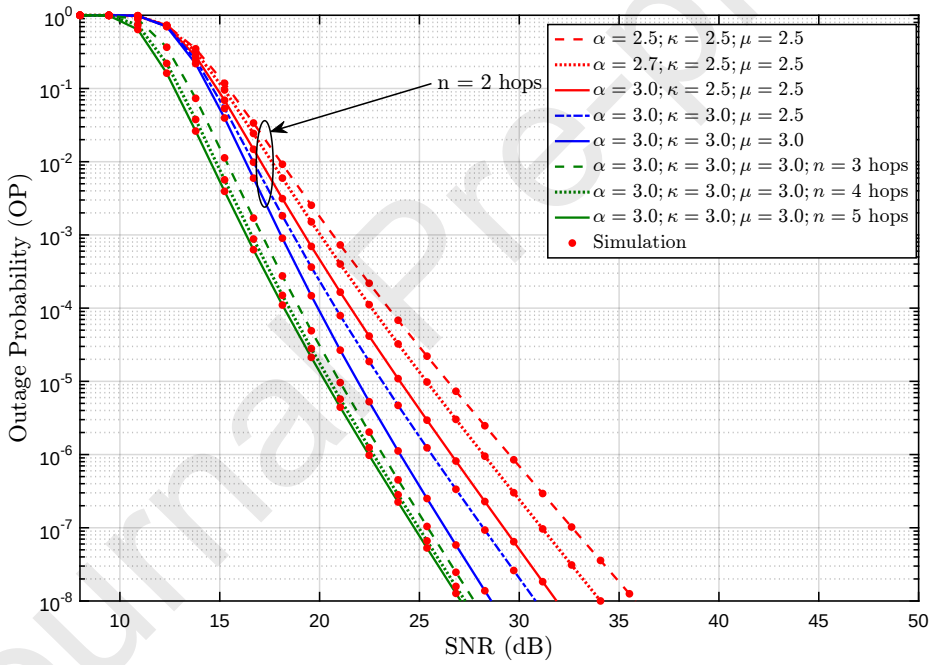


Fig. 2. OP curves for the $\alpha - \kappa - \mu$ case under moderate fading showing the effects of increments in hop number and fading model parameter values.

Figure 2 illustrates the OP vs average SNR curves for **Case I** under moderate fading conditions. From the figure, we can deduce that the increment in the value of any of the fading parameters results in notable system performance enhancement, with steeper increments noted with the increase in μ, value. Increasing the number of hops also improves system performance, especially

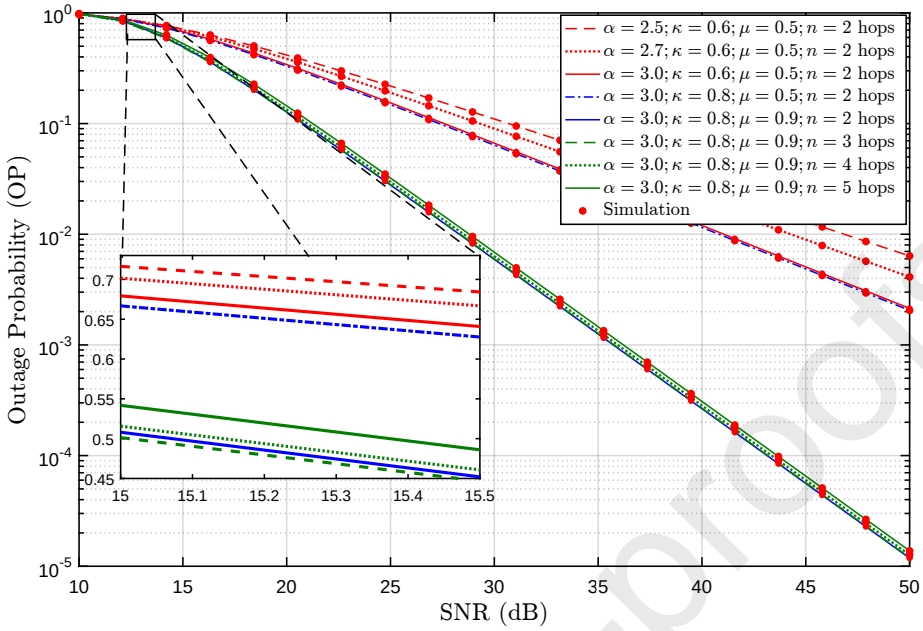


Fig. 3. OP curves for the $\alpha - \kappa - \mu$ case under severe fading showing the effects of increments in hop number and fading model parameter values.

at lower SNR levels which is a particularly desirable trait. The increase in performance is albeit of a diminishing type. This implies that there is an optimal number/range of hops that maximizes system performance. That is, relaying for the sole purpose of improving outage is only fruitful up to a certain number of relays upon which any additional relaying is futile. Another important result which was reported in [8] and has been noted to apply to the present model is that OP decreases with decrease in the spectrum efficiency $\mathcal{R}$. We avoid displaying the result here for the sake of brevity. Figure 3 shows the effect of severe fading on OP for Case I. Even though there is still a considerable system improvement for increments in parameter values, little is gained or there is actual performance degradation with the increase in number of hopes. This means that an engineer who is trying to increase coverage under severe fading conditions can afford to sacrifice a negligible loss in system performance with addition of a few hops, but will pay a steeper price in outage probability as the number of hops increases considerably.

Finally, Figure 4 shows the OP curves under Case II for severity fading scenario. It is clear

Journal Pre-proofs

that the system performs slightly better under $\alpha-\kappa-\mu$-Extreme fading model with less steeper penalties paid as a result of increase in the hop number than $\alpha-\kappa-\mu$ model. Another important result which was reported in [8] and has been noted to apply to the present model is that OP decreases with decrease in the spectrum efficiency. We avoid displaying the result here for the sake of brevity..

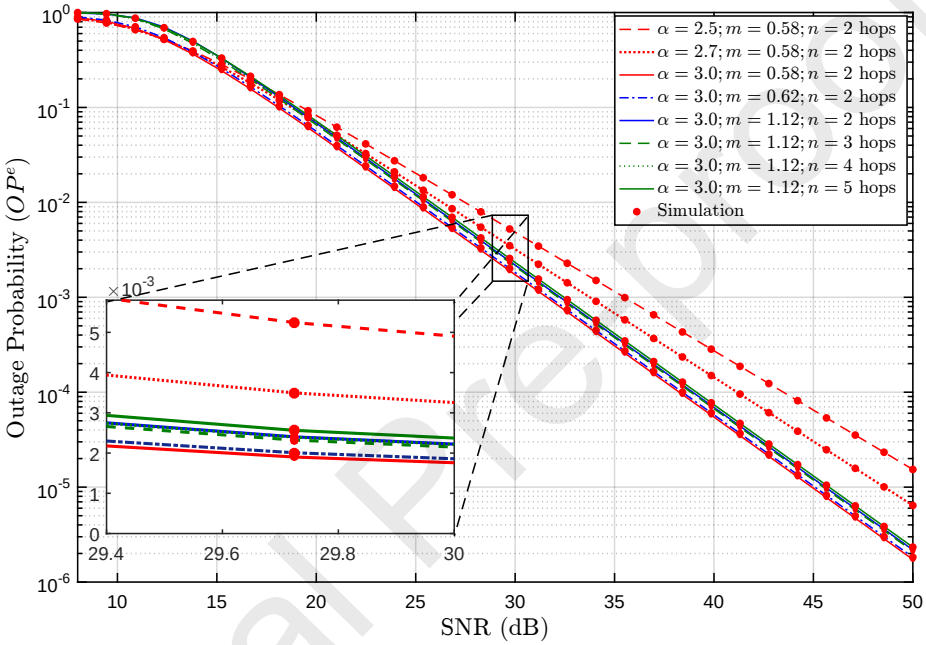


Fig. 4. OP curves for the $\alpha-\kappa-\mu$-Extreme case under severe fading conditions similar to Case I showing the effects of increments in hop number and fading model parameter values.

We now discuss simulation results of what is undoubtedly the most revealing of performance metrics, the BER. Figure 5 depicts the behaviour of BER under non-coherent DBPSK modulation scheme for **Case I** in moderate fading conditions. Consistent with observations from OP analyses, there is notable system improvement with increments in fading parameter values and a diminishing system performance improvement with the increase in hop number at low SNR levels. An additional system behavior is observed for high SNR level which shows that there is a non-null BER even as signal power increases indefinitely. This characteristic was observed by Rabelo et al. [13] for the linear $\kappa-\mu$-Extreme fading model where they showed that as SNR increases,

Journal Pre-proofs

BER under non-coherent DPSK modulation approaches the constant $(1/2)\exp(-2m)$. It is also worth pointing out that the application of multi-hops in the system is advantageous in the SNR window of 10 dB to 20 dB in the studied scenarios. Figure 6 presents BER curves under

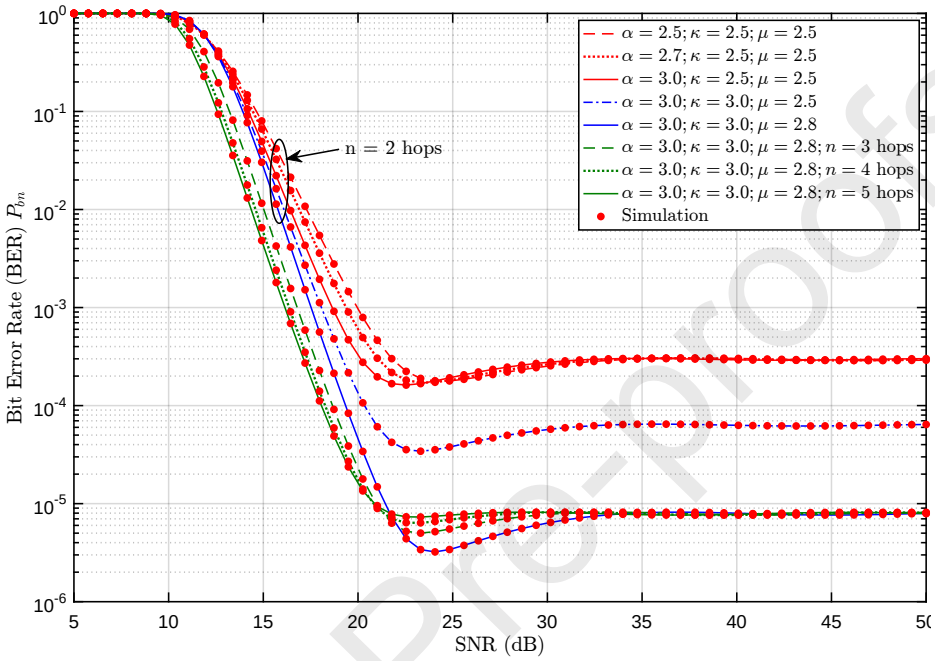


Fig. 5. BER curves for the $\alpha - \kappa - \mu$ case under non-coherent DBPSK modulation showing the effects of increasing model parameters and the number of hops and N in Poincare approximations for moderate fading conditions

non-coherent DBPSK modulation for **Case II**. We once more observe a consistent behaviour in terms of parameter and hop number variation. Furthermore, BER levels are considerably lower as compared to **Case I** even under moderate fading. This suggest that the $\alpha - \kappa - \mu$-Extreme fading model is much more robust and more flexible than its counterpart the, $\alpha - \kappa - \mu$ model. Figure 6 also shows that Poincare series-based asymptotic BER approximations perform quite well at low SNR regimes for the choice in parameter values, but the increase in N does not seem to improve convergence or guarantee improvement in accuracy of approximations. Moreover, an increase in hop number or μ results in a higher non-null BER level under **Case I** than for increases in m under Case II.

This is further supported by the system behaviour via BER curves under extremely severe fading conditions in Figure 7 and Figure 8 for **Case II** and **Case I**, respectively. From the two

Journal Pre-proofs

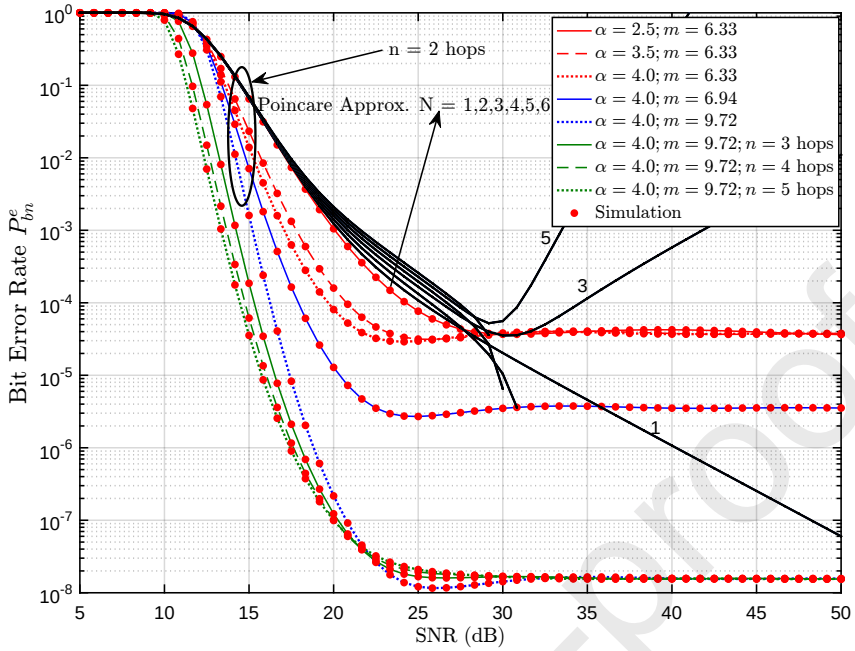


Fig. 6. BER curves for the $\alpha - \kappa - \mu$-Extreme case under non-coherent DBPSK modulation showing the effects of increasing model parameters, the number of hops and N in Poincare approximations for moderate fading conditions

figures, we note that while there appears to be some gain in system performance for increments in parameter values, system performance actually degrades as more relays are added to the system, with steeper penalties apparent for **Case I** than for **Case II**. Figure 9 depicts BER curves under Case II using coherent BPSK modulation scheme for moderate fading. From the figure, we note a consistent system behaviour as reported above with respect to parameter value and number of hop . Additionally, it is clear that one-term Poincare approximations improve with decrease in BER levels. It is also worth pointing out that the shapes of the curves for OP and BER are very similar to the shape of the Marcum Q-function curves. This can be attributed to the fact that both performance metrics are expressed as linear combinations and products of Marcum Q-functions. One of the interesting observation in all the BER plots is that irrespective of the hops number, the BER level is constant for initial values of SNR, as the SNR is further increased the benefit of using multi-hops system on the BER performance is quite visible and finally at higher SNR values all the plots become saturated at the same BER level. This gives a window of SNR values

Journal Pre-proofs

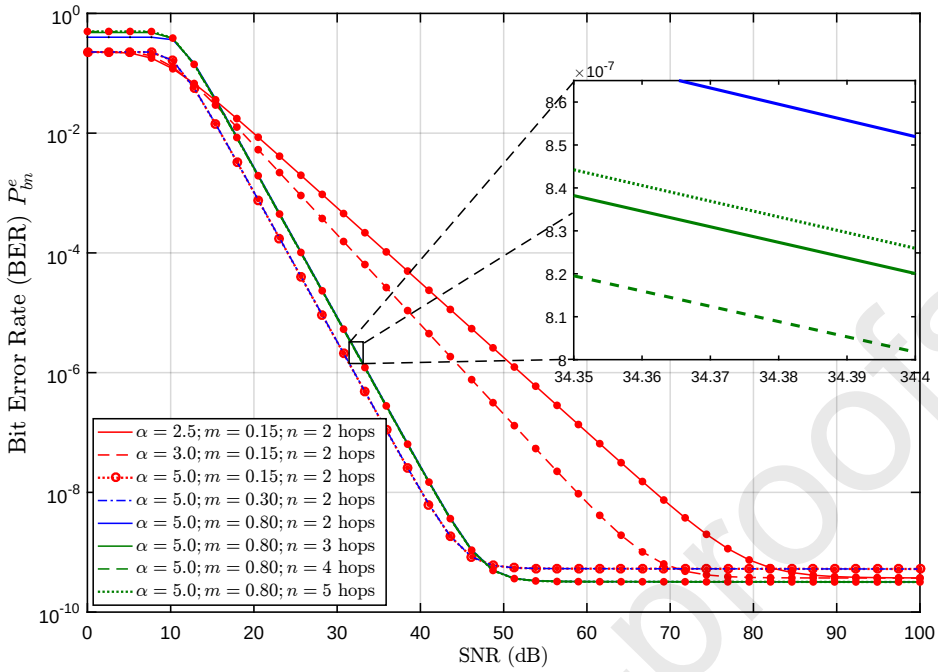


Fig. 7. BER curves for the $\alpha - \kappa - \mu$-Extreme case under non-coherent DBPSK modulation showing the effects of increasing model parameters and number of hops for extremely severe fading conditions

where the performance gains of multi-hop communication system can be explored.

VII. Conclusions

We have presented the closed-form expressions for the OP and BER for an end-to-end DF-based multi-hop relaying system over generalized $\alpha - \kappa - \mu$ and $\alpha - \kappa - \mu$-Extreme fading channels. The effects of the variation of the fading parameters and the number of hops on the system performance is demonstrated. As reported in similar literature, it was found that system performance improves with increase in any of the three parameters for both fading models, with increments in μ, the multi-path cluster parameter, resulting in steeper improvements as compared to increases in α or κ. Additionally, it was found that performance metrics based on the $\alpha - \kappa - \mu$-Extreme fading model appear to be better than those obtained for the $\alpha - \kappa - \mu$ fading model even under non-severe fading conditions. The results conclude that, DF multi-hop technique, which is used to extend the communication range of wireless systems also improves the end-to-end performance of the communication systems.

Journal Pre-proofs

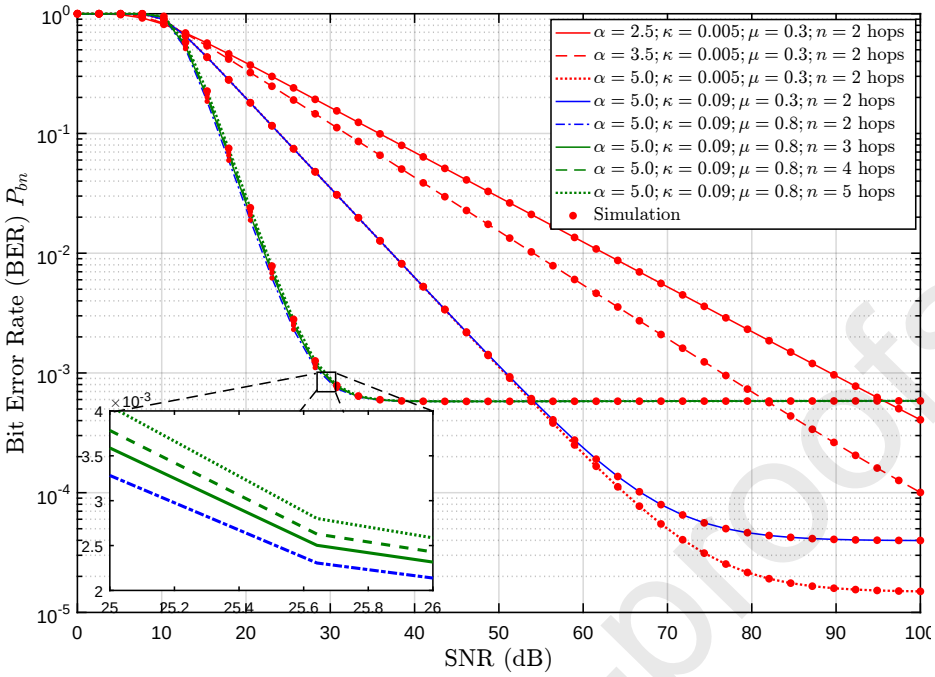


Fig. 8. BER curves for the $\alpha - \kappa - \mu$ case under non-coherent DBPSK modulation showing the effects of increasing model parameters and number of hops for extremely severe fading conditions

REFERENCES

[1] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3037–3063, 2005.

[2] S. N. Islam, P. Sadeghi, and S. Durrani, "Error performance analysis of decode-and-forward and amplify-and-forward multi-way relay networks with binary phase shift keying modulation," *IET Communications*, vol. 7, no. 15, pp. 1605–1616, 2013.

[3] D. T. J. Laneman and G. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

[4] G. Farhadi and N. Beaulieu, "On the ergodic capacity of wireless relaying systems over rayleigh fading channels," *IEEE Transactions on Wireless Communications*, vol. 11, no. 7, pp. 4462–4467, 2008.

[5] M. R. Bhatnagar, "On the capacity of decode-and-forward relaying over rician fading channels," *IEEE Communications Letters*, vol. 17, no. 6, pp. 1100–1103, 2013.

[6] L. Yang, M. O. Hasna, and M.-S. Alouini, "Average outage duration of multihop communication systems with regenerative relays," *IEEE Transactions on Wireless Communications*, vol. 4, no. 4, pp. 1366–1371, 2005.

[7] D. Dixit and P. Sahu, "Performance of multihop communication systems with regenerative relays in η-μ fading channels," in *Vehicular Technology Conference (VTC Spring), 2014 IEEE 79th*. IEEE, 2014, pp. 1–5.

[8] F. H. Tha'er, H. B. Salameh, and T. Aldalgamouni, "Performance study of multi-hop communication systems with decode-and-forward relays over α- μ fading channels," *IET Communications*, vol. 11, no. 10, pp. 1641–1648, 2017.

Journal Pre-proofs

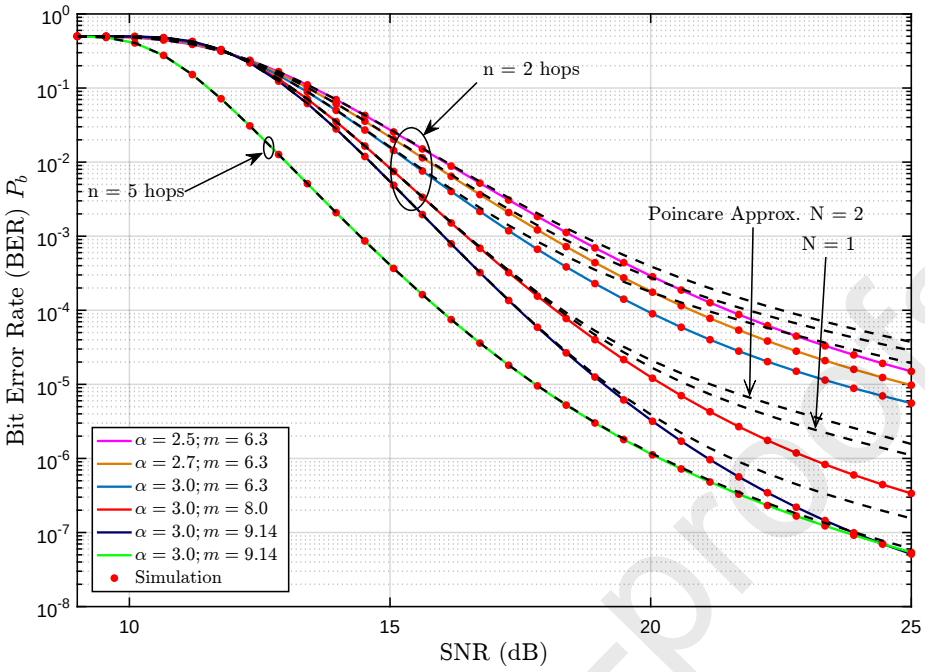


Fig. 9. BER performance for coherent BPSK modulation schemes over $\alpha - \kappa - \mu$-Extreme fading channel showing the variation in parameter values, number of hops and N in Poincare approximation

[9] D. Dixit and P. Sahu, "Exact closed-form aber for multi-hop regenerative relay systems over $\kappa - \mu$ fading," *IEEE Wireless Communications Letters*, vol. 6, no. 2, pp. 246–249, 2017.

[10] J. Cao, L.-L. Yang, and Z. Zhong, "Performance analysis of multihop wireless links over generalized-k fading channels," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 4, pp. 1590–1598, 2012.

[11] S. Kumar, "Performance of ED based spectrum sensing over α–η–μ fading channel," *Wireless Personal Communications*, vol. 100, no. 4, pp. 1845–1857, 2018.

[12] M. D. Yacoub, "The κ-μ distribution and the η-μ distribution," *IEEE Antennas and Propagation Magazine*, vol. 49, no. 1, pp. 68–81, 2007.

[13] G. S. Rabelo, U. S. Dias, and M. D. Yacoub, "The κ-μ extreme distribution: Characterizing severe fading conditions," in *Microwave and Optoelectronics Conference (IMOC), 2009 SBMO/IEEE MTT-S International*. IEEE, 2009, pp. 244–248.

[14] G. Fraidenraich and M. D. Yacoub, "The α-η-μ and α-κ-μ fading distributions," in *Spread Spectrum Techniques and Applications, 2006 IEEE Ninth International Symposium on*. IEEE, 2006, pp. 16–20.

[15] P. C. Sofotasios and S. Freear, "The α-κ-μ extreme distribution: Characterizing non-linear severe fading conditions," in *Telecommunication Networks and Applications Conference (ATNAC), 2011 Australasian*. IEEE, 2011, pp. 1–4.

[16] T. R. Rasethuntsa and S. Kumar, "An integrated performance evaluation of ed-based spectrum sensing over α-κ-μ and α-κ-μ-extreme fading channels," *Transactions on Emerging Telecommunications Technologies*, 2019.

[17] J. M. Moualeu, D. B. da Costa, W. Hamouda, U. S. Dias, and R. A. A. de Souza, "Performance analysis of digital

Journal Pre-proofs

communication systems over α - κ - μ fading channels," *IEEE Communications Letters*, vol. 23, no. 1, pp. 192–195, 2019.

[18] E. Salahat and A. Hakam, "Performance analysis of α - η - μ and α - κ - μ generalizedmobile fading channels," in *20th European Wireless Conference*, 2014, pp. 992–997.

[19] A. K. Papazafeiropoulos and S. A. Kotsopoulos, "Second-order statistics for the envelope of α - κ - μfading channels," *IEEE Communications Letters*, vol. 14, no. 4, pp. 291–293, 2010.

[20] B. Samudhyatha, S. Gurugopinath, and K. Saraswathi, "Maximum eigenvalue-based spectrum sensing over α - κ - μ and α - η - μ fading channels," in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2017, pp. 1–5.

[21] S. Kumar, M. Kaur, N. K. Singh, K. Singh, and P. S. Chauhan, "Energy detection based spectrum sensing for gamma shadowed α-η-μ and α-κ-μ fading channels," *AEU-International Journal of Electronics and Communications*, vol. 93, pp. 26–31, 2018.

[22] S. Kumar, P. Chauhan, P. Raghuwanshi, M. Kaur, and K. Singh, "Ed performance over α-η-μ/ig and α-κ-μ/ig generalized fading channels with diversity reception and cooperative sensing: A unified approach," *AEU-International Journal of Electronics and Communications*, vol. 97, pp. 273–279, 2018.

[23] N. C. Beaulieu and J. Hu, "A closed-form expression for the outage probability of decode-and-forward relaying in dissimilar rayleigh fading channels," *IEEE Communications Letters*, vol. 10, no. 12, 2006.

[24] M. K. Simon and M.-S. Alouini, *Digital communication over fading channels*. John Wiley & Sons, 2005, vol. 95.

[25] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series, and Products*. Elsevier Science, 2014.

[26] A. M. Mathai, R. K. Saxena, and H. J. Haubold, *The H-function: theory and applications*. Springer Science & Business Media, 2009.

[27] J. Proakis and S. M., *Digital Communications*. McGraw-Hill, 2005.

[28] B. Kumbhani and R. S. Kshetrimayum, *MIMO Wireless Communications over Generalized Fading Channels*. CRC Press, 2017.

[29] Y. A. Brychkov, O. Marichev, and A. Prudnikov, *Integrals and Series, vol 3: more special functions*. Gordon and Breach science publishers, 1990.

[30] A. M. Mathai and R. K. Saxena, *The H function with applications in statistics and other disciplines*. Wiley, 1978.

[31] K. Cao and X. Gao, "Solutions to generalized integrals involving the generalized marcum Q-function with application to energy detection," *IEEE Communications Letters*, vol. 20, no. 9, pp. 1780–1783, 2016.

[32] O. S. Badarneh and M. S. Aloqlah, "Performance analysis of digital communication systems over α-η-μ fading channels." *IEEE Trans. Vehicular Technology*, vol. 65, no. 10, pp. 7972–7981, 2016.

[33] A. Annamalai and E. Adebola, "Asymptotic analysis of digital modulations in $\kappa - \mu$, $\eta - \mu$ and $\alpha - \mu$ fading channels," *IET Communications*, vol. 8, no. 17, pp. 3081–3094, 2014.

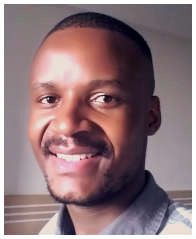
[34] R. Souza, A. Ribeiro, and D. A. Guimaraes, "On the efficient generation of α-κ-μ and α-η-μ white samples with applications," *International Journal of Antennas and Propagation*, vol. 2015, pp. 1–13, 2015.

[35] M. D. Yacoub, "The α-η-κ-μ fading model," *IEEE Transactions on Antennas and Propagation*, vol. 64, no. 8, pp. 3597–3610, 2016.

[36] H. Chergui, M. Benjillali, and M. Alouini, "Rician k-factor-based analysis of xlos service probability in 5g outdoor ultra-dense networks," *IEEE Wireless Commun. Lett.,*, vol. 8, no. 2, pp. 428–431, 2019.

Journal Pre-proofs

Author's Biographies



Tau Raphael Rasethuntsa received his BSc. in Mathematics and Statistics from the National University of Lesotho, Maseru, Lesotho in 2014 and MSc. in Mathematical Engineering from Istanbul Technical University, Istanbul, Turkey in 2017. His research interests include reliability, applied statistics as well as mathematical and statistical aspects of signal processing with special focus on performance evaluation of wireless communication systems.



Manpreet Kaur received her B. Tech. in Electronics and Communication Engineering from Punjab Technical University, Punjab, India in 2005 and Master of Engineering in Electronics and Communication from Thapar University, Patiala, India in 2007. She is pursuing her PhD. from Delhi Technological University, Delhi, India. She is currently working as Member (Senior Research Staff) at Central Research Laboratory, Ghaziabad, Bharat Electronics Limited. Her research interest includes the wireless communication and cognitive radios networks.



Sandeep Kumar received his B. Tech. in electronics and communication from Kurukshetra University, India in 2004 and Master of Engineering in Electronics and Communication from Thapar University, Patiala, India in 2007. He received his PhD. from Delhi Technological University, Delhi, India in 2018. He is currently working as Member (Senior Research Staff) at Central Research Laboratory, Bharat Electronics Limited Ghaziabad, India. His research interests include the study of wireless channels, performance modeling of fading channels and cognitive radio networks. He is also serving as a reviewer for several international journals of IEEE, Springer, Elsevier etc.

Journal Pre-proofs

Physical Communication 41 (2020) 101114

Contents lists available at ScienceDirect

Physical Communication

journal homepage: www.elsevier.com/locate/phycom

Full length article

Performance analysis of beamforming aided uncorrelated mm-Wave MIMO system for IEEE 802.15.3c standard

Anand Agrawal [a,*], Vimal Kumar Mishra [b], Ashish Gupta [b], Bajrang Bansal [b]

[a] Department of Electronics and Communication Engineering, Indian Institute of Information Technology, Kota, Rajasthan, 302017, India
[b] Department of Electronics and Communication Engineering, Jaypee Institute of Information Technology, Noida, U.P, 201301, India

ARTICLE INFO

Article history:
Received 25 October 2019
Received in revised form 17 February 2020
Accepted 1 May 2020
Available online 12 May 2020

Keywords:
Mm-wave communication
Analog beamforming
IEEE802.15.3c standard

ABSTRACT

Millimeter wave (mm-wave) at 60 GHz is a revolutionary technology for ultra-high data rate indoor communications. Unfortunately, due to unfavorable channel characteristics, communication at 60 GHz suffers from poor link budget, high penetration loss and severe signal attenuation. To improve the link budget and the quality of the received signal, the beamforming techniques must be incorporated. In this paper, we derive analytically tractable bit error rate (BER) for the analog beamforming aided mm-wave communication system over the modified IEEE 802.15.3c channel model. It also takes into account the impact of the Rake receivers, which are used to capture the signal energy carried by the multipath components. Monte Carlo simulations have also been carried out to validate the accuracy of the analytical framework. In addition, the impact of Rake's fingers, power decay factors and arrival times on the BER of the 60 GHz system is also highlighted.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Millimeter wave (mm-wave) communication at 60 GHz has shown enormous potential for the bandwidth-hungry applications, such as high definition video streaming and live gaming. It has also drawn great attention because of the large available bandwidth (7 GHz) [1,2]. In 2005, the Federal communications commission formally affirmed the 57–64 GHz frequency spectrum for unlicensed utilization of commercial applications. Specifically, at 60 GHz, this commission also formed a task group (TG) which has published several standards, targeting wireless personal and local area networks such as IEEE 802.15.3c, IEEE 802.11ad and ECMA 387 [3,4]. In 2007, TG (TG3c) carried out several measurements of the IEEE 802.15.3c standard and published their parametric specifications for different indoor environments [5]. Although there are quite a few merits of using a large continuous spectrum, 60 GHz spectrum faces various challenges. Some of them include unfavorable channel characteristics, large path loss (68 dB/m), attenuation due to rain, foliage and atmospheric oxygen [6,7]. This reduces the coverage area of the mm-wave technology and makes the spectrum not suited for long-range communication. Often, the required antenna spacing ($\lambda/2$) in mm-wave communication is of the order of millimeters which facilitates in integrating a large number of antennas array into a smaller area [8–11]. In mm-wave communication, these antenna arrays are combined at the transceivers which on using antenna beamforming technique can potentially overcome the poor propagation characteristics of the channel and can also improve the coverage area [12]. One of the major challenges in designing such transceivers is the implementation of the fully digital beamforming because it requires a separate radio frequency (RF) chain for each antenna element. This leads to a high hardware complexity and excessive power consumption. Therefore, people have suggested the analog and digital hybrid beamforming (HBF) architecture [13,14]. This architecture helps in directionalizing the signal transmission and reception by changing the amplitude and phase of the signal with the help of power variation and beam steering, respectively. The principle of HBF is to partition the conventional beamforming scheme into the digital as well as the analog beamforming. In digital beamforming, the amplitude or phase of the digital signal is varied at the transmit end. It also requires multiple RF chains which are proportional to the antenna elements in use. On the other hand, the analog beamforming is implemented by a set of variable phase shifters. Here, the amplitude/phase of the analog signal is varied at the transmit end. Unlike digital beamforming, it is restricted to one RF chain even when the large number of antenna arrays are being used. Hence, analog beamforming is best suited for coverage due to low power consumption & low cost characteristics. Moreover, it effectively addresses the hardware limitation of the fully-digital beamforming scheme [15].

* Corresponding author.
E-mail address: anand.ece@iiitkota.ac.in (A. Agrawal).
URL: http://www.iiitkota.ac.in/anand (A. Agrawal).

https://doi.org/10.1016/j.phycom.2020.101114
1874-4907/© 2020 Elsevier B.V. All rights reserved.

1.1. Related work

Several recent works have investigated the various performance measures of HBF aided mm-wave system. Hybrid precoding in mm-wave system has effectively reduced the number of RF chains but it limits the spatial multiplexing gain whereas, spatial modulation can convey an extra data stream via the index of the active antennas group. To further improve the achievable spectral efficiency (SE), the performance of spatial modulation aided mm-wave multiple-input-multiple-output (MIMO) system was studied in [16,17]. With this emphasis, in [18] authors have studied the spatially sparsed precoding based mm-wave MIMO system. Furthermore, the simulation and semi-analytical results of bit error rate (BER) for mm-wave system over narrowband fading channels were presented in [19–21]. The authors in [22–24] investigated the BER of mm-wave MIMO system over simplified IEEE 802.15.3c channel. The BER analyses discussed above have either considered the conventional narrow-band fading channels or simplified version of IEEE 802.15.3c channel models. It did not take into account the random nature of the total number of clusters, arrival times and the lognormal behavior of multipath gain coefficients [5]. This questions on the sustainability of the presented BER analyses, as these are very important aspects of wireless networks and can not be ignored while evaluating the performance of mm-wave MIMO systems at 60 GHz spectrum. Moreover, the original IEEE 802.15.3c channel model does not support beamforming technique because it fails to define the angle of departure (AOD) of the channel paths. Therefore, the modified version of the IEEE 802.15.3c channel model (define in the Section 2) was suggested to evaluate the performance of the 60 GHz system with beamforming architecture [25,26]. In [27], the authors calculated the average signal-to-noise ratio for an orthogonal frequency division multiplexing (OFDM) system over the modified IEEE 802.15.3c channel model.

1.2. Main contribution

To the best of the authors' knowledge, the BER analysis of mm-wave MIMO system with beamforming architecture has not been reported in the literature. The contribution of the paper can be summarized in the following bullet points:

- Theoretical analysis is carried out to drive a computable BER formula for analog beamforming aided mm-wave MIMO system over the modified version of the IEEE 802.15.3c channel model by using Gauss-Quadrature rule of numerical integration. This numerical rule facilitates in finding the computable solutions of mathematically intractable integrals [28].
- We have considered the impact of the Rake receiver which gives an insight of the number of Rake's fingers (L) required to capture the maximum signal energy carried by the multipath components (MPCs) in the given time window. Rake's fingers are nothing more than the set of Rake receiver, where each finger can independently tune a single multipath component.
- Characteristic function based probability density function (PDF) for approximate signal energy collected by the fingers of the Rake receiver in the given time window is also presented in this paper.
- We have also discussed the impact of various parameters of IEEE 802.15.3c channel on the BER. In addition, Monte Carlo simulations of the investigating system for IEEE 802.15.3c channel have also been carried out to validate the accuracy of the analysis.

1.3. Paper organization and notations

The rest of the paper is organized as follows: System and channel model are described in Section 2. Analytical BER expression is derived in Section 3. Section 4 presents the simulation results and discussions followed by the conclusion in Section 5.

Notations. Lowercase boldface letters represent the vectors. $\mathbb{C}^{m\times n}$ denotes an $m \times n$ matrix. $\psi_X(\cdot), f_X(\cdot)$ and $F_X(\cdot)$ respectively represent the characteristic function, PDF and cumulative distribution function of a random variable X. $\mathcal{N}(\mu, \sigma^2)$ represents Gaussian random variable with mean μ and variance σ^2, and $\mathrm{E}[\cdot]$ is the expectation operator. Furthermore, $|.|$, $\|.\|_2$ and $(.)^*$ are the absolute value, 2-norm and conjugate of a vector respectively; while $I_A(.)$ denotes an indicator function.

2. System and channel model

In this section, modified version IEEE 802.15.3c channel model and mm-wave MIMO system with antenna beamforming have been described.

2.1. Modified IEEE 802.15.3c channel model

IEEE 802.15.3c channel model is known as the modified Saleh-Valenzuela (SV) indoor channel model. It combines a line-of-sight (LOS) and non-LOS components of the S-V model using a two-path model [29] and gives the information about the spatial temporal phenomenon and angle-of-arrival (AOA), which have been widely used for describing the multi-path propagation behavior in the indoor environment. Besides, this model has been failed to define AOD of channel paths which is necessary for antenna beamforming [30–33]. Therefore, this analysis considers the modified version of the IEEE 802.15.3c channel, which has four parameters, including multipath gain coefficients, time of arrival (TOA), AOA and AOD. Here, AOD is assumed to have same statistics as that of the AOA. During propagation, multipath coefficients fall in the form of rays and clusters. If M is the total number of clusters and R is the number of rays falling in each cluster then the impulse response of the modified version of IEEE 802.15.3c channel is given by [34]

$$h(t,\theta,\phi)=\beta\delta(t)\,\delta(\theta)\,\delta(\phi)+\sum_{m=1}^{M}\sum_{r=0}^{R}\alpha_{r,m}\delta\left(t-T_m-\tau_{r,m}\right)\times\ \delta\left(\theta-\theta_m-\omega_{r,m}\right)\delta\left(\phi-\phi_m-\varpi_{r,m}\right) \quad (1)$$

where β represents the gain coefficient of the LOS path which depends on the antenna-height, antenna gain, Tx–Rx separation and path loss of the first impulse. The parameter $\alpha_{r,m}$ denotes the complex MPCs of the rth ray in the mth cluster and it is distributed as lognormal distribution. The parameters T_m, θ_m and ϕ_m are the TOA, AOA and AOD of the mth cluster respectively; while $\tau_{r,m}$, $\omega_{r,m}$ and $\varpi_{r,m}$ are the TOA, AOA and AOD for the rth ray of the mth cluster respectively. According to TG3c, the total number of clusters (M) is assumed to be random whose probability mass function (PMF) is given as

$$f_M(m)=\begin{cases}\left(1+\bar{L}\right)e^{-\bar{L}}, & m=1\\ \bar{L}^m e^{-\bar{L}}/m!, & m\geq 2\\ 0, & otherwise\end{cases} \quad (2)$$

where $\bar{L}=E[m]$. The propagation behavior of rays and clusters in the IEEE 802.15.3c channel is shown in Fig. 1. The figure clearly shows that the arrival times of the channel paths are random and independent to each other. The parameter Ω_0 is the average power of the first ray of the first cluster and Γ and γ are the

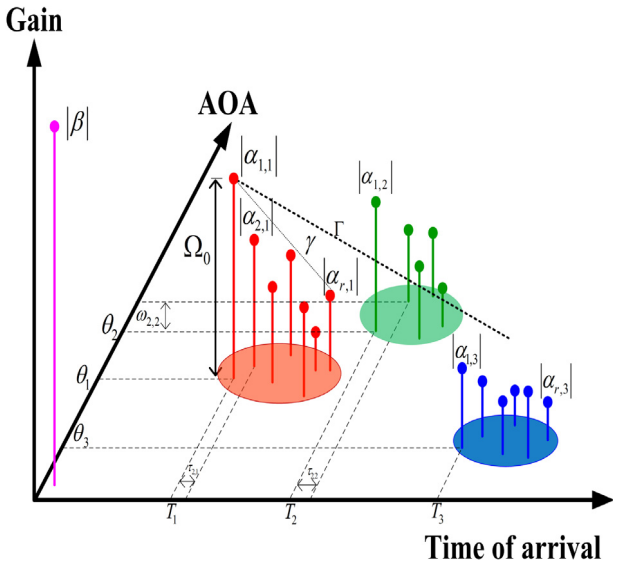


Fig. 1. Graphical representation of the multipath propagation in IEEE 802.15.3c channel.

inter-cluster and intra-cluster power decay factors respectively. According to TG3c channel report, the arrival times of the cluster and ray are modeled as Poisson processes. Therefore, the distribution of the cluster inter-arrival time and the ray inter-arrival time are modeled as exponential random variables, whose PDFs are given as

$$f\left(T_m/T_{m-1}\right) = \Lambda \exp\left[-\Lambda\left(T_m - T_{m-1}\right)\right], \quad m > 0 \tag{3}$$

and

$$f\left(\tau_{r,m}/\tau_{r-1,m}\right) = \lambda \exp\left[-\lambda\left(\tau_{r,m} - \tau_{r-1,m}\right)\right], \quad r > 0 \tag{4}$$

respectively, where Λ and λ denote the arrival rates of the cluster and ray respectively. In the angular domain, AOA of the LOS component, $\{\theta_{LOS}\}$, is assumed to be zero and the distribution of the cluster AOAs, $\{\theta_m\}$, is assumed to be independent and identically distributed (i.i.d.) over $[0,\ 2\pi]$. On the other hand, the distribution of the relative AOAs, $\{\omega_{r,m}\}$, is assumed to be independent Gaussian random variable with zero mean and standard deviation σ_ϕ, and its PDF is given as

$$f\left(\omega_{r,m}\right) = \frac{1}{\sqrt{2\pi}\sigma_\phi} \exp\left(-\omega_{r,m}^2/2\sigma_\phi^2\right) \tag{5}$$

The second-order moment of $\left(\alpha_{r,m}\right)$ can be calculated as (see Appendix for its proof)

$$\mathrm{E}\left[\left|\alpha_{r,m}\right|^2\right] = \Omega_0 \exp\left(-\frac{T_m}{\Gamma} - \frac{\tau_{r,m}}{\gamma} - \tilde{\kappa} I_{[\tilde{\kappa}\neq 0]}\right) \tag{6}$$

where $\tilde{\kappa}$ is the Rician factor.

On the basis of practical measurements, IEEE P802.15 wireless personal area network (WPAN) channel modeling subcommittee has studied 60 GHz network over various environments. These channel environments are further categorized according to the scenarios type (LOS/ NLOS) and the Half-Power Beam-Width (HPBW) of the transceiver antenna array. The detail description and the measured values of the channel's parameters for each environment are given in [5].

2.2. System model

An analog beamforming aided mm-wave MIMO system having N_t transmitting and N_r receiving antennas with a bank of Rake receivers is shown in Fig. 2. Initially, the mm-wave signal is split by the power diver and passed through the analog beamformer. The analog beamforming is implemented via phase shifters which have the provision to change the phase of the signals in each of the paths going towards the uniform linear antennas arrays (ULAs).

Note that the mm-wave technology at 60 GHz exclusively supports the rank one transmission, i.e., all transmitting antennas are fed with the same symbol at one time instant. The mathematical expression of binary phase shift keying (BPSK) modulated signal is given by [35]

$$s\left(t\right) = \sqrt{E_b} \sum_{i=-\infty}^{\infty} b_i\, p\left(t - iT_s\right) \tag{7}$$

where E_b denotes the signal energy per bit, $b_i \in \{\pm 1\}$ represents the ith transmitted symbol, $p\left(t\right)$ is the unit energy of the raised-cosine pulse of duration T_p, and T_s is the symbol duration. We denote $R_g\left(.\right)$ as the normalized autocorrelation function of $p(t)$, where $R_g\left(\delta\right) \approx 0$ for $|\delta| \geq T_p$. In the derivation, it is assumed that $T_s = T_p$ and the transmitting and receiving antennas are perfectly aligned to each other. Therefore, the effective mm-wave channel impulse response (CIR) after incorporating the ULAs at the transmitter and the receiver can be expressed as [18]

$$\boldsymbol{h}\left(t\right) = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \boldsymbol{h}\left(t,\theta,\phi\right) \boldsymbol{\vartheta}\left(\theta\right) \boldsymbol{\zeta}\left(\phi\right) d\theta d\phi \tag{8}$$

where $\boldsymbol{\vartheta}\left(\theta\right)$ and $\boldsymbol{\zeta}\left(\phi\right)$ represent the array response vector of ULAs at the transmitter and the receiver respectively. After satisfying $\|\boldsymbol{\vartheta}\left(\theta\right)\|_2 = 1$ and $\|\boldsymbol{\zeta}\left(\phi\right)\|_2 = 1$, these are respectively defined as

$$\boldsymbol{\vartheta}\left(\theta\right) = \frac{1}{\sqrt{N_t}}\left[1, e^{\left(\frac{j2\pi d \sin(\theta)}{\lambda_c}\right)}, \ldots, e^{\left(\frac{j2\pi (N_t-1) d \sin(\theta)}{\lambda_c}\right)}\right]^T, \tag{9}$$

and

$$\boldsymbol{\zeta}\left(\phi\right) = \frac{1}{\sqrt{N_r}}\left[1, e^{\left(\frac{j2\pi d \sin(\phi)}{\lambda_c}\right)}, \ldots, e^{\left(\frac{j2\pi (N_r-1) d \sin(\phi)}{\lambda_c}\right)}\right]^T, \tag{10}$$

where $d = \lambda_c/2$, λ_c is the wavelength of the carrier frequency and $\boldsymbol{h}\left(t\right) \in \mathbb{C}^{N_r \times N_t}$ is the effective CIR of mm-wave system. The received signal vector, $\boldsymbol{r}(t)$, can be obtained as

$$\boldsymbol{r}(t) = \boldsymbol{h}(t)\boldsymbol{s}(t) + \boldsymbol{\eta}(t) \tag{11}$$

where $\boldsymbol{r}(t) \in \mathbb{C}^{N_r}$, $\boldsymbol{s}(t)$ is a $(N_t \times 1)$ vector of the same transmitted signal $s(t)$, and $\boldsymbol{\eta}(t) \in \mathbb{C}^{N_r}$ is the Gaussian noise vector whose samples are $\mathcal{N}\left(0, \frac{N_t N_0}{N_r}\right)$. Since, all the transmitting antennas are independent and transmit one symbol at one time instant, the received signal at the vth antenna is given by

$$r_v\left(t\right) = \sqrt{\frac{E_b}{N_t N_r}} \sum_{i=-\infty}^{\infty} \sum_{u=1}^{N_t} b_i \left\{\beta_{u,v} p\left(t - iT_s\right) + \sum_{m=1}^{M}\sum_{r=0}^{R} \alpha_{r,m}^{u,v} p\left(t - \delta_{r,m}^{u,v} - iT_s\right) e^{\left(j\Phi_{r,m}^{u,v}\right)}\right\} + n_v\left(t\right) \tag{12}$$

where $\Theta_{r,m}^{u,v} = \frac{2\pi d}{\lambda_c}\left[(u-1)\sin\left(\theta_m^{u,v} + \omega_{r,m}^{u,v}\right) + (v-1)\sin\left(\phi_m^{u,v} + \varpi_{r,m}^{u,v}\right)\right]$ and $\delta_{r,m} = T_m + \tau_{r,m}$, $n_v\left(t\right) \sim \mathcal{N}\left(0, \frac{N_t N_0}{N_r}\right)$.

3. Proposed BER analysis

In this section, using the characteristic function based approach, we derive the BER of binary signals for analog beamforming aided mm-wave MIMO system with the bank of Rake

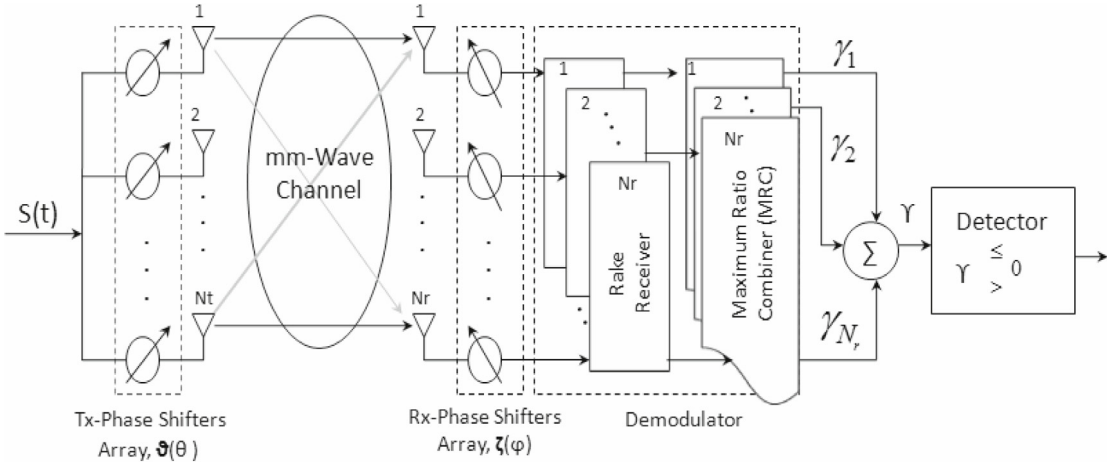


Fig. 2. Block diagram of phase shifters aided mm-wave MIMO system.

receivers. The average BER of binary signals with a coherent Rake receiver can be computed as [36]

$$p_e = \int_0^\infty Q\left(\sqrt{(1-\rho_r)\,\gamma}\right) f_\Upsilon(\gamma)\, d\gamma \tag{13}$$

where $\rho_r = -1$ and $\rho_r = 0$ for the antipodal signals and the orthogonal binary signals respectively, and $f_\Upsilon(\gamma)$ is the PDF of the received SNR (Υ).

3.1. Calculation of the total received instantaneous SNR (Υ)

According to [37], the template signal of rake receiver $\chi(t)$, under the assumption of perfect channel state information is available at the receiver and is given by

$$\chi(t) = \beta p\left(t - i'T_s\right) + \sum_{m'=1}^{M}\sum_{r'=0}^{R} \alpha_{r',m'} p\left(t - \delta_{r',m'} - i'T_s\right) e^{\left(j\Phi_{r',m'}\right)} \tag{14}$$

The output of the Rake receiver can be computed as $R_v = \int_{i'T_s}^{(i'+1)T_s} r_v(t)\,\chi_v^*(t)\,dt$, where $\chi_v^*(t)$ denotes the complex conjugate of the Rake receiver's template. Under the assumption of perfect synchronization, i.e., $(m = m')$ & $(r = r')$ and using the relation $R_g(\delta) = R_g(-\delta)$, the output of the correlator at the vth antenna can be evaluated as

$$R_v = \sqrt{\frac{E_b}{N_t N_r}} \sum_{i'=0}^{\infty}\sum_{u=1}^{N_t} b_{i'} \left(\beta_{u,v}^2 + \sum_{m=1}^{M}\sum_{r=0}^{R} 2\beta_{u,v}\,\alpha_{r,m}^{u,v} R_g\left(\delta_{r,m}^{u,v}\right) \times \cos\left(\Phi_{r,m}^{u,v}\right) + \sum_{m=1}^{M}\sum_{r=0}^{R}\left|\alpha_{r,m}^{u,v}\right|^2\right) + \tilde{n}_v(t) \tag{15}$$

where $\tilde{n}_v(t) = \int_{i'T_s}^{(i'+1)T_s} n_v(t)\,\chi^*(t)\,dt$ is the noise component. Assuming inter-pulse interference free transmission i.e. $R_g\left(\delta_{r,m}^{u,v}\right) \approx 0$, the above equation can be written as

$$R_v = \sqrt{\frac{E_b}{N_t N_r}} \sum_{i'=0}^{\infty}\sum_{u=1}^{N_t} b_{i'} \left(\beta_{u,v}^2 + \sum_{m=1}^{M}\sum_{r=0}^{R}\left|\alpha_{r,m}^{u,v}\right|^2\right) + \tilde{n}_v(t) \tag{16}$$

Considering all the receiving antennas to be sufficiently far away, the approximate received SNR (Υ) can now be viewed as the summation of the individual received SNRs at N_r antennas and is given by

$$\Upsilon \approx \frac{E_b}{N_t^2 N_0} \sum_{v=1}^{N_r}\sum_{u=1}^{N_t} \left(\beta_{u,v}^2 + \sum_{m=1}^{M}\sum_{r=0}^{R}\left|\alpha_{r,m}^{u,v}\right|^2\right) \tag{17}$$

3.2. PDF of the total received instantaneous SNR (Υ)

Let us define $\mathcal{E}_{u,v} = \sum_{m=1}^{M}\sum_{r=0}^{R}\left|\alpha_{r,m}^{u,v}\right|^2$ as the energy carried by all multipath rays excluding the LOS component in the $(u-v)$th link of modified IEEE 802.15.3c channel. According to [38], characteristic function, $\boldsymbol{\Psi}_\Upsilon(\nu)$, of Υ in i.i.d. mm-wave channel links can be computed by

$$\boldsymbol{\Psi}_\Upsilon(\nu) \approx \prod_{v=1}^{N_r}\prod_{u=1}^{N_t} \psi_{\mathcal{E}}^{u,v}\left(\frac{E_b}{N_t^2 N_0}\nu\right) \exp\left(\frac{jE_b\,\beta_{u,v}^2}{N_t^2 N_0}\nu\right) \tag{18}$$

where $\psi_{\mathcal{E}}^{u,v}(.)$ is the characteristic function of $\mathcal{E}_{u,v}$ which is defined in the next subsection. Using (16), PDF of the received SNR is calculated as $f_\Upsilon(\gamma) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \boldsymbol{\Psi}_\Upsilon(\nu)\, e^{(-j\nu\gamma)} d\nu$. This integration has no closed form solution and therefore, by using the Gauss–Hermite numerical quadrature rule, the approximate expression of $f_\Upsilon(\gamma)$ can be written as

$$f_\Upsilon(\gamma) = \frac{1}{2\pi}\sum_{k=1}^{N^H} \left\{w_k^H\right\} e^{\left(-j\gamma x_k^H + \left(x_k^H\right)^2\right)} \times \left(\prod_{v=1}^{N_r}\prod_{u=1}^{N_t} \psi_{\mathcal{E}}^{u,v}\left(\frac{E_b}{N_t^2 N_0} x_k^H\right) e^{\left(\frac{jE_b\beta_{u,v}^2}{N_t^2 N_0} x_k^H\right)}\right) \tag{19}$$

where x_k^H, w_k^H and N^H represent the kth root, kth weight and the number of nodes of the Gauss–Hermite quadrature rule respectively.

3.3. Calculation for $\psi_{\mathcal{E}}^{u,v}(\nu)$

For simplicity, we omit the antenna indexing (u, v) in the calculation of the characteristic function, $\mathcal{E}_{u,v}$, which is already defined in the above subsection. From the definition of parameter M which is given in (2), the average received energy $(\tilde{\mathcal{E}})$ with respect to parameter M can be computed as

$$\begin{aligned}\tilde{\mathcal{E}} &= \mathrm{E}_M[\mathcal{E}] \\ &= \sum_{m=1}^{\infty}\left(\sum_{\ell=1}^{m}\sum_{r=0}^{R}\left|\alpha_{r,\ell}\right|^2\right) f_M(m) \\ &\approx \left(1+\bar{L}\right)e^{-\bar{L}}\mathcal{E}_1 + \sum_{m\geq 2}\mathcal{E}_m\left(\frac{\bar{L}^m e^{-\bar{L}}}{m!}\right)\end{aligned} \tag{20}$$

where $f_M(m)$ is the PMF of M and parameters $\mathcal{E}_1 = \sum_{r=0}^{R}\left|\alpha_{r,1}\right|^2$ and $\mathcal{E}_m = \sum_{\ell=2}^{m}\sum_{r=0}^{R}\left|\alpha_{r,\ell}\right|^2$ are the energies carried by the

MPCs of the first cluster and the remaining clusters respectively. In this analysis, we conjecture that $\mathcal{E}_1$ is close to $\tilde{\mathcal{E}}_1$, the sum of the squared path gains of the first cluster fall within $[0, LT_c]$ time window [36], where L is the number of fingers of the Rake receiver and T_c is the chip duration between two consecutive fingers. Similarly, $\mathcal{E}_m$ is close to $\tilde{\mathcal{E}}_m$, and the sum of squared path gains of rays in the remaining clusters are defined as $\tilde{\mathcal{E}}_1 = \sum_{(r,1):0<(T_1+\tau_{r,1})\leq LT_c} |\alpha_{r,1}|^2$ and $\tilde{\mathcal{E}}_m = \sum_{(r,m):0<(T_m+\tau_{r,m})\leq LT_c} |\alpha_{r,m}|^2$. Note that $\tilde{\mathcal{E}}_1$ and $\tilde{\mathcal{E}}_m$ are assumed to be statistically independent to each other. Therefore, the characteristic function, $\psi_{\tilde{\mathcal{E}}}(\nu)$, of $\tilde{\mathcal{E}}$ can be evaluated as

$$\psi_{\tilde{\mathcal{E}}}(\nu) \approx \psi_{\tilde{\mathcal{E}}_1}(\zeta_1) \left\{ \prod_{m\geq 2} \psi_{\tilde{\mathcal{E}}_m}(\zeta_2) \right\} \tag{21}$$

where $\psi_{\tilde{\mathcal{E}}_1}(.)$ and $\psi_{\tilde{\mathcal{E}}_m}(.)$ are the characteristic functions of $\tilde{\mathcal{E}}_1$ and $\tilde{\mathcal{E}}_m$ respectively. The parameters $\zeta_1 = \left(1+\bar{L}\right) e^{-\bar{L}}\nu$ and $\zeta_2 = \frac{\bar{L}^m e^{-\bar{L}}\nu}{m!}$. From [39], $\psi_{\tilde{\mathcal{E}}_1}(\zeta_1) = \exp\left[-\lambda\tilde{\psi}_{\zeta_1}(0,L)\right]$ and $\psi_{\tilde{\mathcal{E}}_m}(\zeta_2) = \exp\left[-\Lambda\tilde{J}(\zeta_2, L)\right]$. According to [30, eq.(14)], function $\tilde{\psi}_{\zeta_1}(T, L)$ can be computed as

$$\tilde{\psi}_{\zeta_1}(T, L) = \begin{cases} \int_T^{LT_c} \left[1 - \mathcal{L}_{T,t}(\zeta_1)\right] dt, & \leq LT_c \\ 0, & > LT_c \end{cases} \tag{22}$$

where $\mathcal{L}_{T,t}(.)$ represents the characteristic function of the square of single path coefficient in the IEEE 802.15.3c channel, with ray arrival time $t = T + \tau$. Using the Gauss–Hermite quadrature, the Hermite solution of $\mathcal{L}_{T,t}(.)$ can be evaluated as

$$\mathcal{L}_{T,t}(\nu) \approx \frac{1}{\sqrt{\pi}} \sum_{k=1}^{N^H} w_k^H \exp\left(j\nu 10^{\left(\frac{\sqrt{2}\sigma x_k^H + \mu_{T,t}}{10}\right)} \right) \tag{23}$$

where $\mu_{T,t} = \frac{10}{\ln(10)}\left[\ln\Omega_0 - \frac{T}{\Gamma} - \frac{t-T}{\gamma} - \tilde{\kappa} I_{[\tilde{\kappa}\neq 0]} - \left(\frac{\ln 10}{10}\right)^2 \frac{\sigma^2}{2}\right]$ and $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$. The proof of $\mu_{T,t}$ is given in the Appendix. Now using Gauss–Legendre quadrature, the above characteristic function can be written as

$$\tilde{\psi}_{\zeta_1}(T, L) \approx \frac{LT_c - T}{2} \sum_{b=1}^{N^L} w_b^L \left[1 - \mathcal{L}_{T,t}(\zeta_1)\right]\Big|_{t=\frac{LT_c-T}{2}x_b^L + \frac{LT_c+T}{2}} \tag{24}$$

Similarly, function $\tilde{J}(\zeta_2, L)$ can be computed as

$$\tilde{J}(\zeta_2, L) \approx \frac{LT_c}{2} \sum_{b=1}^{N^L} w_b^L \left[1 - \mathcal{L}_{T,T}(\zeta_2)\, e^{-\lambda\tilde{\psi}_{\zeta_1}(T,L)}\right]\Big|_{T=\partial_b^L} \tag{25}$$

where $\partial_b^L = \frac{LT_c}{2}\left(x_b^L + 1\right)$ and $\left\{x_b^L, w_b^L, N^L\right\}$ are the bth root, bth weight and number of nodes of the Gauss–Legendre quadrature rule respectively. Combining (11), (17), and (19) and using Gauss Laguerre quadrature $\int_0^\infty e^{-x} f(x)\, dx = \sum_{a=1}^{M^L} w_a f\left(x_a^L\right)$, the derived BER expression is given by

$$\bar{p}_e \approx \frac{1}{2\pi} \sum_{i_1=1}^{M^L} \sum_{k=1}^{N^H} \left\{ w_{i_1}^L w_k^H e^{\left(x_k^H\right)^2} Q\left(\sqrt{\frac{(1-\rho_r)\, x_{i_1}^L}{j x_k^H}}\right) \times \prod_{v=1}^{N_r} \prod_{u=1}^{N_t} e^{\left[\frac{jE_b}{N_t^2 N_0}\beta_{u,v}^2 x_k^H - \lambda\tilde{\psi}_{\tilde{\kappa}_1}^{u,v}(0,L)\right]} \prod_{m\geq 2} e^{\left[-\Lambda\tilde{J}^{u,v}(\tilde{\kappa}_2, L)\right]} \right\} \tag{26}$$

where $\tilde{\kappa}_1 = \frac{E_b(1+\bar{L})e^{-\bar{L}}}{N_0 N_t^2} x_k^H$ and $\tilde{\kappa}_2 = \frac{E_b}{N_0 N_t^2}\frac{\bar{L}^m e^{-\bar{L}}}{m!} x_k^H$. The parameters $x_{i_1}^L$, $w_{i_1}^L$ and M^L represent the kth root, kth weight and the number of nodes of the Gauss–Legendre quadrature respectively.

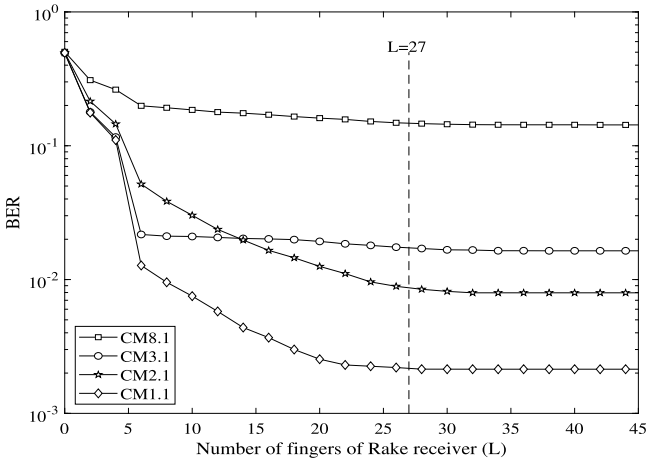


Fig. 3. Fingers of the Rake receiver vs BER of mm-wave SISO communication system at 5 dB SNR.

4. Simulation results and discussions

This section presents the simulated and analytical BERs of the mm-wave MIMO communication system for four different environments of the IEEE 802.15.3c channel model: Residential LOS (CM1), Residential NLOS (CM2), Office LOS (CM3) and Desktop NLOS (CM8). The parametric specifications of IEEE 802.15.3c model for these environments are given in the report published by the TG3c [5]. The simulation has been carried out for 10^6 samples of signal of interest and same has been averaged for 10^6 times. These signals are first transmitted by the highly directional antennas and then passed through the modified version of IEEE 802.15.3c faded channel. For analysis, we set $N^H = N^L = 60$, $T_c = 1$ ns, $M = 4$, $T_p = 0.5$ ns, $T_s = 0.5$ ns, sampling time $T_{samp} = 0.01$ ns and $\rho_r = -1$. The carrier frequency $f_c = 60$ GHz and the roll-off factor of the raised-cosine pulse is 0.8. The HPBW at the Tx is 360°, 360°, 30°, 30° while at the Rx it is 15°, 15°, 30°, 30° for CM1.1, CM2.1, CM3.1 and CM8.1 respectively. For simplicity, we have considered the statistics of AOD same as that of the AOA.

Fig. 3 shows the effect of L on the BER of SISO-mm-wave communication system, over various IEEE 802.15.3c channel models, at 5 dB SNR. It is observed that for different channel models, the BER becomes flat after certain values of L. These values are $L \geq 27$, $L \geq 27$, $L \geq 10$ and $L \geq 10$ for CM1.1, CM2.1, CM3.1 and CM 8.1 respectively. From this observation it is inferred that for CM1.1, the rays which carry most of the energy fall within first 27 ns, whereas for CM3.1, they fall within 10 ns. Therefore, for further simulations, we choose $L = 27$, for all four channel models so that the performance of all these four cases can be fairly compared.

Fig. 4 illustrates the simulated and analytical BER plots of the mm-wave SISO communication system for different environments of the IEEE 802.15.3c channel model. Clearly, the proposed BER (24) is in good agreement with the corresponding simulated plot, which indeed validates our analysis. It should be noted that there are some marginal differences between the analytical and the simulated BERs which are caused by: (a) the use of numerical integration rules like Gauss–Hermite and Gauss–Legendre quadrature and (b) while calculating the energy at the Rake receiver, the error due to the approximations used ($\varepsilon_1 \approx \tilde{\varepsilon}_1$ and $\varepsilon_m \approx \tilde{\varepsilon}_m$). From (24), it is clear that the aforementioned errors increase with increase in the size of the antenna array. For better understanding regarding the accuracy of the approximation, the approximate error plots (difference between the analytical and the simulated BERs) for different environments of IEEE 802.15.3c channel model are also shown in Fig. 5.

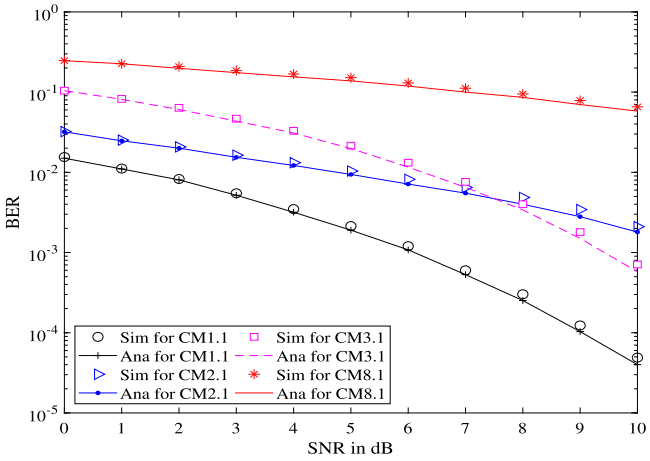


Fig. 4. BER (simulation and analytical) of mm-wave SISO communication system (Ana $--> $ Analytical and Sim $--> $ Simulation).

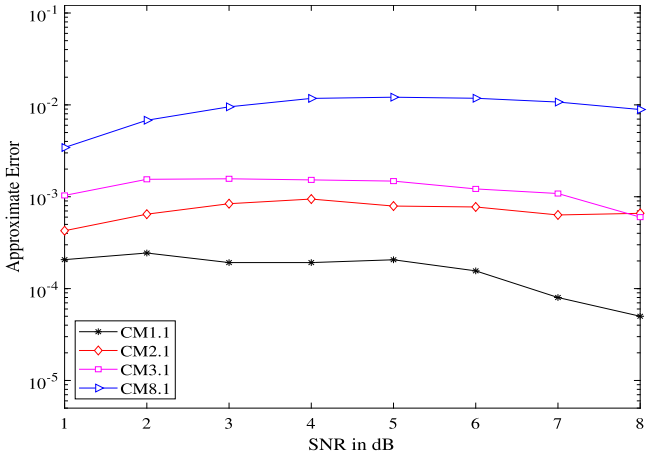


Fig. 5. Approximate error vs SNR: for mm-wave SISO system.

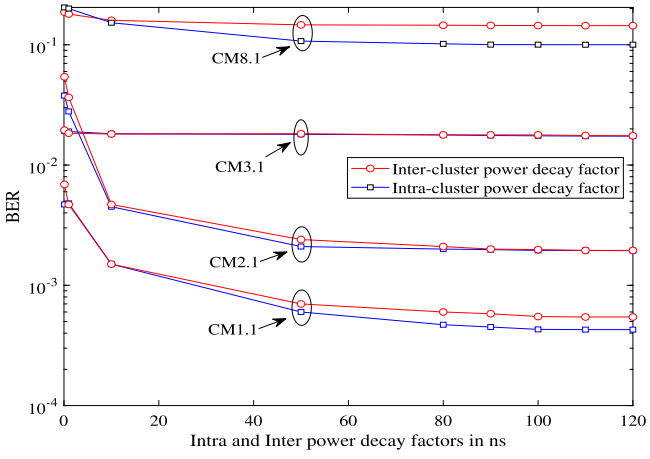


Fig. 6. BER vs $\{\Gamma$ & $\gamma\}$ in mm-wave SISO communication system at 5 dB SNR.

Fig. 6 shows the effect of power decaying factors on the BER of mm-wave SISO communication system, over various environments of the IEEE 802.15.3c channel model. It can be seen that initially the BER decreases with increase in Γ or γ, but after certain values of the power decay factor, the BER does not show any improvement. This phenomenon can be explained as follows:

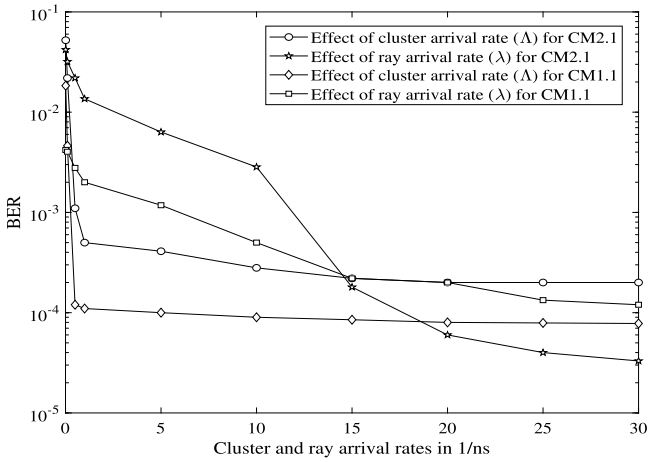


Fig. 7. BER vs $\{\Lambda$ & $\lambda\}$ in mm-wave SISO communication system at 5 dB SNR.

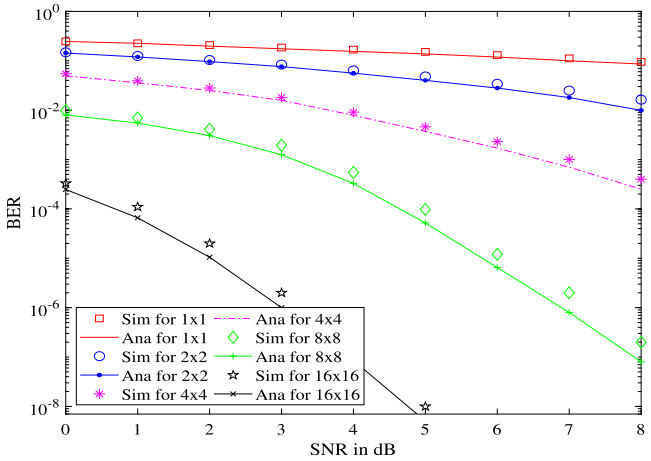


Fig. 8. BER (simulation and analytical) for $N_t \times N_r$ mm-wave MIMO system over IEEE 802.15.3c CM8.1.

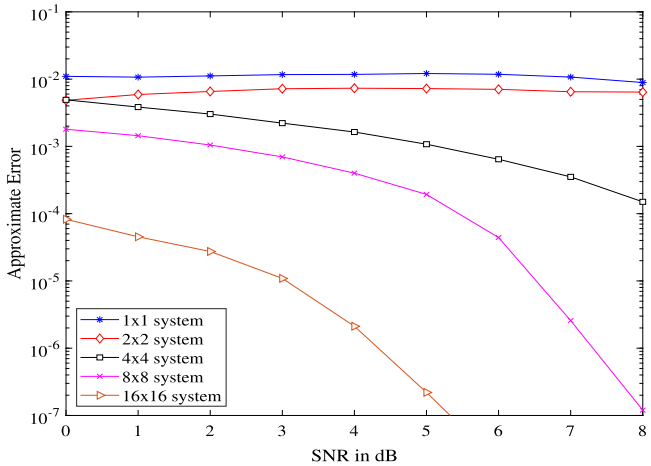


Fig. 9. Approximate error vs SNR: for $N_t \times N_r$ mm-wave MIMO system over IEEE 802.15.3c CM8.1.

From (6), it is clear that the energy carried by the MPCs increases with Γ or γ. As a result, the total energy captured by the Rake receiver (ε) also increases, which reduces the BER. However, after certain point of time, ε remains unchanged as most of the

energy carried by the MPCs has already been captured by the Rake receiver. Therefore, BER curves become flat.

The similar trends can be observed through Fig. 7, which shows the effect of the arrival rates on the BER. This is due to the fact that as Λ or λ increases, the arrival times of MPCs decrease and therefore more MPCs arrive in $[0,\ LT_c]$ time window which increases ε. But for large Λ or λ, most of the energy carrying MPCs arrive in $[0,\ LT_c]$ time interval, which causes the BER curves to get flattened. The results also indicate that the BER curves are less variant for a change in λ and reach the lower saturation value slower than the saturation value obtained while changing Λ. This is because the ray arrival rate influences the rays which are within the cluster whereas cluster arrival rate influences all the rays of the cluster.

For different antenna array configurations, Fig. 8 compares the simulated and the analytical BER for mm-wave MIMO system over CM8.1 channel model. It is noticed that as the size of the antenna array increases, the BER of the investigated system improves; however, the normalized difference between the simulated and the analytical BERs also increases with increase in the antenna array size. This is because the derived BER expression (24) is the function of the number of antennas used i.e. (N_r, N_t). As we increase the number of the transmitting and receiving antennas, the normalized error introduced between the simulated and the analytical BER is added up. Thus, an increase in the number of antennas, increases the error introduced by the analysis. Fig. 9 further gives an insight of this approximated error MIMO-mm-Wave system for CM8.1.

5. Conclusion

This paper derived a computable formula for the BER of binary signals incorporating analog beamforming aided mm-wave MIMO system using Gauss Quadratures rules of numerical integration. In the beamforming architecture, we have considered Tx-ULAs and Rx-ULAs to address the high path loss scenario experienced at 60 GHz. Apart from the analytical results of the BER, the simulations have also been carried out for different antenna arrays over the different environments of IEEE 802.15.3c channel model. From the results, we found that the analytical BER is in excellent agreement with the corresponding simulated results, which validates the suitability of the analysis. We have also considered the impact of the Rake receivers which are employed at the receiver, and suggested the sufficient number of Rake fingers required to capture the maximum energy of the mm-wave signal in various environments of the IEEE 802.15.3c channel model, at 60 GHz. In addition, we have also discussed the impact of various parameters of IEEE 80.15.3c channel on the performance of the proposed system. The presented analysis can be extended for Non-Orthogonal Multiple Access [40,41].

Acknowledgments

Authors would like to thank Prof. Rakhesh singh Kshetrimayum from Indian Institute of Technology Guwahti, India and Prof. Chi Chao-Chao from National Tsing Hua University in Hsinchu city, Taiwan for their valuable remarks and guidance.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

The complex multipath coefficient $(\alpha_{r,m})$ is written as

$$\alpha_{r,m} = \sqrt{\Omega_0}\, e^{\left(j\theta - \frac{\tilde{\kappa}}{2} I_{[\tilde{\kappa}\neq 0]}\right)}\, \xi_m\, \varrho_{r,m} \tag{A.1}$$

where θ is the uniform random variable over the interval $[0, 2\pi]$, parameters ξ_m, $\varrho_{r,m}$ are the amplitude coefficients of the mth cluster and the rth ray within the mth cluster, respectively and are distributed as independent lognormal random variables, i.e., $\xi_m \sim \ln\mathcal{N}\left(-\frac{T_m}{2\Gamma} - \sigma_1^2, \sigma_1^2\right)$ and $\varrho_{r,m} \sim \ln\mathcal{N}\left(-\frac{\tau_{r,m}}{2\gamma} - \sigma_2^2, \sigma_2^2\right)$. Parameters σ_1, σ_2 are the lognormal standard deviations of the cluster and ray respectively. Due to the random phase, the first order moment of complex MPCs is zero, i.e. $\mathrm{E}\left[\alpha_{r,m}\right] \approx 0$. The second order moment of complex MPCs can be computed as

$$\begin{aligned}\mathrm{E}\left[\left|\alpha_{r,m}\right|^2\right] &= \Omega_0 e^{-\tilde{\kappa} I_{[\tilde{\kappa}\neq 0]}} \mathrm{E}\left[\xi_m^2\right] \mathrm{E}\left[\varrho_{r,m}^2\right] \\ &= \Omega_0 e^{-\tilde{\kappa} I_{[\tilde{\kappa}\neq 0]}} e^{2\left[\left(\frac{-T_m}{2\Gamma} - \sigma_1^2\right) + \sigma_1^2\right]} e^{2\left[\left(\frac{-\tau_{r,m}}{2\gamma} - \sigma_2^2\right) + \sigma_2^2\right]} \\ &= \Omega_0 e^{\left(-\frac{T_m}{\Gamma} - \frac{\tau_{r,m}}{\gamma} - \tilde{\kappa} I_{[\tilde{\kappa}\neq 0]}\right)}\end{aligned} \tag{A.2}$$

If $X \sim \ln\mathcal{N}\left(\mu, \sigma^2\right)$, then the second order moment in the log scale is given as

$$\mathrm{E}\left[X^2\right] = e^{\left(2\left(\frac{\ln 10}{20}\right)\mu + 2\left(\frac{\ln 10}{20}\right)^2 \sigma^2\right)} \tag{A.3}$$

From (A.2) and (A.3), we have

$$\mu = \frac{10}{\ln 10}\left[\ln \Omega_0 - \frac{T_m}{\Gamma} - \frac{\tau_{r,m}}{\gamma} - \tilde{\kappa} I_{[\tilde{\kappa}\neq 0]} - \left(\frac{\ln 10}{10}\right)^2 \frac{\sigma^2}{2}\right] \tag{A.4}$$

Now, in order to fit in our formula, we just substitute T_m by T and $\tau_{r,m}$ by $(T - t)$ in (A.4), which yields

$$\mu = \frac{10}{\ln 10}\left[\ln \Omega_0 - \frac{T}{\Gamma} - \frac{T - t}{\gamma} - \tilde{\kappa} I_{[\tilde{\kappa}\neq 0]} - \left(\frac{\ln 10}{10}\right)^2 \frac{\sigma^2}{2}\right] \tag{A.5}$$

References

[1] P. Zhouyue, K. Farooq, An introduction to millimeter-wave mobile broadband systems, IEEE Commun. Mag. 49 (6) (2011) 101–107.
[2] B. Tuncer, S.C. Sean, L. Zhou, W. Junyi, R.M. Azizur, H. Hiroshi, K. Shuzo, IEEE 802.15. 3c: the first IEEE wireless standard for data rates over 1 Gb/s, IEEE Commun. Mag. 49 (7) (2011) 114–124.
[3] Y.S. Khiong, X. Pengfei, V.G. Alberto, 60 GHz Technology for Gbps WLAN and WPAN: From Theory to Practice, John Wiley & Sons, 2011.
[4] P. Cheolhee, R.S. Theordore, Short-range wireless communications for next-generation networks: UWB, 60 GHz millimeter-wave WPAN, and ZigBee, IEEE Wirel. Commun. 14 (4) (2007) 70–78.
[5] Y.S. Khiong, TG3c Channel Modeling Sub-Committee Final Report, IEEE802. 15-06-0195-07-003c, 2006.
[6] S. Harkirat, O. Jisung, K. Changyeul, Q. Xiangping, S.H. Rong, N. Chiu, A 60 GHz wireless network for enabling uncompressed video communication, IEEE Commun. Mag. 46 (12) (2008) 71–78.
[7] S.K. Mathew, R.S. Theodore, 3-D statistical channel model for millimeter-wave outdoor mobile broadband communications, in: Proc. IEEE Int. Conf. Commun., ICC, 2015, pp. 2430–2436.
[8] H. Sooyoung, K. Taejoon, L. David, K.V. James, A.T. Timothy, G. Amitava, Millimeter wave beamforming for wireless backhaul and access in small cell networks, IEEE Trans. Commun. 61 (10) (2013) 4391–4403.
[9] P. Liu, T. Jiang, Q. Zhang, M. Matthaiou, Pilot power allocation through user grouping in multi-cell massive MIMO systems, IEEE Trans. Commun. 65 (4) (2017) 1561–1574.
[10] H.Q. Ngo, L.-N. Tran, T. Duong, M. Matthaiou, E.G. Larsson, On the total energy efficiency of cell-free massive MIMO, IEEE Trans. Green Commun. Netw. 2 (1) (2018) 25–39.
[11] E. Bjornson, M. Matthaiou, M. Debbah, Massive MIMO with non-ideal arbitrary arrays: Hardware scaling laws and circuit-aware design, IEEE Trans. Wirel. Commun. 14 (8) (2015) 4353–4368.
[12] R. Wonil, S.J. Yun, P. Jeongho, L. Byunghwan, L. Jaekon, K. Yungsoo, C. Jaeweon, C. Kyungwhoon, A. Farshid, Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results, IEEE Commun. Mag. 52 (2) (2014) 106–113.

[13] R. Kilian, H. Pirzadeh, A.L. Swindlehurst, J.A. Nossek, A comparison of hybrid beamforming and digital beamforming with low-resolution ADCs for multiple users and imperfect CSI, IEEE J. Sel. Top. Sig. Process. 12 (3) (2018) 484–498.
[14] F. Sohrabi, W. Yu, Hybrid analog and digital beamforming for mmWave OFDM large-scale antenna arrays, IEEE J. Sel. Areas Commun. 35 (7) (2017) 1432–1443.
[15] R. Rakshith, H. Lajos, Hybrid beamforming in mm-wave MIMO systems having a finite input alphabet, IEEE Trans. Commun. 64 (8) (2016) 3337–3349.
[16] L. Peng, B. Jiri, P.N. Stefan, D.R. Marco, S. Andreas, Performance of generalized spatial modulation MIMO over measured 60 GHz indoor channels, IEEE Trans. Commun. 66 (1) (2018) 133–148.
[17] H. Longzhuang, W. Jintao, S. Jian, Spatial modulation for more spatial multiplexing: RF-chain-limited generalized spatial modulation aided MM-wave MIMO with hybrid precoding, IEEE Trans. Commun. 66 (3) (2018) 986–998.
[18] E.A. Omar, R. Sridhar, A.S. Shadi, P. Zhouyue, H.W. Robert, Spatially sparse precoding in millimeter wave MIMO systems, IEEE Trans. Wirel. Commun. 13 (3) (2014) 1499–1513.
[19] Y. Candy, S. Suresh, Empirical capacity of mmWave WLANs, IEEE J. Sel. Area Commun. 27 (8) (2009).
[20] X. Zhu, A. Doufexi, T. Kocak, On the performance of IEEE 802.15. 3c millimeter-wave WPANs: PHY and MAC, in: Proc. 6th IEEE Conf. on Wireless Advanced, WiAD, 2010, pp. 1–6.
[21] T.V. Nguyen, M. Elias, M.B. Laurence, Channel model and performance analysis of QAM multiple antenna systems at 60 GHz in the presence of human activity, in: Proc. IEEE Global Telecommun. Conf., GLOBECOM, 2011, pp. 1–6.
[22] Z. Liang, O. Yoji, Performance analysis of mmWave LOS-MIMO systems with uniform circular arrays, in: Proc. IEEE 81st Veh. Tech. Conf., VTC Spring, 2015, pp. 1–5.
[23] Y. Yavuz, S. Predrag, Reducing the LOS ray beamforming setup time for IEEE 802.11 ad and IEEE 802.15. 3c, in: Proc. IEEE Military Commun. Conf., MILCOM, 2016, pp. 448–453.
[24] N. Preeti, Analysis of Multiple Antenna Ultra-Wideband and Millimeter Wave Communication Systems (Ph.D. thesis), UC San Diego, 2007.
[25] C. Liu, S. Efstratios, P.S. Tony, E.J. Robin, Angle of arrival extended SV model for the 60 GHz wireless desktop channel, in: Proc. 17th IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun., 2006, pp. 1–6.
[26] G. Carl, H. Katsuyuki, W. Shurjeel, T. Fredrik, On mm-wave multipath clustering and channel modeling, IEEE Trans. Antennas and Propagation 62 (3) (2014) 1445–1455.
[27] H. Yueh, T. Hua, W.D. Wei, C.C. Chao, Performance analysis of OFDM systems over 60 GHz indoor Channels, in: Proc. IEEE Veh. Tech. Conf., VTC Fall, 2012, pp. 1–5.
[28] G.H. Gene, W.H. John, Calculation of Gauss quadrature rules, Math. Comp. 23 (106) (1969) 221–230.
[29] M. Matthaiou, P. de Kerret, G.K. Karagiannidis, J.A. Nossek, Mutual information staistics and beamforming analysis of optimized LOS MIMO systems, IEEE Trans. Commun. 58 (11) (2010) 3316–3329.
[30] A.M. Saleh, V. Reinaldo, A statistical model for indoor multipath propagation, IEEE J. Sel. Area Commun. 5 (2) (1987) 128–137.
[31] S.H. Quentin, J.D. Brian, J.A. Michael, S.A. Lee, Modeling the statistical time and angle of arrival characteristics of an indoor multipath channel, IEEE J. Sel. Area Commun. 18 (3) (2000) 347–360.
[32] S. Hirokazu, S. Yozo, C.C. Soon, Proposal of novel statistic channel model for millimeter wave WPAN, in: Proc. Asia-Pacific Microwave Conf., APMC, 2006, pp. 1855–1858.
[33] Gilb, P.K. James, IEEE Standards 802.15. 3c–Part 15.3: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for High Rate Wireless Personal Area Networks (WPANs) Amendment 2: Millimeter-Wave-Based Alternative Physical Layer Extension [S], IEEE Computer Society, New York, 2009.
[34] Y. Seokhyun, J. Taehyun, L. Wooyong, Hybrid beam-forming and beam-switching for OFDM based wireless personal area networks, IEEE J. Sel. Area Commun. 27 (8) (2009) 1425–1432.
[35] S.K. Marvin, A.S. Mohamed, Digital Communication Over Fading Channels, Vol. 95, John Wiley & Sons, 2005.
[36] W.L. Chun, L.W. Cheng, Bit error rate analysis in IEEE 802.15. 3a UWB channels, IEEE Trans. Wirel. commun. 9 (5) (2010) 1537–1542.
[37] B. Zhao, C. Yunfei, G.J. Roger, Average SINR analysis of DS-BPSK UWB systems with IPI, ICI, ISI, and MAI and its application, IEEE Trans. Veh. Technol. 58 (8) (2009) 4690–4696.
[38] A. Agrawal, R.S. Kshetrimayum, Analytical BER calculation of TAS/MRC-based two hop UWB communication system over IEEE 802.15.4a channel, IETE Tech. Rev. 35 (5) (2018) 467–475.
[39] H. Kei, G.A. John, The distribution of sums of path gains in the IEEE 802.15. 3a UWB channel model, IEEE Trans. Wirel. Commun. 6 (3) (2007) 811–816.
[40] X. Zhenyu, Z. Lipeng, C. Jinho, X. Pengfei, X. Xiang-Gen, Joint power allocation and beamforming for non-orthogonal multiple access (NOMA) in 5G millimeter wave communications, IEEE Trans. Wireless Commun. 17 (5) (2018) 2961–2974.
[41] Z. Lipeng, Z. Jun, X. Zhenyu, C. Xianbin, W.D. Oliver, X. Xiang-Gen, Millimeter-wave NOMA with user grouping, power allocation and hybrid beamforming, IEEE Trans. Wireless Commun. 18 (11) (2019) 5065–5079.



Dr. Anand Agrawal received his B.Tech degree from Samrat Ashok Technological Institute, Vidisha, India, in 2008, M.Tech degree from National Institute of Technology Bhopal, India, in 2010 and Ph.D from Indian Institute of Technology Guwahati, India in 2018. Since June 2019, he is working as an Assistant Professor in the department of Electronics and Communication Engineering at Indian Institute of Information Technology Kota, India. From Jan. 2018 to May 2019, he worked as an Assistant Professor in the department of Electronics and Communication Engineering at Jaypee Institute of Information Technology Noida , India. During PhD, he was a visiting research scholar at National Tsing Hua University (NTHU), Taiwan from Sept 2015 to Jan 2016. He has been involved in organizing several national/ international conferences in various capacities. He is a member of Journals Pub editorial board of, International Journal of Radio Frequency Design, International Journal of Broadband Cellular Communication and International Journal of Telecommunications & Emerging Technologies. He is also a reviewer of various reputed national and international journals including, Physical Communication, Wireless Personal Communication, Transaction of emerging Telecommunication technologies, IEEE communication letter and so on. His current area of research interest are UWB Communication, mmWave Communication at 60 GHz, Cooperative Communication, IEEE802.15 Standards, 5G Technology, Massive MIMO and D2D Communication. technology.



Dr. Vimal Kumar Mishra received his B.Tech. Degree in Electronics and Communication Engineering from Uttar Pradesh Technical University, Lucknow, India in 2009, M.Tech. Degree in Microelectronics and Embedded Technology from the Jaypee Institute of Information Technology, Noida, India in 2012, and Ph.D. in VLSI Design from Madan Mohan Malaviya University of Technology, Gorakhpur, India in 2017. Currently, he is an Assistant Professor at Jaypee Institute of Information Technology, Noida, India. His research interests include semiconductor physics and device modeling. He has filed 2 patents and published more than 25 papers in the reputed journals and conferences.



Dr. Ashish Gupta was born on 4th August 1986. Currently, he is an Assistant Professor in Jaypee Institute of Information Technology, Sector 128, Noida. He has done his Ph.D. in RF and Microwave Engineering from Department of Electronics Engineering, Indian Institute of Technology (Indian School of Mines) Dhanbad in 2017. He did his MTech from ShriGovindramSeksaria Institute of Technology and Science, Indore in 2010, where he got fellowship from MHRD Govt. of India. He won the best paper award in ICAICV-2010 at Coimbatore. He has done his B.Tech. from Rustumji Institute of Technology, B.S.F. Academy, Tekanpur in year 2008. From year 2010 to 2013, he was associated with different engineering institutions as an Assistant Professor, in Gujarat, India. He is the author of more than 15 articles in the relevant area. His current research interests involve Metamaterials and its applications in active and passive devices. He is also serving as reviewer in few peer-reviewed journals of repute.



Dr. Bajrang Bansal received his B.E. Degree in Electronics and Communication Engineering from Institute of Technology and Management, Gurgaon, India in 2005, M.Tech. Degree in VLSI Design and CAD from the Thapar University, Punjab, India in 2008, and Ph.D. in Wireless Communication from Delhi Technological University, Delhi, India in 2017. Currently, he is an Assistant Professor at Jaypee Institute of Information Technology, Noida, U.P, India. His research interests include wireless channel modeling and UWB.

International Journal of Progressive Research in Science and Engineering
Volume-1, Issue-3, June-2020
www.ijprse.com

Performance of Masonry Infill Framed Structures During Earthquake: A State-Of-The-Art Review

Priyanka[1], Alok Verma[2], Shobha Ram[3]

[1]*Student, Department of Civil Engineering, Gautam Buddha University, India.*

[2]*Professor, Department of Civil Engineering, Delhi Technological University, India.*

[3]*Professor, Department of Civil Engineering, Gautam Buddha University, India.*

Corresponding Author: priyankap780@gmail.com

Abstract: - Masonry infill walls are widely known to increase lateral stiffness of the structure and for this reason it is accepted all over the world. This paper presents a review work on performance of infill framed structures that were damaged during several earthquakes. Study of behavior of damaged building during different earthquakes in world has been carried out. The mentioned earthquakes substantially caused damaged to the RC buildings. The RC buildings were damaged primarily because of improper design and reinforcement detailing at the design phase and improper workmanship and quality control at the construction phase. The main objective of this paper is to describe and analyze the failure patterns observed in reinforced concrete frame buildings with masonry infill walls and without masonry infill walls all over the world.

Key Words: — ***Infill Walls, Earthquake, Non Structural Elements, Failure Mechanism.***

I. INTRODUCTION

The present paper is concerned with the performance of buildings all over the world primarily due to ground shaking. From the past researches and field investigations by various researches it can be noted that building with masonry infill walls (or unreinforced masonry infill walls) perform better in earthquake. As compared to bare RC frame masonry infill frames can carry more seismic effect. From study it can be concluded that masonry infill walls or panels play very crucial role with respect to severe ground motions. Many structural deficiencies are highlighted here which were caused by earthquake, which consists of soft and weak stories, strong-beam and weak column, etc.

This paper mentions five great earthquakes that struck the world. Among these, two are Bhuj (2001) and Sumatra (2006) earthquake which happened in India. Bhuj earthquake is the largest earthquake that struck India. The remaining earthquakes are being mentioned here, Wenchuan Earthquake (2008), L'Aquila (2009) and Lorca (2011) earthquake.

II. LITERATURE REVIEW

Humar J. M. et al. (2001) and Sudhir K. Jain et al. (2001) reported on the performance of building during Bhuj earthquake that were damage during the earthquake. They both individually investigated that the open storey were severely damage as compare to that of infill storey. Hemant B. Kaushik and Sudhir K. Jain et al. (2004) reported that the great Sumatra earthquake of 26 December 2004 caused damages to RC buildings at Port Blair in the Andaman Islands in India. Li, B., Wang, Z., Mosalam, K. M., & Xie, H. et al. (2008, October) carried out field reconnaissance after Wenchuan earthquake in Dujiangyan, Hanwang, and Yingxiu, to investigate the types of damage occurred in the infill walls and/or the bounding frames. Manfredi and masi reported (Manfredi, V., & Masi, A. et al., (2011, 2014)) on the key role of the non-structural element in building during L' Aquila 2009 earthquake. They put forward an infill model capable of accounting for the effects arising from in-plane (IP) damage on the out-of –plane (OOP) capacity of infill panels. Hermanns, L., Fraile, A., Alarcón, E., & Álvarez, R. et al., (2014) analyzed the failure patterns observed in reinforced concrete frame buildings with masonry infill walls.

III. AN OVERVIEW

A. Bhuj Earthquake:

It is seen that most of the modern RC buildings that were collapsed in the damaged area were open at ground storey for the purpose of parking. The buildings which had few or no infill walls that were damaged are shown in figure 1(a). Open storey led to insufficient strength and stiffness of the structure resulting in failure of structure as shown in figure 1(b). Whereas it was observed from the study that the buildings with full scale infill walls at ground storey withstand earthquake without collapse.

IJPRSE
Progressive Research
International Journal of Progressive Research in Science and Engineering
Volume-1, Issue-3, June-2020
www.ijprse.com

Earthquake Name	Time And Date	Location	Magnitude
Bhuj Earthquake	8:46 Am, January 26, 2001	Bhuj, Gujarat	M7.9
Sumatra Earthquake	6:28:53 Am, December 26, 2004	Port Blair, Andaman Islands	M9
Wenchuan Earthquake	2:28 Pm, May 12, 2008	Northwestern Sichuan, China	M8
L' Aquila Earthquake	3:32:52 (Local Time) , April 6, 2009	Abruzzo, Italy	M6.3
Lorca Earthquake	6:47pm(Local Time), May 11, 2011	Lorca, South-East Spain	M5.1



Fig.1(a). Precariously balance severely damaged RC frame buildings with open ground storey in Ahmadabad



Fig.1(b). Ground storey collapse of a four storey building with open ground storey at Bhuj

B. Sumatra Earthquake:

In the great Sumatra earthquake many damaged pattern were observed due to infill panel among these the common damages were open storey damage, OOP failure of masonry infill walls.

The buildings that were constructed without engineering supervision and were located in silts were collapsed partially or completely. In a collapsed 3-storey building at Naya Gaon, the upper stories suffered only nominal damages in the RC frame and masonry infill (Fig. 2a). Some recently constructed buildings on stilts at Bamboo Flat collapsed completely (Fig. 2b). It was also noticed that a 2-storey Police-Barrack Government owned building, which had the open first storey for storage purpose collapsed at Haddo wharf (Fig. 2c), whereas the adjacent infill buildings in the first storey survived the ground motions. The passenger terminal building at Haddo Wharf (Fig. 2d), was constructed under the supervision of a construction firm gets partially collapsed as it has open masonry infill panels for the purpose of ventilation and in the inner periphery infill walls were not present for the purpose of open space. Due to poor masonry quality and loose frame joints OOP failure was observed.

Figure 3 shows OOP failure of masonry infills in buildings at (a) Mohanpura (school building), (b) Bamboo flat, and (c) Passenger Terminal Building at Haddo wharf.

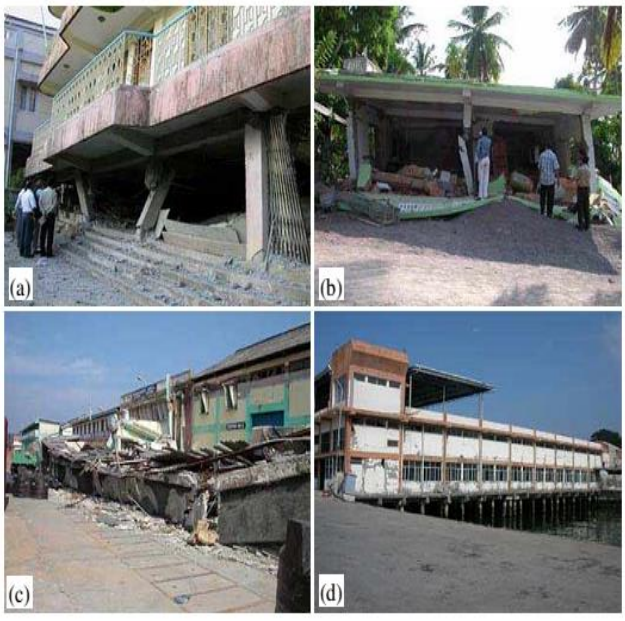


Fig.2. Collapse of RC buildings at (a) Naya Gaon, (b) Bamboo Flat, and (c) Haddo wharf, and (d) partially collapsed Passenger Terminal Building at Haddo wharf.

IJPRSE
Progressive Research
International Journal of Progressive Research in Science and Engineering
Volume-1, Issue-3, June-2020
www.ijprse.com



Fig.3. shows OOP failure of masonry infills in buildings at (a) Mohanpura (school building), (b) Bamboo flat, and (c) Passenger Terminal Building at Haddo wharf.

C. Wenchuan Earthquake:

It was observed that many residential and commercial buildings in the epicenter suffered severe damage due to soft storey effect of infill walls. Many buildings had open storey which serve as the purpose of parking. Many buildings had less or no infill walls. Many buildings were made using hollow clay tiles and gas concrete masonry infill walls. It was observed that damage to masonry infill walls was concentrated in the lower stories of buildings where less or no infill walls were provided. It can be noticed from the results that hollow infill walls contribute to stiffness but less strength. Arrangement of infill panels in orthogonal direction to create residential space resulted in creating stiffness discontinuities. Figure 4 shows a six-storey building where the first storey was available for the purpose of parking and had less infill walls while the upper stories are residential and has many infill walls.



Fig.4. Formation of weak storey: leaned 6-story building in Dujiangyan

D. L' Aquila Earthquake:

During past earthquakes in Italy these non-structural elements have contributed significantly in resisting lateral loads. Figure 5 shows example of masonry panel expulsions from the structural frame, resulting from a poor connection.



Fig.5. Examples of out-of-plane collapse of infills poorly connected to the RC frame

E. Lorca Earthquake:

Figure 6 shows one of the parapet walls that collapsed during lorca earthquake which shows the importance of studying the behaviour of non-structural components. It can be noted that acceleration at roof level is significantly higher than that at ground level. Figure 7 shows soft storey mechanisms due to infill panels with lower stiffness at ground storey.



Figure.6. Building with a partially collapsed parapet wall



Fig.7(a). Building with soft storey



Fig.7(b). Severely damaged column

IV. CONCLUSION

From the above study it can be contributed that masonry infill walls contributes significantly in increasing the lateral stiffness of the structure and increasing the strength of the structure. Also it can be noticed that connection between frames and infill walls play important role in failure pattern. No infill panel or less infill panel leads to soft storey mechanism in which flexibility of the structure is reduced. Hence from the observation it can be mentioned that it is important to provide good quality masonry infill panels under engineer's supervision.

REFERENCES

[1]. Li, B., Wang, Z., Mosalam, K. M., & Xie, H. (2008, October). Wenchuan earthquake field reconnaissance on reinforced concrete framed buildings with and without masonry infill walls. In The 14th World Conference on Earthquake Engineering. Beijing, China.

[2]. Braga, F., Manfredi, V., Masi, A., Salvatori, A., & Vona, M. (2011). Performance of non-structural elements in RC buildings during the L'Aquila, 2009 earthquake. Bulletin of Earthquake Engineering, 9(1), 307-324.

[3]. Hermanns, L., Fraile, A., Alarcón, E., & Álvarez, R. (2014). Performance of buildings with masonry infill walls during the 2011 Lorca earthquake. Bulletin of Earthquake Engineering, 12(5), 1977-1997.

[4]. Murty, C. V. R., Dayal, U., Arlekar, J. N., Chaubey, S. K., & Jain, S. K. A National Technological Crisis: Structural and Geotechnical damages sustained during M7. 9 Bhuj earthquake force. Department of Civil Engineering, IIT Kanpur.

[5]. Kaushik, H. B., & Jain, S. K. Performance of Buildings In Port Blair (India) During The Great Sumatra Earthquake Of 26 December 2004.

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

PMS-Sharing : Framework for Automatically Authenticating users in a Group to Allow Sharing Storage

Nitesh Suthar
System R&D – Convergence Group
Samsung Electronics
Noida, India
s.nitesh@samsung.com,
suthar.nitesh@gmail.com

Rajesh Kr. Yadav
Dept. of Computer Science & Engineering
Delhi Technological University
Delhi, India
rkyadav@dtu.ac.in

***Abstract*—Cloud storage is a very famous buzz word nowadays. There are many cloud storage service providers in the market and millions of mobile users are accessing cloud storage to store and access various personal data such as photos, videos and documents. In this era of smart mobile devices, every user is having devices with various storage capacity i.e. 8, 16, 32, 64, 128 GB and even more storage is going to be embedded in next generation of devices. These storage are widely used for storing and sharing data among multiple people. In this paper, a model and framework is proposed to automatically authenticate users within a group to allow sharing data while connected to same wireless network where initial group is formed by user (group owner) to allow set of users to access shared contents thereafter users in a group are automatically authenticated. The proposed solution is called personal mobile storage sharing (PMS-Sharing). The application and framework is developed and tested on android mobile phones supporting Android version 4.4 and above.**

Keywords—shared storage; network; mobile computing; cloud storage; IoT

I. INTRODUCTION

In this era of smart mobile devices, millions of users are using smartphones and tablets; these smartphones are equipped with high-performance processors, RAM for computing, gigabytes of inbuilt storage memory and additional SD card slot to extend the storage capacity. Substantial growth of smartphone devices also increases the need of Cloud Storage, with the roll-out of 4G and 5G technology, it is assumed that there will be no issue of connectivity and smartphone market is also increasing day by day. Many companies are engaged in developing new smartphones with new and improved capabilities with reasonable price. This motivates smartphone users to upgrade their smartphone to the advanced one.

PMS-Sharing provides a shared space for a group of users to share files (documents, images, videos etc.) among each other while connected to the same network. The Personal Mobile Storage Sharing (PMS-Sharing) allows user to form a local group of users and allows them to access shared storage directory. Currently if user needs to share any file, he has to do it manually by transferring file via Bluetooth or Wi-Fi even though all users are connected to the same local network (WLAN Network). PMS-Sharing overcomes this problem of manual transferring and authenticating. In PMS-Sharing, a framework is introduced for automatically authenticating users in a group to allow sharing storage. RSA encryption algorithm is systematically used along with randomly generated pair key. Pair key is unique for each user, helps Group Owner device to identify and ensures that an authenticate user is accessing the group's data. Public key cryptography based on Rivest–Shamir–Adleman (RSA), is used for encrypting messages to securely exchange pair key with designated user. It is using two level of encryption one at group level and another at user level.

In PMS-Sharing, the user can share the directory among all other users in a group and other user can download and view the files of their choice easily by clicking on file name at displayed in list of files. User can do multiple functions like add, delete and download view or play files. PMS-Sharing will introduce a new framework to utilize the unused mobile phones as storage nodes to build big storage space which can be deployed as a shared storage on distributed network. The application is developed and tested on Android mobile phones using existing infrastructure. In section II, Related work –cloud storage types, service providers, existing cloud storage, key challenges faced and limitations imposed by cloud storage system are discussed. In section III, Proposed Solution – the detailed architecture and framework of PMS-Sharing application are discussed, the algorithm for identification and authentication of user who belong to a particular group are also discussed. In section IV, Conclusion and Future Scope – the current challenges and future scope of this proposed work are discussed.

II. RELATED WORK

A. Cloud Storage System

Cloud Storage means storing data in a remote location which is accessible on any device. Cloud Storage improves efficiency and productivity in terms of backing up and securing

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:02:08 UTC from IEEE Xplore. Restrictions apply.

the data. It is beneficial and Business can pay only for the amount of storage they require [1]. There are different kinds of cloud storage system like a private, public, hybrid, many cloud storage providers provide service to store user data, and they charge pricing for these services.

Cloud Storage is categorized, based on how data is managed on servers- 1. File Storage, 2. Block Storage, and 3. Object storage. There are many types of File Storage with different characteristics like structure, access method, speed, security, and size. File storage is applied on local as well as network storage devices. Some prominent network file storage is Network File System (NFS), Server Message Block(SMB)/ Common Internet File System (CIFS) and Plan 9 File system protocol (9P)[2], [3]. Block Storage provides a way to store data in an evenly-sized memory blocks to host the database, which supports random read and write operations [3], [4]. Amazon Elastic Block Storage (EBS) is developed for use with Amazon Elastic Compute Cloud (EC2) to deploy a broad category of data loads such as databases, containerized applications, big data analytics engine, file system, media data, and enterprise application data [5]. Object Storage provides an easy way to store unstructured data in the form of Objects. An object is a simple repository with the unique ID number, metadata, and actual data [6]. Amazon Simple Storage Service (Amazon S3) is capable of managing a significant amount of unstructured data. Table I represents the basic differences between file storage, block storage, and object storage system.

B. Existing Cloud Storage Providers and Concept

IDrive [7], pCloud [8], Mega [9], OneDrive [10], iCloud [11], Google [12], Box [13], NextCloud [14], and SpiderOak [15] are some of the cloud storage service providers which provide few GB's of storage free of cost. Cloud storage demand is increasing in proportion to the number of mobile device users are increasing each day.

GrandStore [16] proposed a system where user can store the credential of all cloud storage accounts to expand the free storage capacity, they provide a way to manage their all free storage accounts using GrandStore application.

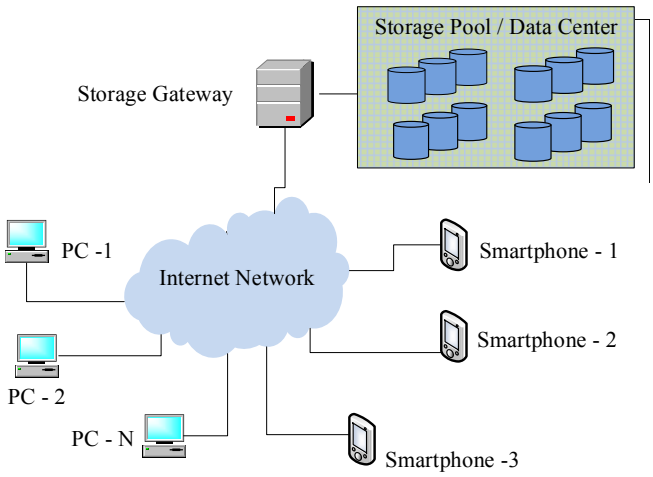


Fig. 1: Exsting Cloud Storage concept

TABLE I. DIFFERENCE BETWEEN FILE, BLOCK AND OBJECT STORAGE SYSTEM

Attribute	Storage System		
	File	***Block***	***Object***
Use Cases	File Sharing, Local Archiving, Data protection	Databases, Email Servers, RAID, Virtual Machines	Big Data, Web Applications, Backup Archives
Access	Full path should be known , Directory, sub directory , folder and file name	Provides Direct Access to the block	No direct access to object. An Object storage system places a file-system bridge [Gateways] in-between the two.
Metadata	No metadata, full path should be known.	No additional information, provides direct access to block	expandable and variable amount of metadata
Ease of Use	Simple to use and setup	Block Storage provide ability to incrementally edit any part of the file	Object storage is powerful and customizable because of variable amount of metadata of any type
Client Side Visibility	Visibility of files and folders is same on client and server	supports formatting of file systems like NTFS or SMB, NFS, VMFS (VMware) as need	Expose object storage as Network File System via Storage Gateways.
Recommendation	Recommended for shared storage	Recommended for dedicated storage.	Recommended for unstructured data storage

Chetan et al [17] proposed a framework using OpenStack [18] software to deploy storage as a service private cloud model that secure data using credential and Advanced Encryption Standard (AES) algorithm [19].

Adiseshu et al [20] proposed a solution to build the personal cloud using unused computers as nodes. Virtual Machine is setup on each unused computer node to build a datacenter-less, distributed cloud.

Alginahi et al [21] explained the various methods of authenticating to guarantee the trusted flow of data, they described the various challenges of authentication methods and usage of cryptographic algorithm in Cloud and IoT.

"Fig. 1" draws the basic representation of existing cloud network, where a smartphone participates as a client like other peer nodes such as personal computers, Tablets, and Laptops. Internet Network represent the Internet Service Provider and Backbone Network. There exists one Storage Gateway Server which basically redirects the user's data GET and PUT request to data center, then data center retrieve and stores the requested data.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:02:08 UTC from IEEE Xplore. Restrictions apply.

C. *Key Challanges with Cloud Storage*

Even though cloud is emerging very fast, there are various challenges related to security, ownership of data, capability, efficiency and cost. Table II describes these major challenges in using Cloud Storage. Our proposed solution provides solution to some of these challenges by empowering user to control their own local cloud and sharing.

TABLE II. CLOUD STROAGE CHALLANGES

Key Challanges	Brief Description
Compliance and Security	• Threats of data leakage • Unique Crendential • Security Threats during Data upload and download • No global policy to protect users cloud storage rights.
Ownership of data	• Only Cloud Storage service provider has full control on data. • No idea about how many copies of data are created by service provider for backup and restore.
Capability	• Rehousing of data is slow • Complex process to change service provider.
Efficency	• Downtime of service provider • Dependency on Network Speed.
Cost	• No pure private or personal cloud storage unless user has dedicated storage in his own premises which leads to high cost.

III. PROPOSED WORK

Our proposed system is called Personal Mobile Storage Sharing (PMS-Sharing). This systemremoves dependencies on cloud storage service providers to store the data on servers in order to send/share files/documents to any other user. In our proposed solution data is stored on the devices belonging to the known members of group. In our proposed way user's data are safe and reachable physically in case of network failure/unavailability.

In this era of wireless data transmission, Wi-Fi data throughput will increased far better with the deployment of 5G and 6G Wi-Fi routers. Wi-Fi Alliance [22] also introduced WiFi-6, next generation of highly efficient Wi-Fi network with theoretical speed of 11 Giga bits per second. These make proposed solution feasible to deploy on private as well as on personal local network. Unlike traditional cloud storage "Fig. 1", in proposed solution smartphones can participate in both ways as peer as well as hosting device as storage provider. "Fig. 2" draws the high level concept of the proposed framework.

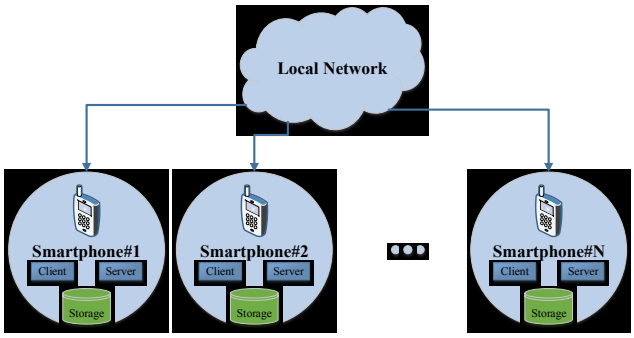


Fig. 2: PMS-Sharing Concept

Objective of PMS-Sharing is to provide a shared storage. PMS-Sharing achieves its objective by operating as both client and server mode. In case of unused smartphones the full storage can be utilized by PMS-Sharing and in case of smartphones devices which are in use, user can limit the size of storage to use as shared storage.

A. *Architecture Overview*

"Fig. 3" represents the high level architecture diagram of the PMS Sharing application. PMS Architecture has following main modules

1) *Login Activity*

Login Activity control the user interaction with Login Screen, Sign up Screen and Reset Password Screen. Login Activity is responsible for authenticating user to enter into the PMS Sharing application and use it. Login Activity is dependent on Account Manager.

2) *Dashboard Activity*

Dashboard Activity control the user interaction with Dashboard Screen, Create Group Screen, and Edit Group Screen. Dashboard Activity is mainly dependent on Group Manager.

3) *File Explorer Activity*

File Explorer Activity controls the user interaction with File Explorer Screens (Remote and Local Directory View). File Explorer is responsible for providing rich experience to end user while using, sharing and accessing the files which are made available by the File Repository

4) *Account Manager*

Account Manager provides services to the Login Activity to authenticate the user and granting Login Activity and the necessary permissions to enter into the PMS Sharing

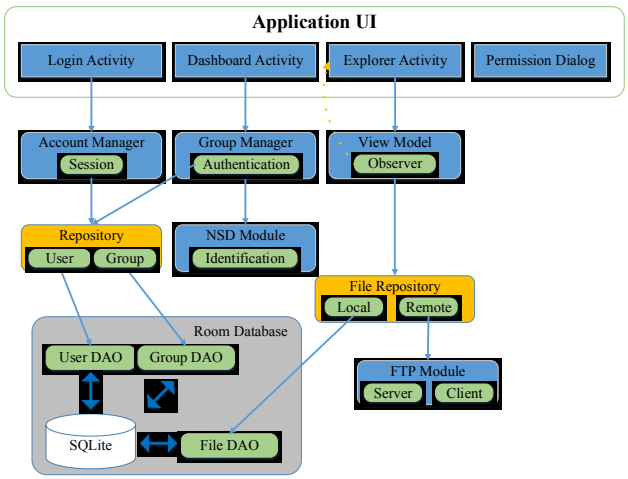


Fig. 3: Architecture PMS-Sharing Application

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:02:08 UTC from IEEE Xplore. Restrictions apply.

Application. Account Manager also manages the user session. Once the user is authenticated, he is granted the access for 12 hours or until the user logout manually.

5) Group Manager

Group Manager is at the core of PMS Sharing. Group Manager provides all the services that enable all the function supported by the Dashboard Activity. Group Manager controls function of Discovering Client, Creating Group, Editing the Group and showing group details.

6) NSD Module

Android provides the API to use Network Service Discovery (NSD). Using NSD PMS-Sharing application can easily identify and detect the users which are using PMS-Sharing services on the local Network. NSD Module help Group Manager to populate all the users available on the same Network and using PMS-Sharing services. User device will start broadcasting PMS-Sharing service on local network after registering the PMS-Sharing services, User device can also discover similar services broadcasted by other user on the same network. On discovering any service, NSD resolves the discovered services to get IP and Port of the peer discovered device. Then NSD can stop discovery any time by un-registering services.

7) File Repository

File repository provides a protection wall between actual storage media and the shared folder thus the remote user is capable to access the files and folder added into the shared storage directory path. File repository also keeps track of file added by any remote user.

id(primary_key)	user_id(string)	group_id(long)	file_name(String)	file_path(String)

Here user_id is the id of user who added the file/directory in the group folder, group_id is the id of the group in which this file is added, file_name is the name of file or directory. file_path is the complete local path of the directory in which file is kept. In case it is directory, it is the complete local path of the directory in which this directory is kept. File owner information helps to authenticate and validate the actual source of the user and in case of shared file it may store multiple user's information to provide access to multiple users.

B. Communication Interface

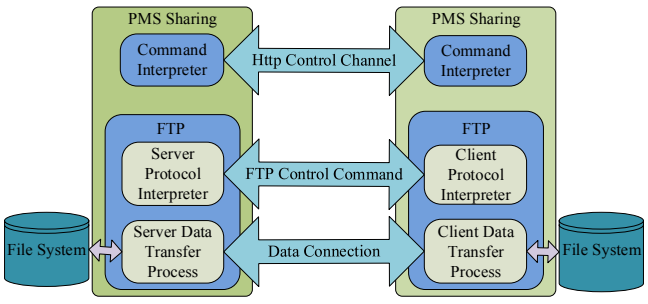


Fig. 4: High level representation of Communication Interface

PMS Sharing application uses FTP for Data Transfer Channel and HTTP for Control channel. "Fig. 4" depicts the communication interface between two PMS-Sharing applications, installed on two different devices belonging to different user. Command Interpreter is used to send and receive PMS-Sharing specific commands and control messages.

C. User, Group and File Repository

User repository is developed to manage the group management and user management. It also incorporated three entity tables:-

1) Entity Name - "NetworkGroupDetails"

In this entity each record will store detail about group id, group name, and total number of users in the group and user id of master user.

id(primary_key)	group_id(long)	group_name(String)	total_user(int)	master_user(string)

2) Entity Name - "NetworkGroupUserDetails"

In this entity each record will store information about each user of the group and its associated pair key to authenticate the user to access the group.

id(primary_key)	group_id(long)	user_id(string)	ip(string)	port(int)	pair_key(long)

3) Entity Name - "User"

In this entity each record will store information about the application's user.

id(primary_key)	user_name(string)	email(string)	password(string)	sec_q(string)	sec_a(string)

Where sec_q is security question and sec_a is security answer. Password, Security Question and Security Answer are encrypted before storing in database. It has used apaches salted password encryption.

D. Authentication Protocol

Authentication method is composed of multiple steps as Device Registration for service discovery, Identification of devices registered for same service. Complete process is discussed in below section.

1) Device Registration

Each device registering PMS-Service, will be visible to all other devices registering same PMS-Service. When any device gets connected to the network, then all other user will automatically detect the device using NSD service discovery. Default service name for registering in NSD Module is "pms-service".

Private static final String DEFAULT_SERVICE_NAME = "pms-service";

2) *Device Identification*

Using "*pms-service*" service name tag, that will be able to identify all the device in the network which are registered with same service name in NSD Module. Now next challenge is to identify the role of the device like which device is operating in Group Owner mode and which device is just participating as a

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:02:08 UTC from IEEE Xplore. Restrictions apply.

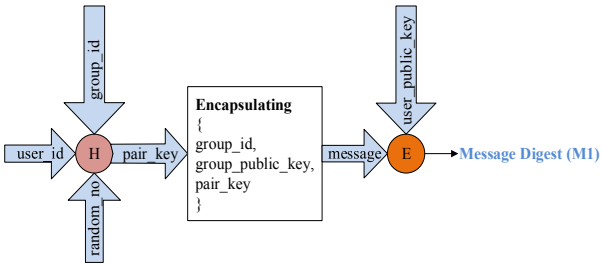


Fig. 5: Generation of Message Digest (M1)

user. Further, it extended the service registered to identify server and client. The device registering with server role is Group Owner and all other devices are normal users as shown in below code snippet.

```
Public String mServiceName = DEFAULT_SERVICE_NAME;

Public static enum TYPE {
   CLIENT,
   SERVER
}

Public void registerService(..., TYPE type) {
   ...
  If (type == TYPE.CLIENT)
      mServiceName = mServiceName + ".client";
  else if (type == TYPE.SERVER)
      mServiceName = mServiceName + ".server";
  ...
}
```

3) *User Identification and Authentication*

The next part is finding the ways to identify users who belong to a particular group and if there are multiple Group Owner in the network then how to identify the Group Owner in which particular user is added because user's IP and Port can change after each connect and disconnect from the Network. It consists of following steps:-

- At the time of group creation when Group Owner selects the users and click on create group, then Group Owner device starts user registration where he first connect with user 'X' device and notify about new group. On receiving notification for new group, User 'X' device will send

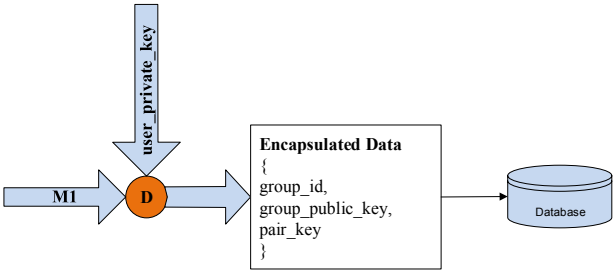


Fig. 6: Decryption of Message Digest (M1)

request to group owner for getting pair key with its public key and user id. Generation of pair_key and message digest is shown on "Fig. 5". Group Owner device will generate a random number, and pass this user_id, group_id and random number to Hash function (H) to generate a 32 bytes secured pair-key. Now it will encapsulate group_id, group_public_key and pair_key into single message and encrypt the message using user's public key to generate message digest (M1) and send to user 'X'.

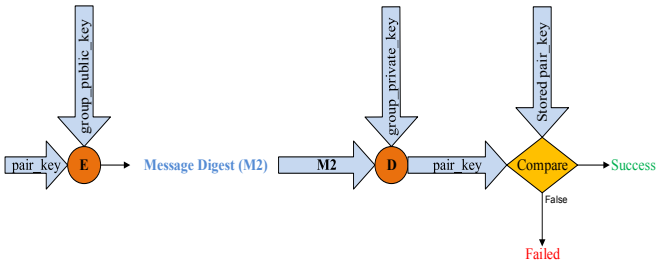


Fig. 7: User Authentication

- As shown in "Fig. 6", on receiving message digest (M1), User 'X' will decrypt the message digest using user 'X' private key to get encapsulated message {group id, group public key, pair key} and store this information in database.
- Now as shown in "Fig. 7", whenever user 'X' wants to access Group. User 'X' device will encrypt this pair_key using previously shared group's public key to generate message digest (M2) and send this to Group Owner.
- Group Owner decrypt this message (M2) using group's private key to get the pair_key and compare this pair key with the stored pair_key, if both are same means user 'X' is authenticated successfully.

IV. Result Analysis

PMS-Sharing application provides following major function:-

- User can create login account with valid email id.
- User can login to the application with valid id.
- User can create group of peer user whom with he wants to share the files.
- Dashboard should list out all the Groups created by user and Groups in which user is added.
- User can modify the group details (Name, and participant user).
- User can select the group to enter in shared directory of that group.
- User can add, remove and view the files in shared directory.
- User who creates the Group will be Group Owner by default for that particular Group.
- The Group Owner can turn ON and turn OFF sharing at any time from Quick Menu Button.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:02:08 UTC from IEEE Xplore. Restrictions apply.

- User can logout from the application any time from any screen

Let's take an example to compare the PMS-Sharing with other Sharing. Suppose there is a family with five members named as Lajjesh, Nitesh, Gunjan, Bhumika and Urav. Suppose Lajjesh wants to share some photo graphs with all other family members. Lajjesh will perform following steps to share files normally:-

- Lajjesh will choose files.
- Lajjesh will share via Wi-Fi Direct/Bluetooth and select user Nitesh, Device pairing request will be sent to Nitesh's device.
- Nitesh will accept device pairing request and both devices are paired.
- After pairing to Lajjesh's Device start transferring files to Nitesh's Device.
- Repeat above steps to share file with Gunjan, Bhumika and Urav also.

So it has been noted that in total 16 steps (4 Members * 4 Steps each) are needed to choose files, need to do manual device pairing, which incorporates active user involvement and file sending can fail also if devices don't have enough space. While in PMS-Sharing Lajjesh will perform following steps:-

- Lajjesh will login in PMS-Sharing application
- Select Family Group
- Add files which he wants to share with Nitesh, Gunjan, Bhumika and Urav.

So in PMS-Sharing only 3 steps in total are needed, whole process of transfer and authentication is automatic once the initial group is created. PMS-Sharing helps in saving overall time and user involvement for sharing. In PMS-Sharing only sender need to add files in shared folder and other user can sync/download the file whenever they want to download unlikely to the normal file transfer where file is transferred immediately even though another user don't want to do so. Similarly if it is compared with Cloud sharing where first user need internet access to use services, user can share uploaded contents URL to any other member, there is no control on data once uploaded to Cloud Storage, All the information is already exposed to the cloud service provider, any person who has account on cloud storage application and has the valid URL of shared files then he can easily access the files. While in case of PMS-Sharing user can use the services on local network without internet and the content is accessible only to the members of the group.

V. CONCLUSION AND FUTURE SCOPE

PMS-Sharing authentication model adds extra security in term of authentic access to the data shared by any other user. Unlikely to the cloud storage, in PMS Sharing user has full control on data. Based on PMS-Sharing concept, personalized cloud storage can be built with marginal cost and more research can be done to mitigate data recovery and security issues. PMS-Sharing is very useful to deploy Private Cloud for a group of know people say a Community and for a Military Services. During the development, various challenges and limitations are observed on which the future work can be done. First, Smartphones which are operating in server mode need more processing and battery power to serve the request from n-number of users. There is threat of losing connectivity with PMS-Sharing when device power down. Second, there is need to have secure and resource effecient method for Data recovery. Third, PMS-Sharing is capable of handling n number of users and groups on local networks but access points allows only limited number of users to connect at a time.

REFERENCES

[1] Yuanyuan Guo, Jianjun Hao, Yijun Guo, Tao Luo, " Research on Data Block Storage Strategy in Cloud Storage System", Proceedings of 2017 3rd IEEE International Conference on Computer and Communications, 2017, pp. 2394-2397.

[2] https://en.wikipedia.org/wiki/File_system, last accessed on: 15/09/2019.

[3] Yuanyuan Guo, Jianjun Hao, Yijun Guo, Tao Luo, " Research on Data Block Storage Strategy in Cloud Storage System", Proceedings of 2017 3rd IEEE International Conference on Computer and Communications, 2017, pp. 2394-2397.

[4] https://www.ibm.com/cloud/learn/block-storage, last accessed on: 15/09/2019.

[5] https://aws.amazon.com/ebs, last accessed on: 16-09-2019.

[6] https://www.ibm.com/cloud/learn/object-storage, last accessed on: 15/09/2019.

[7] https://www.idrive.com/, last accessed on : 21-09-2019

[8] https://www.pcloud.com/, last accessed on : 21-09-2019

[9] https://mega.nz/, last accessed on : 21-09-2019

[10] https://onedrive.live.com/about/en-gb/, last accessed on : 21-09-2019

[11] https://www.icloud.com/, last accessed on : 21-09-2019

[12] https://cloud.google.com/, last accessed on : 21-09-2019

[13] https://www.box.com/, last accessed on : 21-09-2019

[14] https://nextcloud.com/, last accessed on : 21-09-2019

[15] https://spideroak.com/, last accessed on : 21-09-2019

[16] Li Zhang, Bing Tang, "GrandStore: Towards Large-Scale Free Personal Cloud Storage", Proceedings of 2017 International Conference on Computer Network,2017, pp 118-123

[17] Chetan Gaikwad, Bhoomika Churi, Kanad Patil, Tatwadarshi P. N., "Providing Storage as as Service on Cloud using OpenStack", Proceedings of 2017 International Conference on Innovations in information Embedded and Communication Systems, 2017, pp 1-4.

[18] https://www.openstack.org/, last accessed on : 24-09-2019.

[19] B. Varsha and P. Suryateja, "Using Advanced Encryption Standard for Secure and Scalable Sharing of Personal Health Records in Cloud", International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 5 (6) , 2014.

[20] Adiseshu Hari, Ramesh Viswanathan, T. V. Lakshman, Y. J. Chang, "The personal cloud: design, architecture and matchmaking algorithms for resource management", Proceedings of the 2nd USENIX conference on Hot Topics in Management of Internet, Cloud, and Enterprise Networks and Services, 2012, pp 1-6.

[21] Alginahi, Yasser M., and Muhammad Nomani Kabir, eds. Authentication Technologies for Cloud Computing, IoT and Big Data. IET, 2019, chapter 3-4.

[22] https://www.wi-fi.org, last accessd on : 28-09-2019

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:02:08 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Predicting Link Sign in Online Social Networks based on Social Psychology Theory and PageRank

Sanjay Kumar
Department of Computer Science and Engineering
Delhi Technological University
New Delhi, India.
sanjay.kumar@dtu.ac.in

Sudhanshu Shekhar Singh
Department of Computer Science and Engineering
Delhi Technological University
New Delhi, India.
singhsudhanshu204@gmail.com

Vipul Gupta
Department of Computer Science and Engineering
Delhi Technological University
New Delhi, India.
gvipul294@gmail.com

***Abstract*— Recently, the social networks are evolved as the primary platform for internet users to share their ideas and views. Social media interactions can be modelled as links between two users, where the link weight is different for different types of relationships and interactions. Social networks are dynamic in nature where the links are formed and deformed with time. Broadly, two types of connections, i.e. Positive and Negative can exist in most social networks. Positive connections signify trust and friendship, whereas negative connections show distrust and enmity. Reliable prediction of edge signs can greatly influence in enhancing user experience. Prediction of edge signs has been explored previously also. In this paper, the primary aim is to predict the sign of edges based on extracted features of nodes constructed upon theories of social psychology that includes classical balance theory, the status theory and emotional information theory. The ranking of nodes is used to extract features of nodes that quantify the popularity and trust of nodes on the network. These features are combined with the features extracted from socio psychological theories and try to improve the performance of our sign prediction model. Our results show that the proposed methodology achieves a reasonable performance when implemented on two real-life datasets namely Slashdot and Epinions.**

Keywords—Emotional information theory; Online Social Networks; Signed Network; Social Psychology; Node Ranking;

I. INTRODUCTION

Social networks have become a part of our daily life. They have gained enormous popularity in the past few years and play an important role in various aspects of people's lives. Social networks are not only the prime medium of communication between users, but also the primary source of information for most of the people. Social network analysis is not a formal theory but rather an investigation strategy for exploring social structures [1]. These days, product advertisement has become a huge industry. Effective advertising strategies include promoting the product to targeted audiences on online social networks. The recommendation systems are also playing an important role in deciding what a user buys on a website and also improving user experience by suggesting the relevant products and services. There are various other applications of social networks that rely on reliable and robust sign prediction and link prediction algorithms. A lot of effort in the research field of social networks is directed towards understanding the relationships between social entities [2].

Numerous applications of social network analysis are available in various fields such as social content filtering, recommendation systems, behavior analysis, network modeling, etc. Social network analysis has even demonstrated that it is effective in online collaborative learning also [3]. Some of the popular research fields in social network analysis include community detection, partitioning of social networks based on sign, ranking of nodes in social networks, sign prediction, etc.

Link prediction is one of the major research interests in the field of social network analysis. Link prediction has been at the core of many applications like bioinformatics, e-commerce and information retrieval. Link prediction can be used to predict the likelihood of formation of a link between the nodes in the future which currently have no link between them [4]. Sign prediction has also evolved as a growing research interest in social network analysis. Signed networks provide better understanding of social entities and their behavior than the unsigned networks. Various social networks can be represented as a large scale, enormous signed network. The weights and sign of links between users can be derived from the relationship and interaction of users on social networks [5].

In this paper, social psychology is used to analyze signed social networks. The socio-psychological theories used here are, the social status theory, social balance theory and emotional information theory. It has been further tried to use the modified version of the popular pagerank algorithm to determine the popularity and trust of a node. It is then analyzes on how the introduction of these features impacts the prediction accuracy of the model. In this paper, the features are extracted from these theories that helps to characterize the type of relationship between two nodes of a graph based on the social behavior and relationship with other nodes in the graph. Previous studies have been conducted to predict the signs of nodes in the network. So far, the sign of the network has been predicted based on Balance theory and Social status theory [7]. The pagerank algorithm has also been used many times in the field of sign prediction [20]. In this research paper, it has been tried to make use of both socio-psychological theories and the ranking of nodes to predict signs of nodes in a network. The

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:58:59 UTC from IEEE Xplore. Restrictions apply.

first class of features is based on the socio-psychological theories that include the social balance theory [8], the social status theory and the emotional information theory [9]. The second class of features has been extracted on the basis of the Troll-Trust model [20].

Our study provides insight into the nature of the relationships and also provides a large scale evaluation of socio-psychological theories [10] on real-life datasets namely Slashdot and Epinions.

The background of signed social network theories is divided into the following three subsections.

A. *Balance Theory*

According to social psychology, the signed graphs are incorporated to model the social situations where friendship can be inferred from the edges with the positive sign while the negative edges characterize enmity between the respective nodes. Figure1 (A) illustrates the balanced triads while Figure1 (B) illustrates the unbalanced triads. Balance theory suggests that the positive triads (i.e. triads with all three positive edges or two positive edges and one negative edge) are expected to be in stable social situations and thus, are balanced. On the other hand, negative triads are unbalanced and exist in rare unstable social circumstances. Structural balance can be used to express how the feelings, attitudes, and beliefs, which the social actors have towards each other, promote the formation of stable social groups [11].

B. *Status Theory*

The fundamental principle of status theory implies that the recipient of the positive directed link is supposed to have higher status than the originator node while in the case of the negative directed link, the recipient is considered to have lower status.

C. *Emotional Information Theory*

Emotional Information about the nodes can be derived from the nature of relationship that a node forms. It has also been tried to find out the optimism and pessimism of every node in the network. The node with high value optimism is more optimistic about the relationships in the network and is more likely to form links with other nodes and vice versa. High value of pessimism is attributed to skepticism about other nodes and the nodes with high value of pessimism are less inclined towards formation of relationships. It has also been experimented to exploit the information and figure out the values of optimism and pessimism for each node and use those values for prediction of sign of links in the network [9].

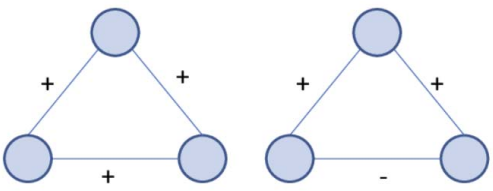


Figure 1 (A) represents a Balanced Triad

is m

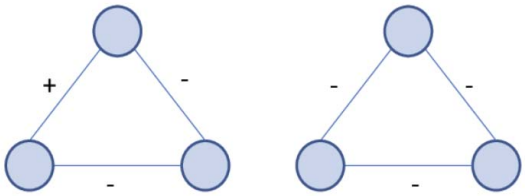


Figure 1 (B) represents an Unbalanced Triad

D. *PageRank algorithm*

The traditional PageRank algorithm is used to rank web pages on the search engine. PageRank counts the number and quality of links to a page and ranks the pages on the basis of popularity [13, 14]. The popular pages are likely to have links from various pages across the internet. The traditional pagerank calculates the probability that a person randomly clicking on links arrives at a particular page. Pagerank iterates over the graph multiple times to adjust the pagerank values of the nodes in the network.

Pagerank values are uniformly distributed in the network at the beginning of computation. Pagerank of every node is adjusted in every iteration using the following formula.

$$\boldsymbol{PR(u) = 1 - d + d \sum_{v \epsilon B^u} \frac{PR(v)}{L(v)}}$$

In the above equation PR (u) is the pagerank of the node u in consideration. B^u is the set of all the nodes that have an edge directed towards node u. The parameter d accounts for damping factor. Damping factor is the probability that the random surfer will continue surfing after the current step. L(v) is the total number of outgoing links from the node v [22].

This explanation provides intuition about the use of pagerank in unsigned networks. The nodes with more number of direct and indirect links directed towards it are given a higher rank, making them more likely to be liked by other nodes. Pagerank effectively captures various relations amongst vertices of graph [21]. Ranking of nodes in unsigned networks is a well known concept [15, 16]. However, use of pagerank for ranking of nodes in a signed network has been explored recently [17, 18]. The traditional pagerank is designed only for unsigned networks. However slight modifications to the existing pagerank algorithm to accommodate negative weights have shown impressive results in prediction of sign of the edges [19].

The rest of the paper is organized as follows: Section II discusses the datasets present. Section III elaborates the proposed methodology and implementation details followed by Section IV which elaborates and analyses the results obtained. Finally Section V concludes our research paper.

II. DATASET DESCRIPTION

In this work, two real-life data sets are considered, they are Slashdot and Epinions, which is publicly available at www.snap.stanford.edu/data. Table 1 summarizes the data sets.

The description of datasets is as follows:

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:58:59 UTC from IEEE Xplore. Restrictions apply.

A. Slashdot

Slashdot is a technology related news website. This website consists of current technology related news. Slashdot allows users to tag each other as friends or enemies.

B. Epinions

Epinions is a general product review website where users are allowed to rate the product. The value of rating indicates how much the product is liked by the user. Higher ratings indicate positive feedback whereas lower ratings indicate negative feedback.

Table 1. Dataset Description

	Epinions	**Slashdot**
Nodes	119,217	82,144
edges	841,200	549,202
+edges	85.0%	77.4%
-edges	15.0%	22.6%

III. PROPOSED METHODOLOGY

The method has a connected graph, which have nodes with both positive and negative links. Initially, it starts to extract the features that model the local relation of the node with the rest of the nodes of the network. It has two types of features i.e. features based on socio-psychological theories and features based on ranking of nodes.

The work has been conducted in following steps:

1. First of all, the features are extracted based on socio-psychological theory. These features can be further classified into three sets of features. These three sets of features are derived from three different theories.

   The features based on socio-psychological theories are enlisted below:

   a. These features include $D^{+}_{in}(v)$, $D^{-}_{in}(v)$, $D^{+}_{out}(u)$ and $D^{-}_{out}(u)$. $D^{+}_{in}(v)$ and $D^{-}_{in}(v)$ represent inward positive and inward negative edges of v while $D^{+}_{out}(u)$ and $D^{-}_{out}(u)$ represent outward positive and outward negative edges on U. And the feature C(u,v) is the total number of common neighbors of u and v, i.e. total number of nodes w that have edges connecting to both u and v, no matter whatever the type of edge is. Other two features include total outdegree of u i.e $D^{-}_{out}(u) + D^{+}_{out}$ as well as total in degree of $D^{+}_{in}(v) + D^{-}_{in}(v)$ [8].

   b. For this feature set, every triad (u, v, w) is considered, such that w has an edge either to or from u and also an edge either to or from v. Since the edges can be of any sign and any direction, there are 2*2*2*2=16 possibilities. Hence, there are sixteen types of triads possible where each triad can provide evidence of different signs [8]. Our sixteen-second class features include the number of occurrences of the sixteen types of features.

   c. For the third class of features, the features are derived by the emotional intelligence theory. Emotional intelligence theory shows the correlation between emotional information and positive/negative links. The measure of pessimism and optimism has been performed for all the nodes.

      A set of OL_i for each node u_i is considered, such that the elements of set OL are the nodes which have positive emotions from nodes u_i but have higher negative emotions received than the average negative emotions received by nodes in the graph [2].

      $$OL_i = \{u_j \mid P_{ij} != 0 \;\&\&\; N_j > N_{avg}\}$$

      It further defines OH such that OH is a subset of OL such that elements present in OH receive lower positive emotions than the average of the entire graph.

      $$OH_i = \{u_k \mid u_k \in OL_i \wedge P_k < P_{avg}\}$$

      The optimism is measured as the ratio of the length of OH to OL.

      $$O_i = \frac{|OH_i|}{|OL_i|}$$

      The high optimism feature shows that the node with high optimism has positive emotions towards the nodes with low positive and high negative emotions across the graph and hence is more likely to form positive relations with other nodes too.

      Pessimism is calculated and defined as a set PH_i for all the edges u_i such that PH_i contains all the nodes that have negative emotions from the node u_i but have higher value of positive emotions from other nodes in the graph than the average positive emotions.

      $$PH_i = \{u_j \mid N_{ij} != 0 \;\&\; P_j > P_{avg}\}$$

      It further defines PL_i for u_i such that PL_i is a subset of PH_i and contains elements of PL_i which

      Have more negative emotions than the average of the entire graph.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:58:59 UTC from IEEE Xplore. Restrictions apply.

$$PL_i = \{u_k \mid u_k \in PH_i \ \& \ N_k < N_{avg}\}$$

Pessimism is given by:

$$P_i = \frac{|PL_i|}{|PH_i|}$$

A high value of pessimism shows that the given node is more likely to form a negative relationship with undiscovered nodes because it already has negative relations with the nodes having positive emotions from other nodes across the graph. The optimism and pessimism of both nodes are considered for participating in the edge as the features for the learning algorithm.

The above three classes of features constitute the features based on socio-psychological theory. In the above feature set, which made use of one or the other socio-psychological theory to extract the features.

2. After this, the features are extracted based on ranking of nodes. Later, the graph is divided into two sub graphs, i.e. A^+ and A^-. A^+ is the positive sub graph containing all positive edges while A^- is a negative sub graph containing the negative edges only.

   The pagerank algorithm is used to calculate ranking scores ($\square_i$) of each node in both the subgraphs. It adopt two measures, reputation, and optimism, as features of nodes for classification purposes. The feature reputation quantifies the popularity of the node in the network, whereas Trust quantifies the amount of trust, the node has in other nodes of the network and the pattern in which the node makes votes.

$$\text{Reputation}_i = \frac{\Sigma\square_{j:(j,i)\epsilon A^+} - \Sigma\square_{j:(j,i)\epsilon A^-}}{\Sigma\square_{j:(j,i)\epsilon A^+} + \Sigma\square_{j:(j,i)\epsilon A^-}}$$

$$\text{Trust}_i = \frac{\Sigma\square_{j:(i,j)\epsilon A^+} - \Sigma\square_{j:(i,j)\epsilon A^-}}{\Sigma\square_{j:(i,j)\epsilon A^+} + \Sigma\square_{j:(i,j)\epsilon A^-}}$$

R_i and T_i represent the reputation and trust of the node (i).
The more popular the node is, the more edges will be directed towards it. Hence it will have a higher reputation. Whereas higher optimism of a node shows that it has many outward edges directed towards other nodes from itself. It use trust and reputation of both nodes in the edge as edge features [20].

It merges both feature sets defined above and train the machine learning model on a combined large dataset.

It chooses as a logistic regression classifier to predict sign prediction as it experiences a good performance of this classifier in social network analysis [12]. This has trained the model on the features extracted from the graphical data. This has extracted three classes of features (i.e. first, second and third class of features) from the data to train the model on each one of them separately.

IV. RESULT

In the following Table 2, values signify the minimum value of embeddedness allowed in the graph and the prediction accuracy of model corresponding to each minimum threshold embeddedness level (e.g., Embeddedness 25 means that only those links with embeddedness greater than or equal to 25 have been considered for prediction)

Table 2: Prediction accuracy of the model when the features based on social psychological theories were used for training

	Embedded-ness(0)	**Embedded-ness(10)**	**Embedded-ness(25)**
Slashdot	87.2	88.4	89
Epinions	93.3	95.9	96.6

The following Figure 2(A) shows the prediction accuracy of the model on Epinions dataset when trained on the features based on social psychological theories whereas the Figure 2(B) shows the prediction accuracy of the model on Slashdot dataset when trained on the features based on social psychological theories.

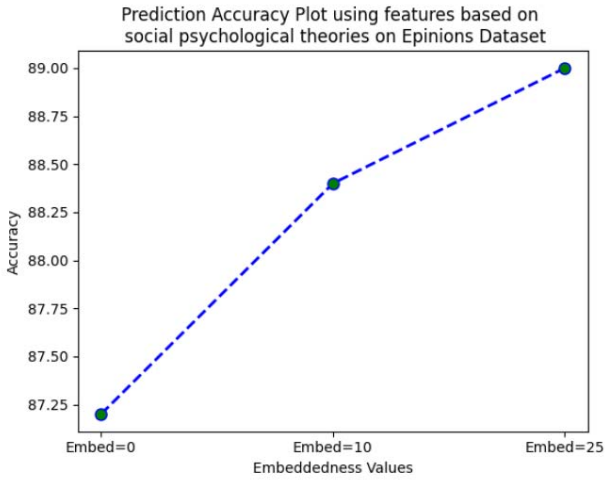


Figure 2(A)Prediction accuracy for Epinions Dataset when the model is trained on the features based on social psychological theories.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:58:59 UTC from IEEE Xplore. Restrictions apply.

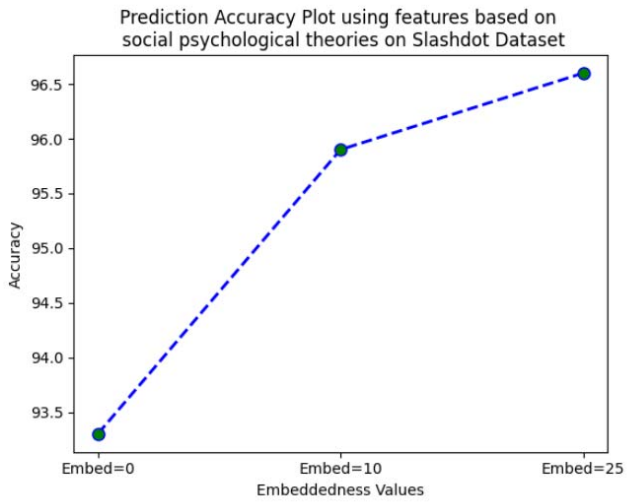


Figure 2(B)Prediction accuracy for Slashdot Dataset when the model is trained on the features based on social psychological theories.

The following table 3 and Figure 3 depict the accuracy values of the model based on the node ranking algorithms on both Epinions and Slashdot datasets.

Table 3: Prediction accuracy of the model when the features based on node ranking algorithms were used for training

	Accuracy
Slashdot	85.7
Epinions	94.4

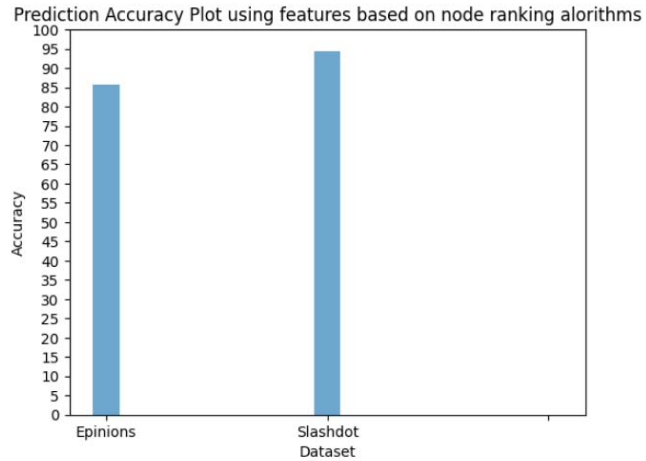


Figure 3 Prediction accuracy for Epinions and Slashdot Dataset when the model is trained on the features based on the node ranking algorithms.

The proposed methods has further used logistic regression as well as artificial neural networks on the feature set obtained by integrating the social psychological theories feature set and the features based on node ranking algorithms on both Epinions and Slashdot datasets. The prediction accuracy of logistic regression on the new integrated feature set is found to be higher than the prediction accuracy of both the first and second set of features when considered separately.

The following Table 4 and Figure 5 show the accuracy comparison of the model before and after the introduction of features based on node ranking algorithms on Epinions dataset.

Table 4: Accuracy Comparison Table of the model before and after the introduction of features based on node ranking algorithms on Epinions dataset.

Algorithm Architecture	**Improved Accuracy**	**Accuracy Earlier**
Embeddedness-0 Logistic Regression	95.5%	93.3%
Embeddedness-0 Neural Network	92.5%	92.4%
Embeddedness-10 Logistic Regression	97.1 %	95.9%
Embeddedness-10 Neural Network	92.5 %	92.5 %
Embeddedness-25 Logistic Regression	97.7 %	96.6%
Embeddedness-25 Neural Network	92.6 %	92.6 %

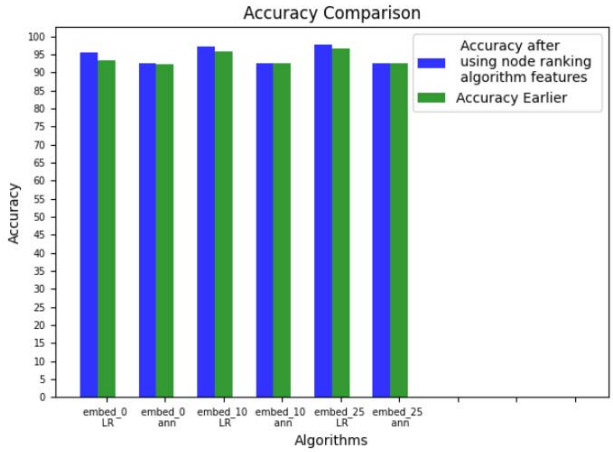


Figure 5: Accuracy comparison depiction for Table 4.

The following Table 5 and Figure 6 shows the accuracy comparison of the model before and after the introduction of features based on node ranking algorithms on Slashdot dataset.

Table 5: Accuracy Comparison Table of the model before and after the introduction of features based on node ranking algorithms on Slashdot dataset.

Algorithm Architecture	**Improved Accuracy**	**Accuracy Earlier**
Embeddedness-0 Logistic Regression	88.5%	87.2%
Embeddedness-0 Neural Network	92.5%	92.3%

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:58:59 UTC from IEEE Xplore. Restrictions apply.

Embeddedness-10 Logistic Regression	89.4%	88.4%
Embeddedness-10 Neural Network	91.5 %	91.6 %
Embeddedness-25 Logistic Regression	89.9 %	89%
Embeddedness-25 Neural Network	91.5 %	91.7 %

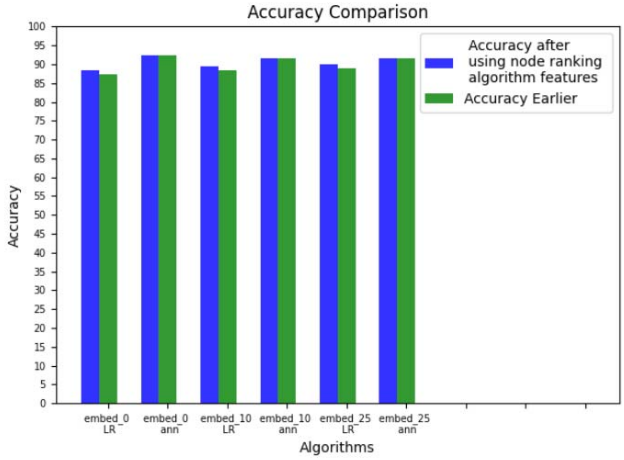


Figure 6: Accuracy comparison depiction for Table 5.

These have analyzed the results and draw insights from the results shown. The analysis of the result is divided into the following two subsections:

A. *Improvement of Accuracy with embeddedness:*

Embeddedness of a particular edge represents the number of common neighbors of the nodes participating in the link. The threshold embeddedness values are used for analyzing the result. From Figure 2(A), Figure 2(B) and Table 2, illustrates that there is a significant improvement in accuracy as the minimum embeddedness threshold is increased for a graph. As it increase the threshold embeddedness of links in the proposed network, where it only select nodes that have more information about their common neighbors. Our result shows that the accuracy of prediction increases with increase in the threshold embeddedness of the graph. As it increases the minimum embeddedness threshold of the graph, it selects only those nodes that have more behavioral information present in the network, which is encoded in the relationships between the nodes of the edge under consideration and their neighbors. This helps our model to make more informed predictions than that on the original graph. The trend of increasing predictive accuracy holds good even with the introduction of the set of features based on ranking of nodes.

B. *Improved accuracy:*

The observations in Table 3 depict the predictive accuracy of training the model on features based on node ranking algorithms. Table 4 and Table 5 show the comparison of prediction accuracies of models before and after introduction of node ranking based features. The prediction accuracy has improved with the introduction of features based on node ranking algorithms. This observation signifies that combining the first and second set of features for prediction increases the accuracy and makes the model more efficient.

The predictive accuracy of the model when trained only in the third class of features using Logistic Regression algorithm is given in Table 2. The features based on node ranking method are integrated with the features based on social psychological theories, and the model is trained on these features. The "Improved Accuracy" column of Table 4 and Table 5 shows the predictive accuracy of the model when trained on the big feature set which contains both the feature sets. There is a significant increase in the predictive accuracy of the model.

V. CONCLUSION

From the above analysis, it has been concluded that the accuracy of the learning model improves with the introduction of features based on node ranking. It is also observed that the accuracy increases by increasing embeddedness of the graph. The above analysis shows that the features based on ranking algorithms are very relevant to signed social networks. Training model on both, features based node ranking and features based on social theories improve the accuracy of prediction of the model. It is also concluded that the reputation of node is a measure of popularity of the node in a signed network. A node with high positive reputation value is likely to receive more positive edges towards it, whereas a node with high negative value of reputation is likely to receive more negative edges towards it. Therefore, nodes are more likely to form a positive relationship with a node with higher positive reputation, whereas other nodes in a network are more likely to dislike a node with high value of negative reputation. Similarly, the trust value of nodes depicts how optimistic a node is about other nodes in the network. A node with high value of positive trust is more likely to form positive relationships with other nodes of the network, whereas a node with high value of negative trust is more likely to form negative links with other nodes in its neighborhood. Introduction of the features, trust and reputation, provides vital information about the behavior of nodes in the network. Adding these features to our feature set enhances the performance of our model. The result also shows that the improvement in prediction accuracy of logistic regression model after introduction of features based on node ranking is more significant than the improvement in artificial neural network model.

The results show that the accuracy of prediction model has reasonably increased after augmenting the features trust and reputation that have been derived from the pageranks of the nodes.

REFERENCES

[1] E. Otte, and R. Rousseau. "Social network analysis: a powerful strategy, also for the information sciences." Journal of information Science 28.6 (2002): 441-453.
[2] S. Wasserman, and K. Faust. Social network analysis: Methods and applications. Vol. 8. Cambridge university press, (1994).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:58:59 UTC from IEEE Xplore. Restrictions apply.

[3] M. Saqr, U. Fors, and J. Nouri. "Using social network analysis to understand online Problem-Based Learning and predict performance." PloS one 13.9 (2018): e0203590.
[4] M. Al Hasan, and M. J. Zaki. "A survey of link prediction in social networks." In Social network data analytics, pp. 243-275. Springer, Boston, MA, 2011.
[5] S. Maniu,B. Cautis, T Abdessalem. " Building a signed network from interactions in Wikipedia." In Databases and Social Networks ,pp. 19-24. ACM, 2011.
[6] M Stojmenovic., G Lindgaard ."Benefits and Limitations of the Social Network Analysis when explaining instances of ineffective communication in two Chemical, Biological, Radiological, Nuclear, and Explosives simulations." In Cloud and Green Computing (CGC), 2013. Third International Conference on (pp. 327-334). IEEE, 2013.
[7] J. Leskovec, D. Huttenlocher, and J. Kleinberg. "Signed networks in social media." Proceedings of the SIGCHI conference on human factors in computing systems. ACM, 2010.
[8] J. Leskovec, D. Huttenlocher, and J. Kleinberg. "Predicting positive and negative links in online social networks." Proceedings of the 19th international conference on World wide web. ACM, 2010.
[9] B. Ghazaleh, J. Tang, and H. Liu. "Signed Link Analysis in Social Media Networks." ICWSM. 2016.
[10] A. Clifton , and G. D. Webster. "An introduction to social network analysis for personality and social psychologists." Social Psychological and Personality Science 8.4 (2017): 442-453.
[11] N. P. Hummon, and P. Doreian. "Some dynamics of social balance processes: bringing Heider back into balance theory." Social Networks 25.1 (2003): 17-49.
[12] J. L.Matthews, P. DeWan, and E. Y. Rula. "Methods for inferring health-related social networks among coworkers from online communication patterns." PloS one 8.2 (2013): e55234.
[13]Chen P, Xie H, Maslov S, Redner S. "Finding scientific gems with Google's PageRank algorithm." Journal of Informetrics. 2007 Jan 1;1(1):8-15.
[14]Franceschet M. "PageRank: Standing on the shoulders of giants." Communications of the ACM. 2011 Jun 1;54(6):92-101.
[15]Chiang KY, Hsieh CJ, Natarajan N, Dhillon IS, Tewari A. "Prediction and clustering in signed networks: a local to global perspective." The Journal of Machine Learning Research. 2014 Jan 1;15(1):1177-213.
[16]Kleinberg JM. Authoritative sources in a hyperlinked environment. Journal of the ACM (JACM). 1999 Sep 1;46(5):604-32.
[17]Kamvar SD, Schlosser MT, Garcia-Molina H. "The eigentrust algorithm for reputation management in p2p networks." InProceedings of the 12th international conference on World Wide Web 2003 May 20 (pp. 640-651).
[18]Kerchove CD, Dooren PV. "The pagetrust algorithm: How to rank web pages when negative links are allowed?." InProceedings of the 2008 SIAM International Conference on Data Mining 2008 Apr 24 (pp. 346-352). Society for Industrial and Applied Mathematics.
[19]Tang J, Chang S, Aggarwal C, Liu H. "Negative link prediction in social media." InProceedings of the eighth ACM international conference on web search and data mining 2015 Feb 2 (pp. 87-96).
[20]Wu Z, Aggarwal CC, Sun J. "The troll-trust model for ranking in signed networks.". InProceedings of the Ninth ACM international conference on Web Search and Data Mining 2016 Feb 8 (pp. 447-456).
[21] Chung F, Zhao W. "PageRank and random walks on graphs." InFete of combinatorics and computer science 2010 (pp. 43-62). Springer, Berlin, Heidelberg.
[22] Brin S, Page L. "The anatomy of a large-scale hypertextual web search engine."

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:58:59 UTC from IEEE Xplore. Restrictions apply.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/341238308

Real Time Experimental control and Design of FOPI λ and L–PID Controllers tuned by Invasiveweed optimization for Trajectory Control of TRAS

Conference Paper · February 2020

DOI: 10.1109/PARC49193.2020.236577

CITATIONS
0

READS
13

2 authors:



Hem Prabha
Delhi Technological University
2 PUBLICATIONS **0** CITATIONS
SEE PROFILE



Rajul Kumar
Delhi Technological University
6 PUBLICATIONS **0** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Fuzzy controller View project

All content following this page was uploaded by Rajul Kumar on 07 July 2020.

The user has requested enhancement of the downloaded file.

2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC)
GLA University, Mathura, UP, India. Feb 28-29, 2020

2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC) 978-1-7281-6575-2/20/$31.00 ©2020 IEEE 10.1109/PARC49193.2020.236577

Real Time Experimental control and Design of FOPI^{λ} and L-PID Controllers tuned by Invasive-weed optimization for Trajectory Control of TRAS

Hem Prabha
Dept. of Electrical Engineering
Delhi Technological University
Delhi-110042, India
prabhasingh1686@gmail.com,

Rajul Kumar
Dept. of Electrical Engineering
Delhi Technological University
Delhi-110042, India
rajul_bt2k17@dtu.ac.in

***Abstract*— This article presents formational designing and real time experimental validation of Fractional order-Proportional integral (FOPI^{λ}) and Linear-Proportional Integral Derivative (L-PID) controllers for trajectory stabilization of Twin Rotor Aerial System (TRAS) which is conceded as helicopter model having pitch and yaw actuators for hovering of TRMS in vertical and horizontal planes respectively. Both the controller parameters are tuned using robust metamorphic algorithm called Invasive Weed Optimization. The control performances of both the designed controllers are examined for trajectory tracking and time taken to stabilize after the given manual disturbances. It is experimentally validated that FOPI^{λ} shows superior and robust control behavior then L-PID for benchmark TRAS plant.**

Keywords—IWO, FOPI$^{\lambda}$, L-PID, TRAS, Stabilization Control

I. INTRODUCTION

In contemporary year's automatic control of aerial vehicles are in curious interest because of their non-linear, complex system dynamics and wide range of applications in sky vigilance, detecting forest fire, natural calamity relief assistance[1]-[4]. Twin Rotor Ariel System (TRAS) plant mimics 2-DOF helicopter model having a two 44 watt DC motor propellers for pitch and yaw angle control and for achieving desired set-point hovering in vertical and horizontal planes without significant cross coupling between the two planes. The pitch and yaw DC motors receives voltage control signals from PC via PCI-1711 card and vertical beam which is mounted over electrical base plate. These control signals decisive the needed pneumatic torque by altering the rpm of pitch and yaw rotors resulting change in pitch and yaw hovering angles[5], [6].

TRAS is recognized as most non linear and complex behavior because of three moments i.e. gravity, friction and gyroscopic acting on pitch and yaw propellers [7], [8]. The stabilization control of such complex plants cannot be fulfilled by adoption traditional controllers like L-PID [9], [10].

To realize these complex system uncertainties two Fractional order-PI$^{\lambda}$ controllers are designed and experimentally validated in real time for pitch and yaw angle stabilization control. The designed FOPI$^{\lambda}$ controller have capability to stabilize close loop dynamics from dynamic environment through optimized fractional power of integral scheme [11], [12].

For optimal utilization and tuning of designed controllers a ecologically inspired catechistic algorithm [13], [14] called invasive weed optimization is used. Invasive weed optimization (IWO)[15]-[17] is influenced from robust pioneering of invasive weeds in plant cultivation. Invasive weed Colonize very vigorously and shows very adaptive nature to environmental diversification. Therefore the robust acclimating behavior of invasive weeds provides an edge for dealing with optimization of complex plants in control engineering practices. After tuning with IWO trajectory control performance both FOPI$^{\lambda}$ and L-PID controller are validated and compared mathematically and experimentally by giving real time manual disturbances for a regular time interval, these manual disturbances provided through counterbalance resembles aerial, wind and pneumatic disturbances on helicopter.

The paper is fragmented into five sections, Introduction then section II describing TRAS plant dynamical Modelling followed by Section III comprises designing of FOPI$^{\lambda}$ and L-PID controllers both for pitch and yaw propellers. Section IV intimates the experimental comparisons and results followed by conclusion and references in Section IV and V.

II. DYNAMICAL MODELLING OF TWIN ROTOR ARIEL SYSTEM (TRAS)

A. Dynamical Momentum Equations

The experimental laboratory setup for TRAS is given in fig.1 with pitch and yaw two angular rotors.

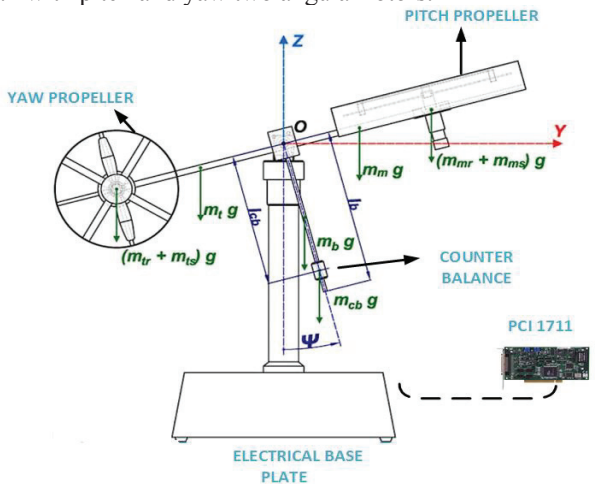


Fig. 1. TRAS Experimental system

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 18,2020 at 16:42:54 UTC from IEEE Xplore. Restrictions apply.

The mathematical statements of pitch angle (Ψ) propeller are provided in equation (1) - (5).

$$I_1\ddot{\Psi} = M_1 - M_{FG} - M_{B\Psi} - M_G \quad (1)$$

$$M_1 = a_1\tau_1^2 + b_1\tau_1 \quad (2)$$

$$M_{FG} = M_g sin\Psi \quad (3)$$

$$M_{B\Psi} = B_{1\Psi}\dot{\Psi} + B_{2\Psi}sign\dot{\Psi} \quad (4)$$

$$M_G = K_{gy}M_1\dot{\varphi}cos\Psi \quad (5)$$

The mathematical statements of yaw angle (φ) propeller are provided in equation (6) - (8).

$$I_2\ddot{\varphi} = M_2 - M_{B\varphi} - M_R \quad (6)$$

$$M_2 = a_2\tau_2^2 + b_2\tau_2 \quad (7)$$

$$M_{B\varphi} = B_{1\varphi}\dot{\Psi} + B_{2\varphi}sign\dot{\varphi} \quad (8)$$

Dynamical systemization of pitch and yaw rotors consist of gyroscopic, frictional and non-linear, gravity characteristics momentums that influence TRAS hovering non-linearities and panorama. The numerical values of system parameters including static, frictional and cross-coupling constants are given in Table.1.

TABLE I. TRAS PARAMETRIC VALUES

TRMS modeling Constants	Values
$\boldsymbol{I_1}$- M.I of Pitch Propeller	$\mathbf{7.9 \times 10^{-2} Kgm^2}$
$\boldsymbol{I_2}$- M.I of Yaw Propeller	$\mathbf{4 \times 10^{-3} Kgm^2}$
$\boldsymbol{a_1}$- Pitch Propeller Static Constant	**0.4134**
$\boldsymbol{b_1}$- Pitch Propeller Static Constant	**0.655**
$\boldsymbol{a_2}$- Yaw Propeller Static Constant	**0.67**
$\boldsymbol{b_2}$- Yaw Propeller Static Constant	**0.74**
$\boldsymbol{M_g}$- Gravity Moment	**0.399**
$\boldsymbol{B_{1\psi}}$- Frictional Moment of Pitch Propeller	$\mathbf{5 \times 10^{-0.6}}$
$\boldsymbol{B_{2\psi}}$- Frictional Moment of Pitch Propeller	$\mathbf{0.7 \times 10^{-0.6}}$
$\boldsymbol{B_{1\varphi}}$- Friction Moment of Yaw Propeller	$\mathbf{2 \times 10^{-0.5}}$
$\boldsymbol{B_{2\varphi}}$- Friction moment of Yaw Propeller	$\mathbf{3.5 \times 10^{-0.5}}$
$\boldsymbol{K_{gy}}$- Gyroscopic Moment	**0.053**
$\mathbf{k_1}$- DC Gain of Pitch Propeller	**1.45**
$\boldsymbol{k_2}$- DC Gain of yaw Propeller	**0.79**
$\boldsymbol{T_{11}}$- Moment Parameter of Pitch Propeller	**1.01**
$\boldsymbol{T_{10}}$- Moment Parameter of Pitch Propeller	**1**
$\boldsymbol{T_{21}}$- Moment Parameter of yaw Propeller	**1**
$\boldsymbol{T_{20}}$- Moment Parameter of yaw Propeller	**1**
$\boldsymbol{T_p}$- Cross-coupling value	**2.1**
$\boldsymbol{T_0}$- Cross-coupling value	**3.85**
$\boldsymbol{k_c}$- Gain of Cross-coupling	**−0.34**

After Solving and substituting TRAS system constants in non-linear pitch and yaw rotor equations gives transfer function in Laplace state-space form in equation (9) and (10).

$$T_P = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \quad (9)$$

$$T_{p(s)} = \begin{bmatrix} \frac{2.346}{(s+0.934)(s^2+0.0765s+3.689)} & 0 \\ \frac{2.567(s+0.365)}{s(s+6)(s+0.944)(s+0.95)} & \frac{4.8}{s(s+9)(s+3)} \end{bmatrix} \quad (10)$$

III. FOPI^λ AND L-PID CONTROLLER DESIGNING

Two FOPI^λ and L-PID controllers are designed for pitch and yaw propellers stabilization. The general control scheme is given in equation (11) consisting proportional, integral and derivative control arrangements where as equation (12) shows error function as difference of desired propeller angle and experimental propeller angle.

$$c(t) = k_p e(t) + k_i \int e(t)dt + k_d \frac{de(t)}{dt} \quad (11)$$

$$e(t) = z_{desired\ angles} - z_{actual\ angles} \quad (12)$$

A. L-PID CONTROL SCHEME

The control structural of L-PID control arrangement with control equation (13) in mathematical form is provided in Fig.2.

$$C(s) = k_p + \frac{k_i}{s} + k_d s \quad (13)$$

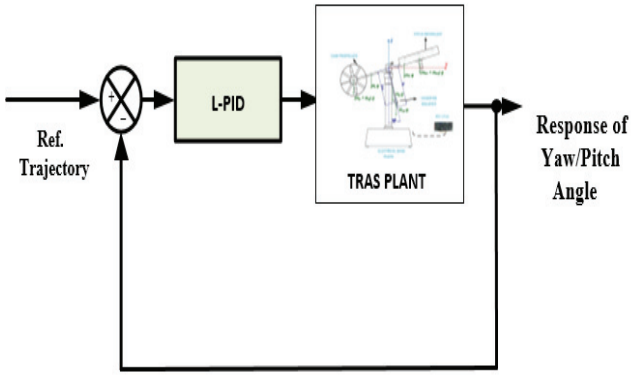


Fig. 2. L-PID control arrangement

B. FOPI^λ CONTROL SCHEME

According to equation (8), the modified transfer function of IO-PID controller structure is written in (10)

$$C(s) = k_p + \frac{k_i}{s^\lambda} \quad (10)$$

The design of IO-PID controller with TRMS plant is shown in Fig.3.

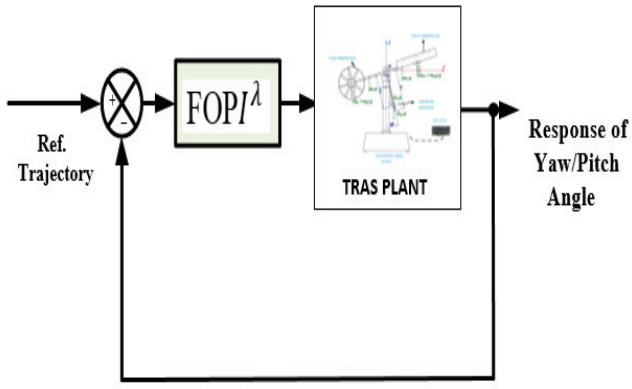


Fig. 3. FOPI$^\lambda$ control arrangement

IV. INVASIVE WEED OPTIMIZATION (IWO) OF DESIGNED FOPI^λ AND L-PID CONTROLLERS

In recent years scientists and industries are fascinated towards solving non-definitive or discontinuous control engineering problems by nature influenced non-gradient

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 18,2020 at 16:42:54 UTC from IEEE Xplore. Restrictions apply.

stochastic algorithms. Invasive weed optimization is one of the ecological influenced metamorphic optimization technique. In scientific language Weeds can be elaborated as any undesirable shrubs, trees, saplings in our cultivation or optimal solution search space. Invasive traits of weeds poses critical hazard to cultivation, so many attempts like tillage, defoliation and machine weeding are rendered to get rid of these nasty and stubborn plants.

They can reproduce by seeds or sex cells, when ancestor weed forms a seed by fertilization of egg through pollen. They flourish and reproduce at the last cycle of their life, although the way of living their entire life cycle depends upon two selections called r and k force criterion. The r criteria includes more abundance, little troop and low value to permanent colonization for winning in competition this cooperates them to succeed in insecure and unpredictable ecological conditions, whereas k criteria consist of bigger troops, slow fecundity and give more emphasis on permanent colonization. After reproduction process seeds are haphazardly dispersed in solution search space according to equations (11)-(13), through random values such that they domicile near their parent weeds shown in Fig.4, Fig.5, Fig.6.



Fig. 4. Distributing of initial spore



Fig. 5. Dispersing of spores in neighborhood orbits



Fig. 6. Rolling down of spores into most distant orbits

Weeds are very adaptive and flexible to their surroundings and environmental changes, they overrun and seize the search space by permanent colonization and also getting improved by having best fitness among plant community through interacting with peer weeds based upon their capability traits. Weeds are allowed to clone based upon the linear fitness enhancement of whole community as shown in Fig 7.

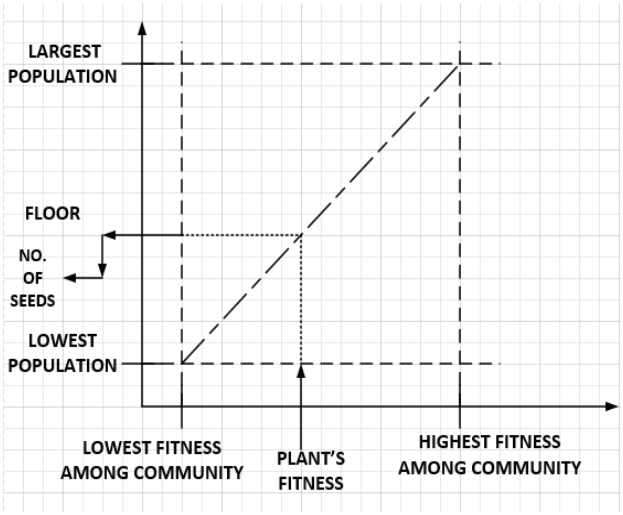


Fig. 7. Linear increase in reproduction of seeds in weeds community

$$Mean = 0 \tag{11}$$

$$Variance = varied\ over\ search\ space \tag{12}$$

Standard deviation is given by σ_{loops} ,

$$\sigma_{loops} = \frac{(loop_{max} - loop)^a}{(loop_{max})^a}(\sigma_{start} - \sigma_{end}) + \sigma_{end} \tag{13}$$

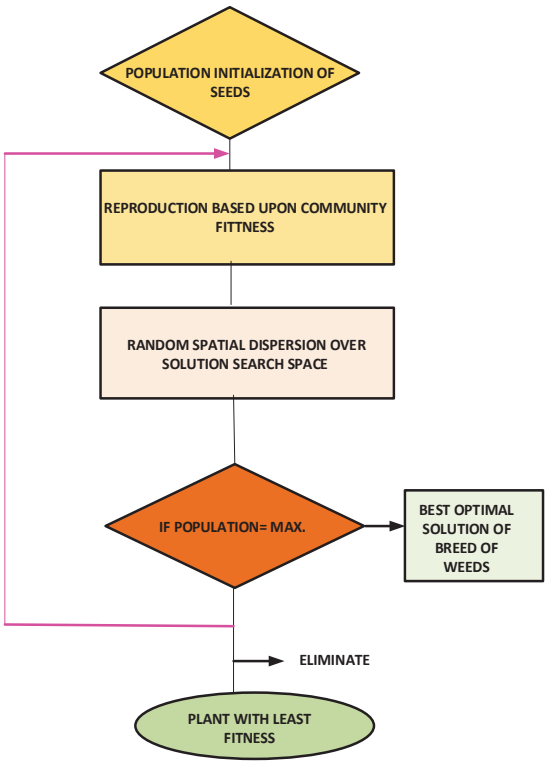


Fig. 8. IWO flowchart

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 18,2020 at 16:42:54 UTC from IEEE Xplore. Restrictions apply.

The loops of dispersion of seeds in search space works according to equitation (13) ensuring less chances to disperse in larger radial orbit in each successive iterations. The process includes elimination of less fit weeds, and revolution of their way of living from r to k force criterion. When maximum population is reached in community, weeds with highest fitness are segregated together and one with lower fitness are disqualified and not allowed to breed further. The outflow of IWO is elaborated in form of flow chart provided in Fig.8.

The tuned values of both the controller parameters are given in table.2.

TABLE II. PITCH AND YAW PROPELLERS OPTIMIZED VALUES FOR L-PID CONTROLLER

Rotor	Optimization parameter	Tuned Values
Pitch Controller	k_{p1}	6.95
	k_{i1}	8.923
	k_{d1}	7.784
Yaw Controller	k_{p2}	2.798
	k_{i2}	4.554
	k_{d2}	3.843

TABLE III. PITCH AND YAW PROPELLERS OPTIMIZED VALUES FOR FOPI^{λ} CONTROLLER

Rotor	Optimization Parameter	Tuned Values
Pitch Controller	k_{p1}	7.213
	k_{i1}	18.101
	λ_1	0.634
Yaw Controller	k_{p2}	4.051
	k_{i2}	19.544
	λ_2	0.863

These IWO tuned values of controller parameters are tested and validated on real time and stabilization graphs are obtained mentioned in section 5.

V. STABILIZATION RESULTS AND EXPERIMENTAL COMPARISION OF CONTROLLERS

The L-PID and FOPI^{λ} controller response on pitch and yaw propellers along with control signal are given in Fig.9, Fig.10, Fig.11, Fig.12

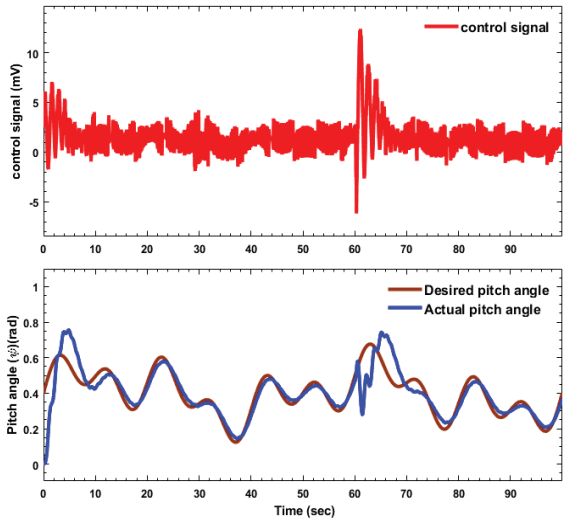


Fig. 9. L-PID controller Pitch angle response

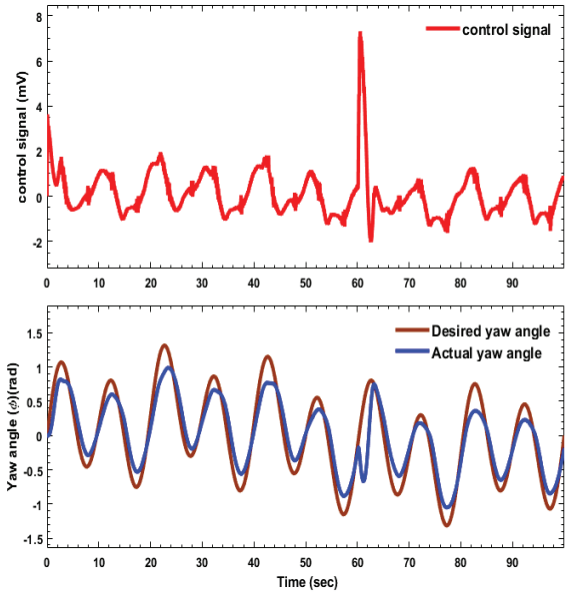


Fig. 10. L-PID Yaw angle Response

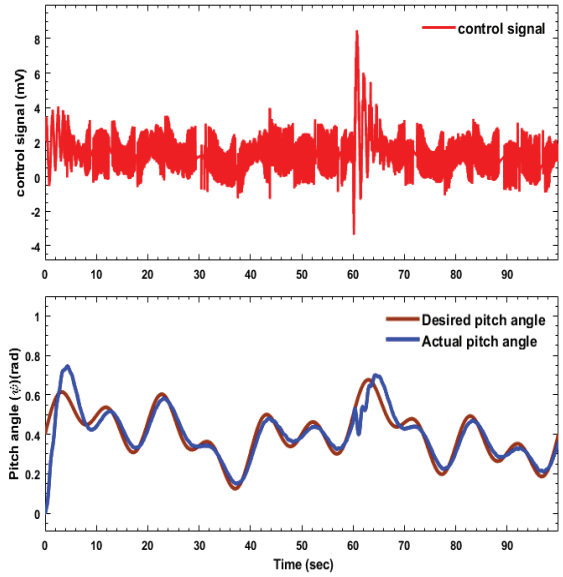


Fig. 11. FOPI^{λ} controller pitch response

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 18,2020 at 16:42:54 UTC from IEEE Xplore. Restrictions apply.

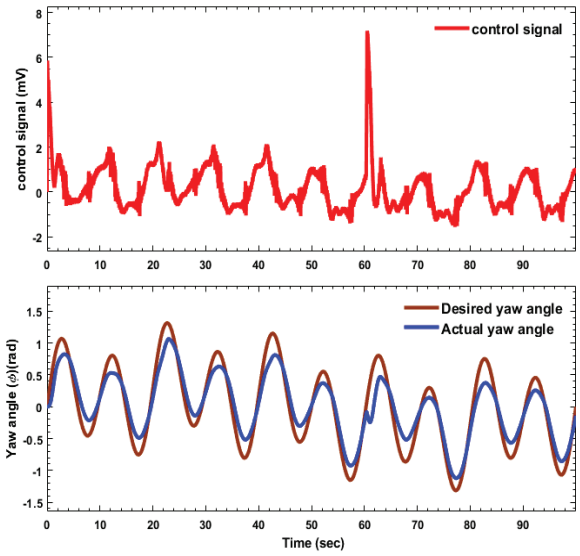


Fig. 12. FOPI^{λ} Controller yaw response

A comparison table describing time taken to stabilize after the given manual disturbance are given in table.4 and table.5 for L-PID and FOPI^{λ} controllers respectively.

TABLE IV. L-PID CONTROLLER RESPONSES

Rotor	**L-PID**		
Pitch	Distrubence Initiated (Sec)	Stabilization Instant (Sec)	Stabilization time (Sec)
	60	74	14
Yaw	60	65	5

TABLE V. FOPI^{λ} CONTROLLER RESPONSES

Rotor	***FOPI*$^{\lambda}$**		
Pitch	Distrubence Initiated (Sec)	Stabilization Instant (Sec)	Stabilization time (Sec)
	60	66	6
Yaw	60	62	2

It is observed from Table IV and Table V, that FOPI^{λ} controller takes less time to stabilize pitch and yaw angles of TRAS after the given manual disturbances and shows robust control signal then L-PID controller

VI. CONCLUSION

In this paper, optimal control of Twin rotor aerial system (TRAS) has been achieved by designing and real time experimental validation of L-PID and FOPI^{λ} controllers. These controller parameters are optimized between constraints through analytical metamorphic algorithm inspired from robust and stubborn colonization traits of invasive weeds. After experimental verification it is found that FOPI^{λ} shows better control behavior than L-PID controller for stabilization of pitch and yaw rotors of TRAS aerial plant, this signifies that it is also capable of receding wind and atmospheric disturbances on real helicopter and UAVs.

ACKNOWLEDGMENT

We would like express our warm gratitude to Indian government and DTU to provide the real time experimental plant of aerial system.

REFERENCES

[1] H. Shakhatreh *et al*., "Unmanned Aerial Vehicles (UAVs): A Survey on Civil Applications and Key Research Challenges," in *IEEE Access*, vol. 7, pp. 48572-48634, 2019.

[2] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, "A tutorial on UAVs for wireless networks: Applications, challenges, and open problems," arXiv preprint arXiv:1803.00680, 2018..

[3] Y. Sun, D. Wing Kwan Ng, D. Xu, L. Dai, and R. Schober, "Resource allocation for solar powered uav communication systems," in 2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), June 2018, pp. 1–5.

[4] G. Chmaj and H. Selvaraj, "Distributed processing applications for UAV/Drones: a survey," in Progress in Systems Engineering. Springer, 2015, pp. 449–454

[5] Ahmad, S. M., Chipperfield, A. J., & Tokhi, M. O. (2000). Dynamic modelling and optimal control of a twin rotor MIMO system. In Proceedings of the 2000 IEEE international conference on national aerospace and electronics (pp. 391–398)..

[6] Ahmad, S. M., Shaheed, M. H., Chipperfield, A. J., & Tokhi, M. O. (2000). Nonlinear modelling of a twin rotor MIMO system using radial basis function networks. In Proceedings of the 2000 IEEE international conference on national aerospace and electronics (pp. 313–320).

[7] Aldebrez, F. M., Mat-Darus, I. Z., & Tokhi, M. O. (2004). Dynamic modelling of a twin rotor system in hovering position. In Proceedings of the 2004 international symposium on control, communications and signal processing (pp. 823–826).

[8] Avila-Vilchis, J. C., Brogliato, B., Dzul, R., & Lozano, R. (2003). Nonlinear modelling and control of helicopters. Automatica, 39(9), 1583–1596.

[9] S. K. Pandey and V. Laxmi, "Control of twin rotor MIMO system using PID controller with derivative filter coefficient," *2014 IEEE Students' Conference on Electrical, Electronics and Computer Science*, Bhopal, 2014, pp. 1-6.

[10] V. S. Rao, V. I. George, S. Kamath and C. Shreesha, "Reliable robust PID controller design for TRMS," *2017 11th Asian Control Conference (ASCC)*, Gold Coast, QLD, 2017, pp. 565-569.

[11] S. K. Mishra and S. Purwar, "To design optimally tuned FOPID controller for twin rotor MIMO system," *2014 Students Conference on Engineering and Systems*, Allahabad, 2014, pp. 1-6.

[12] S. Chaudhary and A. Kumar, "Control of Twin Rotor MIMO System Using 1-Degree-of-Freedom PID, 2-Degree-of-Freedom PID and Fractional order PID Controller," *2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, 2019, pp. 746-751.

[13] H. N. Al-Duwaish and M. Bettayeb, "Control of nonlinear dynamical systems using genetic algorithms," *Proceedings of 12th IEEE International Symposium on Intelligent Control*, Istanbul, Turkey, 1997, pp. 421-424.

[14] P. Biswas, R. Maiti, A. Kolay, K. Das Sharma and G. Sarkar, "PSO based PID controller design for twin rotor MIMO system," *Proceedings of The 2014 International Conference on Control, Instrumentation, Energy and Communication (CIEC)*, Calcutta, 2014, pp. 56-60.

[15] Prithwish Chakraborty, Gourab Ghosh Roy, Swagatam Das and B. K. Panigrahi, "On population variance and explorative power of invasive weed optimization algorithm," *2009 World Congress on Nature & Biologically Inspired Computing (NaBIC)*, Coimbatore, 2009, pp. 227-232.

[16] D. Vyas and G. Parmar, "Reduced order modelling of ltisystems using invasive weed optimization," *2017 International Conference on Information, Communication, Instrumentation and Control (ICICIC)*, Indore, 2017, pp. 1-4.

[17] N. C. Patel, M. K. Debnath, B. K. Sahu, S. S. Dash and R. Bayindir, "Multi-Staged PID Controller Tuned by Invasive Weed optimization Algorithm for LFC Issues," *2018 7th International Conference on Renewable Energy Research and Applications (ICRERA)*, Paris, 2018, pp. 1358-1362

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 18,2020 at 16:42:54 UTC from IEEE Xplore. Restrictions apply.

View publication stats

Proceedings of the Fourth International Conference on Trends in Electronics and Informatics (ICOEI 2020)
IEEE Xplore Part Number: CFP20J32-ART; ISBN: 978-1-7281-5518-0

Realization of Regula-Falsi Iteration based Double Precision Floating Point Division

Riya Jain
Department of Electronics and Communication
Delhi Technological University
Delhi, India
riyajain_bt2k16@dtu.ac.in

Neeta Pandey
Department of Electronics and Communication
Delhi Technological University
Delhi, India
neetapandey@dce.ac.in

***Abstract*— One of the fundamental elements of modern microprocessor architectures is that they provide hardware support for arithmetic operations in floating point. In spite of the fact that division is a moderately rare operation in computer architecture, they are noteworthy and becoming progressively imperative in numerous modern applications. This paper presents design and implementation of Regula-Falsi (False Position) method for double precision floating point division. Floating point numbers are represented using IEEE 754 standard format. The presented algorithm iteratively computes reciprocal of 64-bit floating point divisor where two initial approximations are computed using ROM(s). The algorithm is implemented in Verilog Hardware Description Language (VHDL) and Zybo Zynq – 7000 Development FPGA Board is used to verify the same. For fairer comparison, the proposed method is compared with Newton Raphson counterpart, which is an existing algorithm that performs division operation iteratively. It has been proved that the proposed algorithm shows improvement in terms of number of iterations (computation cycles) and the accuracy of output.**

Keywords*— *Regula-Falsi, Newton-Raphson, division, multiplication, addition, floating-point, double precision, ROM

I. INTRODUCTION

In modern applications, wide range of numbers is processed, from small to large. Thus, for such applications, fixed point numbers and their arithmetic operations are unsuitable. This is where floating point numbers come into play and provide wider range of numbers. Although floating point numbers are just approximated real values but they can facilitate computation of extremely large values (astronomically large numbers) to extremely small values (atomically small numbers). Cost and performance of any system is affected by the type of number representation. Thus, representation of a number is an imperative thing in computer arithmetic. Depending upon the application and required performance, one can make a decision of which representation to be followed. Since floating point arithmetic [1] plays important role in wide range of applications, enhancing them makes a recognizable positive effect on the execution of the entire system. Among all basic arithmetic operations, division takes maximum clock cycles to compute the quotient for a given dividend and divisor. Floating point division algorithms [2] have changed over the last few decades. Most of them can be classified as subtractive (e.g. SRT division [3]-[8] and variable latency division [9]) or multiplicative (e.g. Newton-Raphson division [10], [11]) division algorithms. In subtractive floating point division, number of clock cycles is proportional to the word length of dividend. On the other hand, multiplicative algorithms perform operation by making initial approximations using look-up table [12]. Such algorithms are iterative in nature and their speed of convergence towards final result depends upon initial approximation.

In this paper, a new algorithm for floating point division using Regula-Falsi method is proposed. It computes the quotient by multiplying dividend with reciprocal of divisor. Reciprocal unit of the divider uses multiplication and addition/subtraction operations to compute the reciprocal of divisor. This method makes two initial approximations that are read from two 2^{10} x 10 ROMs and this value is updated in every iteration to compute the final result. Newton-Raphson is an existing iterative division algorithm which acts as a counterpart for the proposed algorithm. Newton-Raphson also makes an initial approximation, whose values are determined as given in [13], using one 2^{10} x 10 ROM. Both the algorithms are implemented in parallel for comparison.

The remaining of this paper is organized as follows: Section 2 comprises of mathematical analysis of the algorithm followed by its hardware implementation. This section also highlights the procedure of initial approximations. Section 3 makes comparison between the proposed algorithm and Newton-Raphson algorithm in terms of required computation cycles (iterations) and accuracy. Simulation results for an example are also placed in this section. Finally, conclusions are given in Section 4.

II. REGULA-FALSI ITERATIVE DIVISION USING LOOK-UP TABLE

Regula-Falsi, also called false position method is one of the equations solving technique used to find roots of any function in one variable. The method is simply a trial and error technique of using test ("false") values for the variable and then adjusting the test value according to the outcome.

A. Mathematical Analysis

Consider a function $f(x)$, shown in Fig. 1, that cuts the x-axis at point O. Two initial approximations x_0 and x_1 are made such that $f(x_0)$ is negative and $f(x_1)$ is positive.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:26:09 UTC from IEEE Xplore. Restrictions apply.

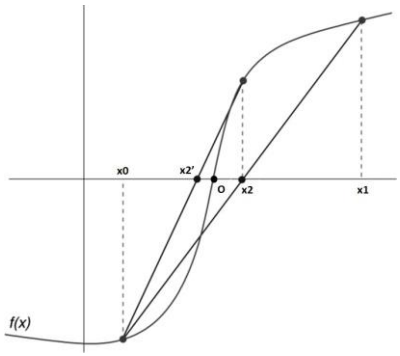


Fig. 1: Geometrical representation of the Regula-Falsi method

Construct the line through the points $(x0, f(x0))$ and $(x1, f(x1))$. This line is a chord of the graph of the function f and is represented in point-slope form (1).

$$y - f(x1) = \frac{f(x1) - f(x0)}{x1 - x0}(x - x1) \quad (1)$$

Considering $x2$ as x-intercept point of this line, i. e., y = 0 for $x = x2$, (1) may be rewritten as

$$f(x1) + \frac{f(x1) - f(x0)}{x1 - x0}(x2 - x1) = 0 \quad (2)$$

After simplifying,

$$x2 = \frac{x0\, f(x1) - x1\, f(x0)}{f(x1) - f(x0)} \quad (3)$$

Values of $x0$ and $x1$ are now updated according to sign of $f(x2)$. Thus there are two possible cases:

1. f(x2) > 0 ; x1 is replaced by x2 such that x0 remains the same.
2. f(x2) <0 ; x0 is replaced by x2 such that x1 remains the same.

After getting new values of $x0$ and $x1$, above process is repeated till two approximations get close together, and coincide (or almost coincide) at point O. To perform floating point division, both dividend (A) and divisor (B) are represented in IEEE 754 double precision binary floating point format. The quotient Q is obtained by multiplying A with reciprocal of B, given as:

$$Q = \frac{A}{B} \quad (4)$$

Or

$$Q = A\frac{1}{B} \quad (5)$$

Now, to compute the value of reciprocal $1/B$, let

$$f(x) = \frac{1}{x} - B \quad (6)$$

such that root of the function is at $x = 1/B$. With $x0$ and $x1$ as two initial approximations, substitute values of $f(x0)$ and $f(x1)$ in (3) to get the value of $x2$ as:

$$x2 = \frac{x0^2\,(1 - B\,x1) - x1^2\,(1 - B\,x0)}{x0 - x1} \quad (7)$$

$$x2 = \frac{x0^2 - x1^2 - B\,x1\,x2\,(x0 - x1)}{x0 - x1} \quad (8)$$

$$x2 = x0 + x1 - B\,x0\,x1 \quad (9)$$

In (9), two multiplications and two addition/ subtraction are used to compute $x2$. To check the sign of $f(x2)$, there are two cases:

1. If $Bx2 < 1$, then $f(x2) > 0$
2. If $Bx2 > 1$, then $f(x2) < 0$

When $x2$ coincides (or almost coincide) with point O, $x2$ $1/B$. Final result is obtained by substituting value of $1/B$ in (5).

B. Block Diagram

The block diagrammatic representation of the method is depicted in Fig. 2. The dotted block is the Reciprocal Unit that computes $1/B$ (1/divisor). The output of this block is multiplied with A (dividend) to get the final result.

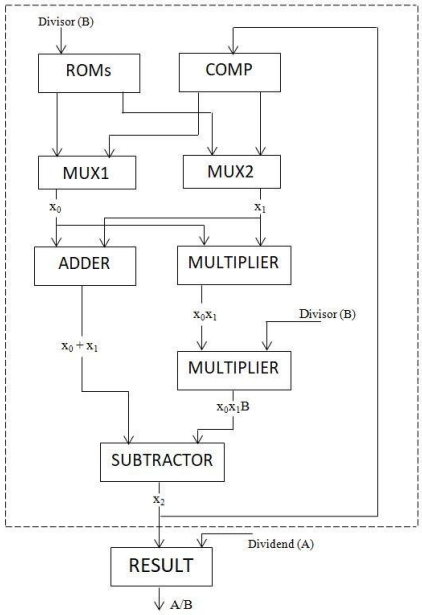


Fig. 2. Regula-Falsi Divider block diagram

In first clock cycle initial approximations $x0$ and $x1$, each of 10 bits, are obtained from two 2_{10} x 10 ROMs. Most significant 10 bits of the divisor (B) are truncated, and are

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:26:09 UTC from IEEE Xplore. Restrictions apply.

given as input to two ROMs that give 64-bit approximations of x_0 and x_1 at their output (detailed explanation of how these approximations are taken is in next section). In second clock cycle, initial values of x_0 and x_1, output of ROMs, are selected by MUX1 and MUX2. In next clock cycle, addition/subtraction and multiplication operations are performed to compute x_2 as in (9). The result obtained for x_2, is then sent as an input to COMP block to check the sign of $f(x_2)$ by computing Bx_2 and comparing the value with 1, as explained in previous section. Values of x_0 and x_1 are then updated and new values of x_0 and x_1 are selected by MUX1 and MUX2 in subsequent clock cycle. The process is repeated till reciprocal of divisor (B) approaches it's true value. Finally, in last clock cycle, end result is obtained at output by multiplying final value of x_2 with dividend (A). Similar approach is followed in Newton-Raphson algorithm where reciprocal of divisor (B) is computed and then multiplied with dividend (A) by taking an initial approximation of 10 bits. Even though they follow similar approach, it will be shown further that Regula-Falsi converges faster than Newton-Raphson and computes result with higher accuracy.

C. Look-Up Table Values

Look-up tables used in this method are of size 2_{10} x 10. Most significant 10-bits of the divisor (B) are truncated and used to find initial values of x_0 and x_1. They are given at input as address of two ROMs. 10-bits output of ROM1 and ROM2, corresponding to the address, determine the values of x_0 and x_1 respectively. 10-bits values stored in ROM can be determined as follows:

$$B = 1.\underbrace{x_1x_2x_3x_4x_5x_6x_7x_8x_9x_{10}}\ldots.x_{52} \quad (10)$$

$$B' = 1.x_1x_2x_3x_4x_5x_6x_7x_8x_9x_{10} \quad (11)$$

$$B' = 1 + x_1 2^{-1} + x_2 2^{-2} + x_3 2^{-3} + \ldots + x_{10} 2^{-10} \quad (12)$$

$$X' = \frac{1}{B'} \quad (13)$$

$$= 1.\underbrace{x'_1x'_2x'_3x'_4x'_5x'_6x'_7x'_8x'_9x'_{10}}\ldots.x'_{52}$$

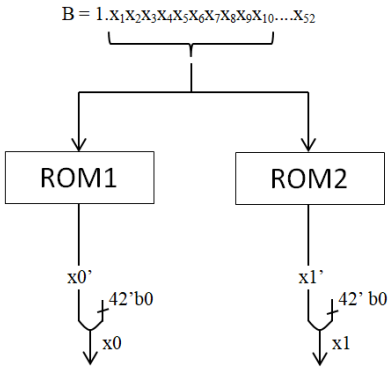


Fig. 3: Initial approximation using ROM

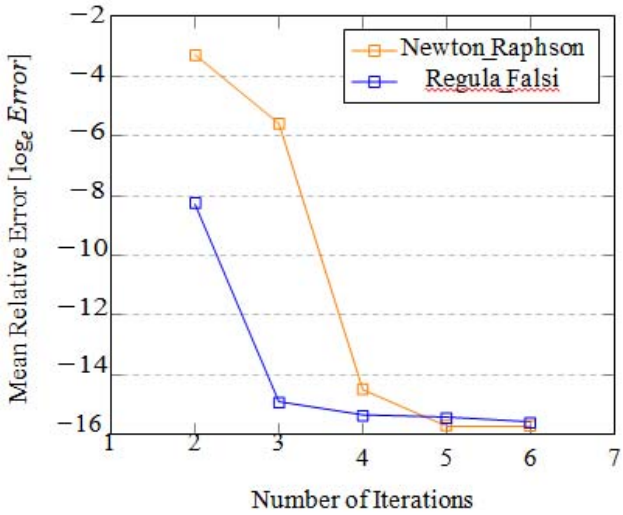


Fig. 4: Mean Relative Error

Values of ROM1:

$$x0' = x'_1x'_2x'_3x'_4x'_5x'_6x'_7x'_8x'_9x'_{10} - 0000000001$$

Values of ROM2:

$$x1' = x'_1x'_2x'_3x'_4x'_5x'_6x'_7x'_8x'_9x'_{10} + 0000000001$$

Finally, as shown in Fig. 3, x0 and x1 are obtained by appending 42 zeros to LSB of the output of ROM1 and ROM2 respectively to make it 52-bit significand of the floating point numbers.

III. Results and Comparison

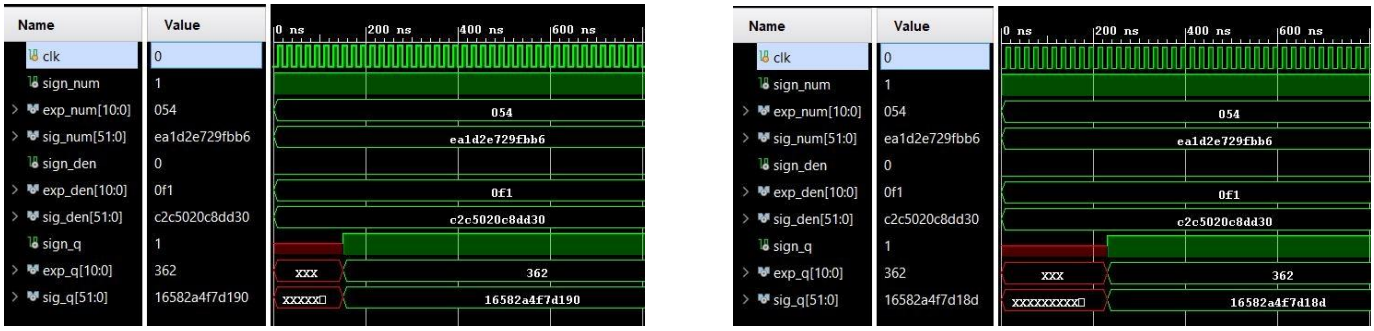
The proposed algorithm is implemented in Verilog HDL and simulated using ISIM-Xilinx. The FPFA board used is Zybo Zynq– 7000. Two examples are stated in Table 1 and Table 2 where N and D represent dividend and divisor respectively, Q_Actual denotes expected value of quotient for given dividend and divisor, Q_Regula denotes the quotient value obtained from Regula-Falsi method in 8 clock cycles, and Q_Newton_m denotes the quotient value obtained from Newton-Raphson method in 'm' clock cycles. In the first example, Newton-Raphson and Regula-Falsi method computes the result with $2.8e^{-16}\%$ error in 12 clock cycles and 8 clock cycles respectively while error increases to $-2.3e^{-10}\%$ for Newton-Raphson method when result is computed in 10 clock cycles. In second example, Newton-Raphson and Regula-Falsi method computes the result in 14 and 8 clock cycles respectively with 0% error while error increases to $-7.5e^{-11}\%$ for Newton-Raphson method when clock cycles are reduced to 12. For ready reference and comparison, simulation results of example 2 for the proposed method and Newton-Raphson method are placed in Fig. 4 and Fig. 5 respectively. It may be seen that in case of Regula-Falsi method, final result is obtained in 8 clock cycles: 7 clock cycles to compute reciprocal and 1 clock cycle to calculate the final result and in case of Newton Raphson method, final result is obtained in 14 clock cycles, which is 50% more than the proposed algorithm. For fairer comparison, simulations are performed for random 15 data sets and it is found that both the methods give similar approximated results

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:26:09 UTC from IEEE Xplore. Restrictions apply.

TABLE I: Example for floating point division by Regula-Falsi and Newton-Raphson method

Iterations	Regula-Falsi		Newton-Raphson	
	Output Quotient	MRE	Output Quotient	MRE
2	$-5.951592619252755e-48$	$5.6969893e-10$	$-5.950389005329300e-48$	$2.0223449e-04$
3	$-5.951592622643376e-48$	$8.1687894e-16$	$-5.951592379230435e-48$	$4.0898789e-08$
4	$-5.951592622643374e-48$	$4.0843947e-16$	$-5.951592622643361e-48$	$1.6337578e-15$
5	$-5.951592622643371e-48$	0	$-5.951592622643372e-48$	$2.0421973e-16$
6	$-5.951592622643371e-48$	0	$-5.951592622643372e-48$	$2.0421973e-16$

*Dividend(A) = $-4.119942981079925e^{-283}$, Divisor(B) = $6.922421009471027e^{-236}$, True value of quotient(Q) = $-5.951592622643371e^{-48}$



*sign xx: sign bit of xx, exp xx: exponent of xx, sig xx: significand of xx for xx = num (dividend), den (divisor), and q (quotient)

Fig. 5: Simulation results for both methods

but the proposed algorithm is better in terms of number of clock cycles and iterations required, that is, proposed method requires 4-6 clock cycles less than that required in Newton-Raphson method.

One of the samples, randomly selected from all samples, is stated as an example for ready reference in Table 1. For given value of dividend(A) and divisor(B), results of both methods for 2, 3, 4, 5, and 6 iterations are recorded in the table. Mean Relative Error (MRE) for these values with respect to true value of quotient(Q) are also mentioned. It is clear from the given example that it took 3 iterations (8 clock cycles) for Regula-Falsi method to nearly converge to the result while Newton-Raphson method took 5-6 iterations (12- 14 clock cycles). Simulation results of the same example (in hexadecimal representation) for Regula-Falsi (3 iterations) and Newton-Rapshson (5 iterations) are placed in Fig. 5 (a) and (b) respectively.

Possible explanation for this faster convergence in case of Regula-Falsi method is utilization of 2 initial approximation. Moreover, there is an advantage of using Regula-Falsi over Newton-Raphson method that it will always converge since two approximations are taken such that one lies on left side of the root and other on the right side. In Newton-Raphson method, the result of a function may diverge if the initial approximation is too far away from the root.

IV. Conclusion

The floating point division using Regula-Falsi iteration algorithm is presented in the paper. The proposed design computes the reciprocal by utilizing two look-up tables to obtain initial approximations. The algorithm takes eight cycles to compute the quotient for given double precision floating- point dividend and divisor with an MRE in a range of $e^{-}16$. A precision of 52-bit is obtained in the conducted experiment, but it is not guaranteed the correctly rounded least significant bit for all cases. The proposed algorithm is compared with Newton-Raphson algorithm of division and is verified to be faster with more accurate results. Number of clock cycles and iterations required in Regula-Falsi method is much less than that in Newton-Raphson method.

References

[1] B. Parhami, “Computer Arithmetic Algorithms And Hardware Designs,” 2nd ed., New York, USA: Oxford University Press, 2010.
[2] S.F. Oberman and M.J. Flynn, “An Analysis of Division Algorithms and Implementations,” Computer Systems Lab., Dept. of Elect. Eng. and Comp. Sci., Stanford Univ., Tech. Rep. CSL-TR-95-675, July 1995.
[3] P. Kornerup, “Digit Selection for SRT Division and Square Root”, IEEE Transactions on Computers, Vol. 54, No. 3, March 2005, pp. 294-303.
[4] L.S. Niwal, S. P. Hajare, “Design of Radix 4 Divider Circuit Using SRT Algorithm”, in ICCSP, Melmaruvathur, India, 2015, pp. 1107-1110.
[5] David M. Russinoff, “Computation and Formal verification of SRT Quotient and Square Root Digit Selection Tables”, in IEEE Transactions on Computers, Vol. 62, No. 5, 2013, pp. 900-913.
[6] D.L. Harris, S.F. Oberman, and M.A. Horowitz, “SRT Division Architectures and Implementations,” in Proceeding of 13th IEEE Sympsoium on Computer Arithmetic, Asilomar, USA, 1997, pp. 18-

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:26:09 UTC from IEEE Xplore. Restrictions apply.

25.
[7] A. Amaricai and O. Boncalo, "SRT Radix-2 Dividers with (5,4) Redundant Representation of Partial Remainder," in NORCHIP, Vilnius, Lithuania, 2013, pp. 1-5.
[8] P. Kornerup, "Revisiting SRT Quotient Digit Selection," in proceeding of 16th IEEE Symposium on Computer Arithmetic, 2003, pp. 38-45.
[9] P.M. Seidel, "Variable Latency Division," in MWSCAS, Windsor, Canada, 2018, pp. 944-947.
[10] N. Singh and T. N. Sasamal, "Design and Synthesis of Single Precision Floating Point Division based on Newton-Raphson Algorithm on FPGA", in MATEC Web of Conferences, ICAET, 2016.
[11] U. Kucukkabak and A. Akkas, "Design and Implementation of Reciprocal Unit Using Table Look-up and Newton-Raphson Iteration", in Euromicro Symposium on Digital System Design, Rennes, France, 2004.
[12] M. Ito, N. Takagi and S. Yajima, "Efficient Initial Approximation for Multiplicative Division and Square Root by a Multiplication with Operand Modification," in IEEE Transactions on Computers, Vol. 46, No. 4, 1997, pp. 495-498.
[13] D. Das Sarma and D.W. Matula, "Measuring The Accuracy Of ROM Reciprocal Tables," in Proceedings of IEEE 11th Symposium on Computer Arithmetic, Windsor, Canada, 2002, pp. 95-102.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:26:09 UTC from IEEE Xplore. Restrictions apply.

Real-Time Implementation of the Optimal Predictor and Optimal Filter: Accuracy Versus Latency

SYED ASEEM UL ISLAM, ANKIT GOEL, and DENNIS S. BERNSTEIN

Summary

The Kalman filter uses noisy sensor measurements to estimate the states of a linear system that is subject to disturbances. An elegant implementation of this algorithm is the optimal two-step filter (OTSF), which involves a forecast step based on the dynamics and a data assimilation step based on sensor data. However, what happens if the order of the steps between state estimates is reversed? This variation of the OTSF, which is the optimal two-step predictor (OTSP), provides slightly earlier state estimates than the OTSF based on slightly older data. The goal of this article is to compare the accuracy of the OTSF and the OTSP to assess whether or not the delayed estimates of the OTSF are more accurate than the faster estimates of the OTSP.

Although the Kalman filter is often presented within a continuous-time context [1]–[4], the original derivation was provided in discrete time [5]. In practice, controllers and observers are invariably implemented digitally; thus discrete-time algorithms deserve special attention. This article focuses on the real-time implementation of the discrete-time Kalman filter and its relation to the discrete-time Kalman predictor. These algorithms differ in subtle ways, as highlighted in this article. Equations for the discrete-time Kalman filter and predictor are given in [6]. Unfortunately, the covariance matrix becomes indefinite after a few steps, which shows that those equations are erroneous. The present article corrects and extends the results from [6].

This derivation is based on necessary conditions and thus does not prove that the optimal one-step predictor (OOSP) is the globally optimal one. However, this derivation is succinct and uses minimal mathematics, making it efficient for classroom presentation. As shown in [2, p. 46], in the absence of Gaussian processes, this derivation yields the optimal linear predictor. Next, the OOSP is recast as an optimal two-step predictor (OTSP). The steps of the OTSP are then reversed to obtain the optimal two-step filter (OTSF), as described in "Summary." Finally, the steps of the OTSF are combined to obtain the optimal one-step filter (OOSF).

The OTSP and OTSF consist of assimilation and forecast updates. The OTSP begins with an assimilation update followed by a forecast update, whereas the OTSF begins with a forecast update followed by an assimilation update. An additional distinction between these two algorithms concerns their real-time implementation. Within the context of constraints on data collection and computation, the OTSP and OTSF use different data to produce state estimates that are available at different times. Overall, the OTSF uses more recent data but produces state estimates at a later time. It is thus expected that the OTSF estimates are more accurate than those of the OTSP, with latency the price paid for the enhanced accuracy of the filter.

To investigate the validity of the expected accuracy/latency tradeoff, the accuracy of the OTSP and OTSF is compared by means of a numerical example. Somewhat surprisingly, the numerical results show that the filter estimates are not uniformly better than the predictor estimates. This discrepancy is traced to the use of the initial data, suggesting a third algorithm: an optimal two-step filter that invokes an assimilation update at start-up (OTSFSU) to utilize the initial data. Aside from the initial start-up step, the OTSFSU is identical to the OTSF. We then revisit the accuracy/latency tradeoff by numerically comparing the accuracy of the OTSP and OTSFSU to determine whether the OTSFSU is uniformly more accurate than the OTSP. Spoiler alert: It is.

OPTIMAL ONE-STEP PREDICTOR

Consider the system

$$x_{k+1} = A_k x_k + B_k u_k + w_k, \tag{1}$$

$$y_k = C_k x_k + D_k u_k + v_k, \tag{2}$$

Digital Object Identifier 10.1109/MCS.2019.2961588
Date of current version: 17 March 2020

1066-033X/20©2020IEEE

Authorized licensed use limited to: University of Michigan Library. Downloaded on March 28,2020 at 16:38:43 UTC from IEEE Xplore. Restrictions apply.

Controllers and observers are invariably implemented digitally; thus discrete-time algorithms deserve special attention.

where, for all $k \geq 0$, $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^m$, $y_k \in \mathbb{R}^p$, $w_k \in \mathbb{R}^n$, and $v_k \in \mathbb{R}^p$, and the real matrices A_k, B_k, C_k, and D_k are of corresponding size. All stochastic processes are assumed to have finite second moments; the details of the densities are not relevant to the derivation below. Specifically, $w_k \in \mathbb{R}^n$ is a zero-mean, white-noise disturbance signal with the covariance $Q_k \triangleq \mathbb{E}[w_k w_k^{\mathrm{T}}] \in \mathbb{R}^{n\times n}$, and $v_k \in \mathbb{R}^p$ is a zero-mean, white sensor noise with the covariance $R_k \triangleq \mathbb{E}[v_k v_k^{\mathrm{T}}] \in \mathbb{R}^{p\times p}$. The cross-covariance between v_k and w_k is denoted by $S_k \triangleq \mathbb{E}[w_k v_k^{\mathrm{T}}] \in \mathbb{R}^{n\times p}$. The input u_k is assumed to be known; in practice, u_k is typically a control input. Consider the OOSP

$$\hat{x}_{k+1} = A_k \hat{x}_k + B_k u_k + K_k(y_k - \hat{y}_k), \tag{3}$$

$$\hat{y}_k = C_k \hat{x}_k + D_k u_k, \tag{4}$$

where $\hat{x}_k \in \mathbb{R}^n$ is an estimate of the state x_k; the optimal gain $K_k \in \mathbb{R}^{n\times p}$ is determined below.

Defining the state error

$$e_k \triangleq x_k - \hat{x}_k, \tag{5}$$

it follows that

$$\begin{aligned} e_{k+1} &= x_{k+1} - \hat{x}_{k+1} \\ &= \tilde{A}_k e_k + \tilde{w}_k, \end{aligned} \tag{6}$$

where

$$\tilde{A}_k \triangleq A_k - K_k C_k, \tag{7}$$

$$\tilde{w}_k \triangleq w_k - K_k v_k. \tag{8}$$

Next, we define the cost

$$\begin{aligned} J_k(K_k) &\triangleq \mathbb{E}[e_{k+1}^{\mathrm{T}} e_{k+1}] \\ &= \mathrm{tr}\,\mathbb{E}[e_{k+1}^{\mathrm{T}} e_{k+1}] \\ &= \mathbb{E}[\mathrm{tr}\,(e_{k+1}^{T} e_{k+1})] \\ &= \mathbb{E}[\mathrm{tr}(e_{k+1} e_{k+1}^{\mathrm{T}})] \\ &= \mathrm{tr}\,\mathbb{E}[e_{k+1} e_{k+1}^{\mathrm{T}}] \\ &= \mathrm{tr}\,P_{k+1}, \end{aligned} \tag{9}$$

where the *error covariance* is defined by

$$P_{k+1} \triangleq \mathbb{E}[e_{k+1} e_{k+1}^{\mathrm{T}}] \in \mathbb{R}^{n\times n}. \tag{10}$$

It thus follows that

$$\begin{aligned} e_{k+1} e_{k+1}^{\mathrm{T}} &= (\tilde{A}_k e_k + \tilde{w}_k)(\tilde{A}_k e_k + \tilde{w}_k)^{\mathrm{T}} \\ &= \tilde{A}_k e_k e_k^{\mathrm{T}} \tilde{A}_k^{\mathrm{T}} + \tilde{A}_k e_k \tilde{w}_k^{\mathrm{T}} + \tilde{w}_k e_k^{\mathrm{T}} \tilde{A}_k^{\mathrm{T}} + \tilde{w}_k \tilde{w}_k^{\mathrm{T}}. \end{aligned} \tag{11}$$

Next, since w_k and v_k are white-noise sequences that affect e_{k+1} but not e_k, it follows that e_k and $\tilde{w}_k$ are uncorrelated. Furthermore, since w_k and v_k have a zero mean, it follows that $\tilde{w}_k$ also has a zero mean. Therefore, $\mathbb{E}[e_k \tilde{w}_k^{\mathrm{T}}] = \mathbb{E}[e_k]\mathbb{E}[\tilde{w}_k^{\mathrm{T}}] = 0$. Taking the expected value of (11) yields

$$\begin{aligned} P_{k+1} &= \mathbb{E}[e_{k+1} e_{k+1}^{\mathrm{T}}] \\ &= \mathbb{E}[\tilde{A}_k e_k e_k^{\mathrm{T}} \tilde{A}_k^{\mathrm{T}} + \tilde{A}_k e_k \tilde{w}_k^{\mathrm{T}} + \tilde{w}_k e_k^{\mathrm{T}} \tilde{A}_k^{\mathrm{T}} + \tilde{w}_k \tilde{w}_k^{\mathrm{T}}] \\ &= \tilde{A}_k \mathbb{E}[e_k e_k^{\mathrm{T}}] \tilde{A}_k^{\mathrm{T}} + \tilde{A}_k \mathbb{E}[e_k \tilde{w}_k^{\mathrm{T}}] + \mathbb{E}[\tilde{w}_k e_k^{T}] \tilde{A}_k^{\mathrm{T}} + \mathbb{E}[\tilde{w}_k \tilde{w}_k^{\mathrm{T}}] \\ &= \tilde{A}_k P_k \tilde{A}_k^{\mathrm{T}} + \mathbb{E}[\tilde{w}_k \tilde{w}_k^{\mathrm{T}}] \\ &= \tilde{A}_k P_k \tilde{A}_k^{\mathrm{T}} + \mathbb{E}[(w_k - K_k v_k)(w_k - K_k v_k)^{\mathrm{T}}] \\ &= \tilde{A}_k P_k \tilde{A}_k^{\mathrm{T}} + \mathbb{E}[w_k w_k^{\mathrm{T}} - w_k v_k^{\mathrm{T}} K_k^{\mathrm{T}} - K_k v_k w_k^{\mathrm{T}} + K_k v_k v_k^{\mathrm{T}} K_k^{\mathrm{T}}] \\ &= \tilde{A}_k P_k \tilde{A}_k^{\mathrm{T}} + \mathbb{E}[w_k w_k^{\mathrm{T}}] - \mathbb{E}[w_k v_k^{\mathrm{T}}] K_k^{\mathrm{T}} - K_k \mathbb{E}[v_k w_k^{\mathrm{T}}] \\ &\quad + K_k \mathbb{E}[v_k v_k^{\mathrm{T}}] K_k^{\mathrm{T}} \\ &= \tilde{A}_k P_k \tilde{A}_k^{\mathrm{T}} + Q_k + K_k R_k K_k^{\mathrm{T}} - \mathbb{E}[w_k v_k^{\mathrm{T}}] K_k^{\mathrm{T}} - K_k [\mathbb{E}[w_k v_k^{\mathrm{T}}]]^{\mathrm{T}}. \end{aligned} \tag{12}$$

The cost (9) is thus given by

$$\begin{aligned} J_k(K_k) &= \mathrm{tr} P_{k+1} \\ &= \mathrm{tr}[(A_k - K_k C_k) P_k (A_k - K_k C_k)^{\mathrm{T}} + Q_k + K_k R_k K_k^{\mathrm{T}} \\ &\quad - S_k K_k^{\mathrm{T}} - K_k S_k^{\mathrm{T}}] \\ &= \mathrm{tr}[A_k P_k A_k^{\mathrm{T}} - K_k C_k P_k A_k^{\mathrm{T}} - A_k P_k C_k^{\mathrm{T}} K_k^{\mathrm{T}} + K_k C_k P_k C_k^{\mathrm{T}} K_k^{\mathrm{T}} \\ &\quad + Q_k + K_k R_k K_k^{\mathrm{T}} - S_k K_k^{\mathrm{T}} - K_k S_k^{\mathrm{T}}] \\ &= \mathrm{tr}[K_k (C_k P_k C_k^{\mathrm{T}} + R_k) K_k^{\mathrm{T}} - K_k C_k P_k A_k^{\mathrm{T}} - A_k P_k C_k^{\mathrm{T}} K_k^{\mathrm{T}} \\ &\quad + A_k P_k A_k^{\mathrm{T}} + Q_k - S_k K_k^{\mathrm{T}} - K_k S_k^{\mathrm{T}}] \\ &= \mathrm{tr}[K_k (C_k P_k C_k^{\mathrm{T}} + R_k) K_k^{\mathrm{T}}] - 2\mathrm{tr}\, K_k (C_k P_k A_k^{\mathrm{T}} + S_k^{\mathrm{T}}) \\ &\quad + \mathrm{tr}(A_k P_k A_k^{\mathrm{T}} + Q_k). \end{aligned} \tag{13}$$

To minimize $J_k(K_k)$, note that

$$\frac{\mathrm{d}J_k(K_k)}{\mathrm{d}K_k} = 2(C_k P_k C_k^{\mathrm{T}} + R_k) K_k^{\mathrm{T}} - 2C_k P_k A_k^{\mathrm{T}} - 2S_k^{\mathrm{T}}. \tag{14}$$

Setting the derivative to zero yields the *optimal one-step predictor gain*

$$K_k = (A_k P_k C_k^{\mathrm{T}} + S_k)(C_k P_k C_k^{\mathrm{T}} + R_k)^{-1}. \tag{15}$$

By substituting (7) and (15) in (12), the *one-step predictor* is thus given by

$$\hat{x}_{k+1} = A_k \hat{x}_k + B_k u_k + (A_k P_k C_k^{\mathrm{T}} + S_k)(C_k P_k C_k^{\mathrm{T}} + R_k)^{-1}(y_k - C_k \hat{x}_k), \tag{16}$$

$$P_{k+1} = A_k P_k A_k^{\mathrm{T}} - (A_k P_k C_k^{\mathrm{T}} + S_k)(C_k P_k C_k^{\mathrm{T}} + R_k)^{-1} \cdot (C_k P_k A_k^{\mathrm{T}} + S_k^{\mathrm{T}}) + Q_k. \tag{17}$$

Authorized licensed use limited to: University of Michigan Library. Downloaded on March 28,2020 at 16:38:43 UTC from IEEE Xplore. Restrictions apply.

Optimal Two-Step Predictor and Two-Step Filter With Correlated Disturbance and Sensor Noise

This sidebar gives the update equations for the optimal two-step predictor (OTSP) and optimal two-step filter (OTSF) in the case where $S_k \neq 0$. For the OTSP, the *assimilation update* is given by

$$x_k^{\mathrm{a}} = x_k^{\mathrm{f}} + P_k^{\mathrm{f}} C_k^{\mathrm{T}} (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1} (y_k - C_k x_k^{\mathrm{f}}), \tag{S1}$$

$$P_k^{\mathrm{a}} = P_k^{\mathrm{f}} - P_k^{\mathrm{f}} C_k^{\mathrm{T}} (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1} C_k P_k^{\mathrm{f}}, \tag{S2}$$

which are identical to (20) and (21). The *forecast update* is given by

$$x_{k+1}^{\mathrm{f}} = A_k x_k^{\mathrm{a}} + B_k u_k + S_k (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1} (y_k - C_k x_k^{\mathrm{f}}), \tag{S3}$$

$$\begin{aligned} P_{k+1}^{\mathrm{f}} = {} & A_k P_k^{\mathrm{a}} A_k^{\mathrm{T}} - S_k (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1} C_k P_k^{\mathrm{f}} A_k^{\mathrm{T}} \\ & - A_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1} S_k^{\mathrm{T}} \\ & - S_k (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1} S_k^{\mathrm{T}} + Q_k, \end{aligned} \tag{S4}$$

which are (22) and (23) with additional terms involving S_k. Defining the *OTSP gain*

$$K_k^{\mathrm{f}} \triangleq P_k^{\mathrm{f}} C_k^{\mathrm{T}} (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1}, \tag{S5}$$

(S1) and (S2) can be written as

$$x_k^{\mathrm{a}} = x_k^{\mathrm{f}} + K_k (y_k - C_k x_k^{\mathrm{f}}), \tag{S6}$$

$$P_k^{\mathrm{a}} = (I - K_k C_k) P_k^{\mathrm{f}}, \tag{S7}$$

which are identical to (25) and (26).

For the OTSF, the *forecast update* is given by

$$x_{k+1}^{\mathrm{f}} = A_k x_k^{\mathrm{a}} + B_k u_k + S_k (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1} (y_k - C_k x_k^{\mathrm{f}}), \tag{S8}$$

$$\begin{aligned} P_{k+1}^{\mathrm{f}} = {} & A_k P_k^{\mathrm{a}} A_k^{\mathrm{T}} - S_k (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1} C_k P_k^{\mathrm{f}} A_k^{\mathrm{T}} \\ & - A_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1} S_k^{\mathrm{T}} \\ & - S_k (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1} S_k^{\mathrm{T}} + Q_k, \end{aligned} \tag{S9}$$

which are (27) and (28) with additional terms involving S_k. The *assimilation update* is given by

$$\begin{aligned} x_{k+1}^{\mathrm{a}} = {} & x_{k+1}^{\mathrm{f}} + P_{k+1}^{\mathrm{f}} C_{k+1}^{T} (C_{k+1} P_{k+1}^{\mathrm{f}} C_{k+1}^{T} + R_{k+1})^{-1} \\ & \cdot (y_{k+1} - C_{k+1} x_{k+1}^{\mathrm{f}}), \end{aligned} \tag{S10}$$

$$P_{k+1}^{\mathrm{a}} = P_{k+1}^{\mathrm{f}} - P_{k+1}^{\mathrm{f}} C_{k+1}^{\mathrm{T}} (C_{k+1} P_{k+1}^{\mathrm{f}} C_{k+1}^{\mathrm{T}} + R_{k+1})^{-1} C_{k+1} P_{k+1}^{\mathrm{f}}, \tag{S11}$$

which are identical to (29) and (30). In terms of the *OTSF gain*

$$K_{k+1}^{\mathrm{f}} \triangleq P_{k+1}^{\mathrm{f}} C_{k+1}^{\mathrm{T}} (C_{k+1} P_{k+1}^{\mathrm{f}} C_{k+1}^{\mathrm{T}} + R_{k+1})^{-1}, \tag{S12}$$

(S10) and (S11) can be written as

$$x_{k+1}^{\mathrm{a}} = x_{k+1}^{\mathrm{f}} + K_{k+1} (y_{k+1} - C_{k+1} x_{k+1}^{\mathrm{f}}), \tag{S13}$$

$$P_{k+1}^{\mathrm{a}} = (I - K_{k+1} C_{k+1}) P_{k+1}^{\mathrm{f}}, \tag{S14}$$

which are identical to (32) and (33).

The OTSF in the traditional notation is

$$\begin{aligned} \hat{x}_{k+1|k} = {} & A_k \hat{x}_{k|k} + B_k u_k + S_k (C_k P_{k|k-1} C_k^{T} + R_k)^{-1} \\ & \cdot (y_k - C_k \hat{x}_{k|k-1}), \end{aligned} \tag{S15}$$

$$\begin{aligned} P_{k+1|k} = {} & A_k P_{k|k} A_k^{T} - S_k (C_k P_{k|k-1} C_k^{T} + R_k)^{-1} \\ & \cdot C_k P_{k|k-1} A_k^{T} - A_k P_{k|k-1} C_k^{T} (C_k P_{k|k-1} C_k^{T} + R_k)^{-1} S_k^{T} \\ & - S_k (C_k P_{k|k-1} C_k^{T} + R_k)^{-1} S_k^{T} + Q_k, \end{aligned} \tag{S16}$$

$$\begin{aligned} \hat{x}_{k+1|k+1} = {} & \hat{x}_{k+1|k} + P_{k+1|k} C_{k+1}^{T} (C_{k+1} P_{k+1|k} C_{k+1}^{T} + R_{k+1})^{-1} \\ & \cdot (y_{k+1} - C_{k+1} \hat{x}_{k+1|k}), \end{aligned} \tag{S17}$$

$$\begin{aligned} P_{k+1|k+1} = {} & P_{k+1|k} - P_{k+1|k} C_{k+1}^{T} (C_{k+1} P_{k+1|k} C_{k+1}^{T} + R_{k+1})^{-1} \\ & \cdot C_{k+1} P_{k+1|k}, \end{aligned} \tag{S18}$$

where (S15) and (S16) are (38) and (39) with additional terms involving S_k, and (S17) and (S18) are identical to (40) and (41). The OTSF gain (S12) can be written as

$$K_{k+1}^{\mathrm{f}} = P_{k+1|k}^{\mathrm{f}} C_{k+1}^{\mathrm{T}} (C_{k+1} P_{k+1|k}^{\mathrm{f}} C_{k+1}^{\mathrm{T}} + R_{k+1})^{-1}, \tag{S19}$$

and thus (S17) and (S18) can be written as

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1}^{\mathrm{f}} (y_{k+1} - C_{k+1} \hat{x}_{k+1|k}), \tag{S20}$$

$$P_{k+1|k+1} = (I - K_{k+1}^{\mathrm{f}} C_{k+1}) P_{k+1|k}, \tag{S21}$$

which are identical to (43) and (44).

Note that the error-covariance propagation equation is independent of data. Furthermore, (16) and (17) can be written in terms of K_k as

$$\hat{x}_{k+1} = A_k \hat{x}_k + B_k u_k + K_k (y_k - C_k \hat{x}_k), \tag{18}$$

$$P_{k+1} = A_k P_k A_k^{\mathrm{T}} - K_k (C_k P_k A_k^{\mathrm{T}} + S_k^{\mathrm{T}}) + Q_k. \tag{19}$$

OPTIMAL TWO-STEP PREDICTOR

As an alternative but equivalent implementation of the optimal predictor, (16) and (17) can be implemented as the OTSP. For simplicity in this and subsequent sections, consider the case where $S_k = 0$. The case where S_k is nonzero is discussed in "Optimal Two-Step Predictor and Optimal Two-Step Filter With Correlated Disturbance and Sensor Noise." The *assimilation update* is given by

$$x_k^{\mathrm{a}} = x_k^{\mathrm{f}} + P_k^{\mathrm{f}} C_k^{\mathrm{T}} (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1} (y_k - C_k x_k^{\mathrm{f}}), \tag{20}$$

$$P_k^{\mathrm{a}} = P_k^{\mathrm{f}} - P_k^{\mathrm{f}} C_k^{\mathrm{T}} (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1} C_k P_k^{\mathrm{f}}, \tag{21}$$

and the *forecast update* is given by

$$x_{k+1}^{\mathrm{f}} = A_k x_k^{\mathrm{a}} + B_k u_k, \tag{22}$$

$$P_{k+1}^{\mathrm{f}} = A_k P_k^{\mathrm{a}} A_k^{\mathrm{T}} + Q_k. \tag{23}$$

Figure 1 shows a timing diagram for the real-time implementation of the OTSP. By substituting (20) into (22) and (21) into (23), it can be seen that $x_k^{\mathrm{f}} = \hat{x}_k$ and $P_k^{\mathrm{f}} = P_k$. Defining the *OTSP gain*

$$K_k^{\mathrm{f}} \triangleq P_k^{\mathrm{f}} C_k^{\mathrm{T}} (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1}, \tag{24}$$

(20) and (21) can be written as

Authorized licensed use limited to: University of Michigan Library. Downloaded on March 28,2020 at 16:38:43 UTC from IEEE Xplore. Restrictions apply.

$$x_k^{\mathrm{a}} = x_k^{\mathrm{f}} + K_k(y_k - C_k x_k^{\mathrm{f}}), \tag{25}$$

$$P_k^{\mathrm{a}} = (I - K_k C_k) P_k^{\mathrm{f}}. \tag{26}$$

It is interesting to note that the computational requirements of the OTSP are more burdensome than the OOSP, and yet the OTSP and OOSP produce identical state estimates. Therefore, there is no implementation advantage of the OTSP over the OOSP. However, the value of the OTSP is in providing a framework for the OTSF, as discussed next.

OPTIMAL TWO-STEP FILTER

In contrast to the OTSP in (20)–(23), the OTSF is given by the *forecast update*

$$x_{k+1}^{\mathrm{f}} = A_k x_k^{\mathrm{a}} + B_k u_k, \tag{27}$$

$$P_{k+1}^{\mathrm{f}} = A_k P_k^{\mathrm{a}} A_k^{\mathrm{T}} + Q_k, \tag{28}$$

and the *assimilation update*

$$x_{k+1}^{\mathrm{a}} = x_{k+1}^{\mathrm{f}} + P_{k+1}^{\mathrm{f}} C_{k+1}^{\mathrm{T}} (C_{k+1} P_{k+1}^{\mathrm{f}} C_{k+1}^{\mathrm{T}} + R_{k+1})^{-1} \cdot (y_{k+1} - C_{k+1} x_{k+1}^{\mathrm{f}}), \tag{29}$$

$$P_{k+1}^{\mathrm{a}} = P_{k+1}^{\mathrm{f}} - P_{k+1}^{\mathrm{f}} C_{k+1}^{\mathrm{T}} (C_{k+1} P_{k+1}^{\mathrm{f}} C_{k+1}^{\mathrm{T}} + R_{k+1})^{-1} C_{k+1} P_{k+1}^{\mathrm{f}}. \tag{30}$$

Note that (20) and (21) are slightly different from (29) and (30); specifically, (20) and (21) use the measurement y_k, whereas (29) and (30) use the measurement y_{k+1}. Moreover, the forecast and assimilation updates of the OTSF appear in reverse order compared to the forecast and assimilation updates of the OTSP. This reversal explains the index k in (20) and (21) and the index $k+1$ in (29) and (30).

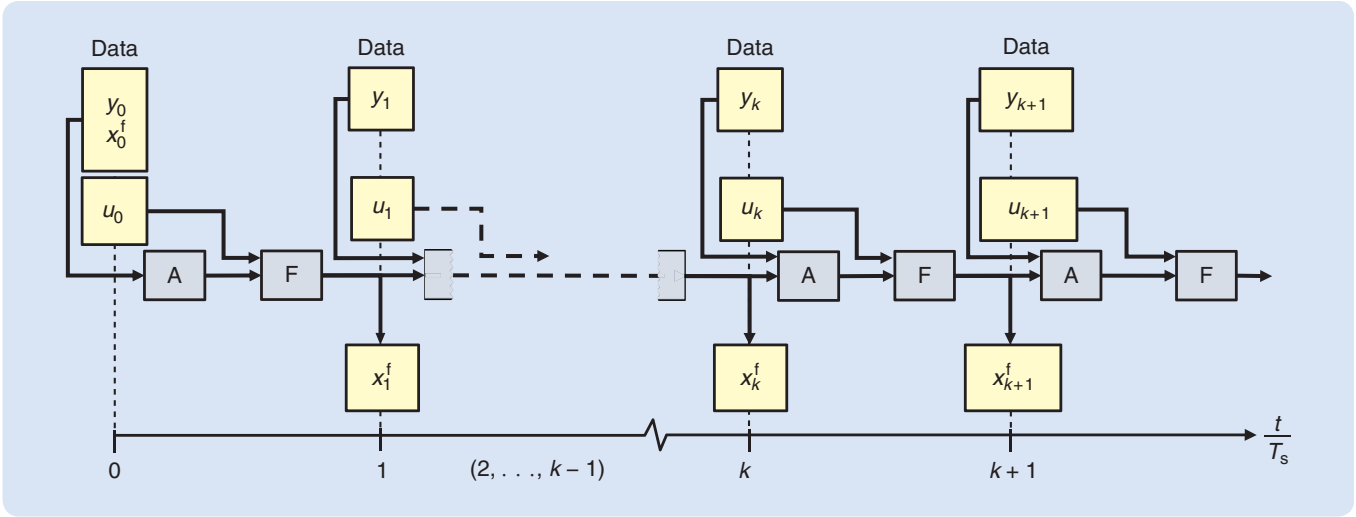


FIGURE 1 The timing diagram for the optimal two-step predictor (OTSP), where A denotes an assimilation update and F denotes a forecast update. The measurement y_k and applied control u_k are available at step k for computation. For all $k \geq 1$, the OTSP produces an estimate of x_k at step k without latency.

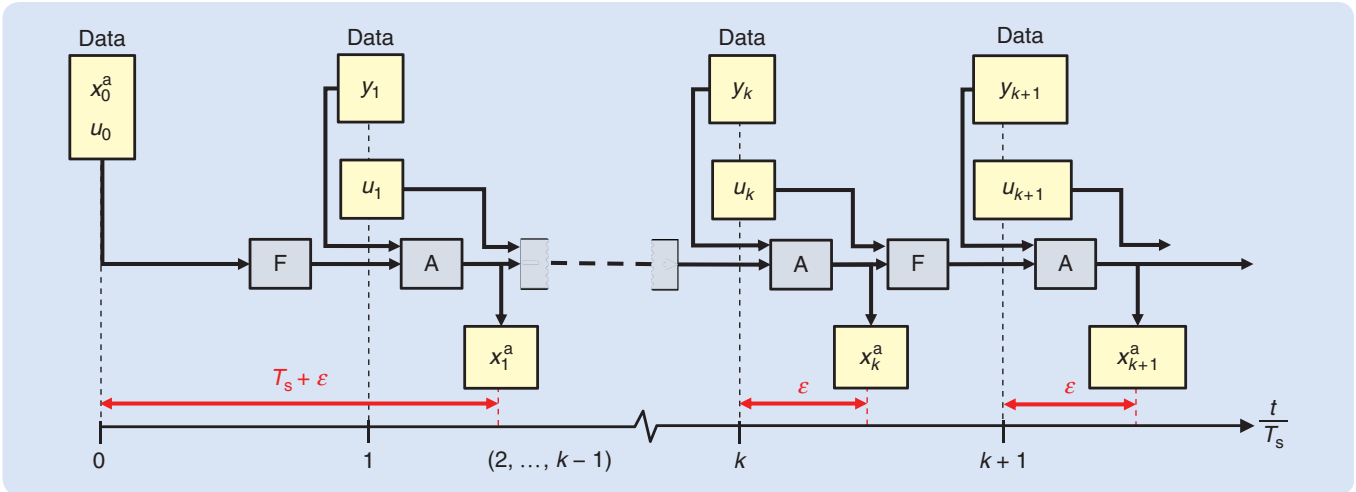


FIGURE 2 The timing diagram for the optimal two-step filter (OTSF), where A denotes an assimilation update and F denotes a forecast update. The measurement y_k and applied control u_k are available at step k for computation. For all $k \geq 1$, the OTSF produces an estimate of x_k with latency ε.

Authorized licensed use limited to: University of Michigan Library. Downloaded on March 28,2020 at 16:38:43 UTC from IEEE Xplore. Restrictions apply.

Next, note that, since x^{a}_{k+1} depends on y_{k+1} and the required computation cannot be performed instantaneously, the estimate x^{a}_{k+1} of x_{k+1} is not available at step $k+1$ but at some time afterward, where the latency depends on the computer speed and architecture. Therefore, if the latency of x^{a}_k as an estimate of x_k is critical in a real-time application, the OTSP may be a better choice than the OTSF. Figure 2 provides a timing diagram for the real-time implementation of the OTSF.

In terms of the *OTSF gain*

$$K^{\mathrm{f}}_{k+1} \triangleq P^{\mathrm{f}}_{k+1}C^{\mathrm{T}}_{k+1}(C_{k+1}P^{\mathrm{f}}_{k+1}C^{\mathrm{T}}_{k+1}+R_{k+1})^{-1}, \tag{31}$$

(29) and (30) can be written as

$$x^{\mathrm{a}}_{k+1} = x^{\mathrm{f}}_{k+1} + K_{k+1}(y_{k+1} - C_{k+1}x^{\mathrm{f}}_{k+1}), \tag{32}$$

$$P^{\mathrm{a}}_{k+1} = (I - K_{k+1}C_{k+1})P^{\mathrm{f}}_{k+1}. \tag{33}$$

Note that, aside from a shift in the index, the OTSF gain is identical to the OTSP gain.

OPTIMAL ONE-STEP FILTER

The OOSF can be obtained by substituting (27) and (28) into (29) and (30). Specifically, the OOSF is given by

$$\begin{aligned} x^{\mathrm{a}}_{k+1} = {} & A_k x^{\mathrm{a}}_k + B_k u_k + (A_k P^{\mathrm{a}}_k A^{\mathrm{T}}_k + Q_k)C^{\mathrm{T}}_{k+1} \\ & \cdot [C_{k+1}(A_k P^{\mathrm{a}}_k A^{\mathrm{T}}_k + Q_k)C^{\mathrm{T}}_{k+1} + R_{k+1}]^{-1} \\ & \cdot [y_{k+1} - C_{k+1}(A_k x^{\mathrm{a}}_k + B_k u_k)], \end{aligned} \tag{34}$$

$$\begin{aligned} P^{\mathrm{a}}_{k+1} = {} & A_k P^{\mathrm{a}}_k A^{\mathrm{T}}_k + Q_k - (A_k P^{\mathrm{a}}_k A^{\mathrm{T}}_k + Q_k) \\ & \cdot C^{\mathrm{T}}_{k+1}[C_{k+1}(A_k P^{\mathrm{a}}_k A^{\mathrm{T}}_k + Q_k)C^{\mathrm{T}}_{k+1} + R_{k+1}]^{-1} \\ & \cdot C_{k+1}(A_k P^{\mathrm{a}}_k A^{\mathrm{T}}_k + Q_k). \end{aligned} \tag{35}$$

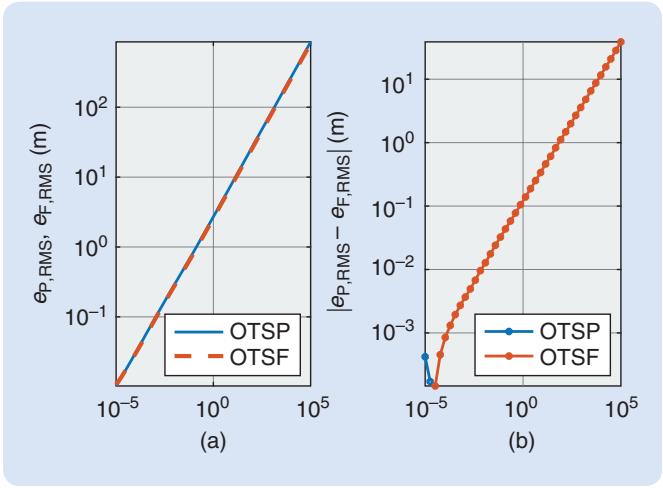


FIGURE 3 The root-mean-square (RMS) position-estimation errors for the optimal two-step predictor (OTSP) and optimal two-step filter (OTSF) versus the disturbance-covariance scaling α_{d}. Part (a) shows the RMS position-estimation errors for the OTSP and OTSF, $e_{\mathrm{P,RMS}}$ and $e_{\mathrm{F,RMS}}$, respectively. Part (b) shows $|e_{\mathrm{P,RMS}} - e_{\mathrm{F,RMS}}|$. In all cases where $e_{\mathrm{P,RMS}} > e_{\mathrm{F,RMS}}$, the OTSF is more accurate than the OTSP; these are shown in red. In all cases where $e_{\mathrm{P,RMS}} < e_{\mathrm{F,RMS}}$, the OTSP is more accurate than the OTSF; these are shown in blue. Note that the OTSP is more accurate than the OTSF for $\alpha_{\mathrm{d}} < 10^{-4.5}$, whereas the OTSF is more accurate than the OTSP for $\alpha_{\mathrm{d}} > 10^{-4.5}$.

OPTIMAL TWO-STEP FILTER AND TRADITIONAL KALMAN FILTER

We now restate the OTSF in standard notation to show that it is precisely the Kalman filter. Defining

$$\hat{x}_{k|k-1} \triangleq x^{\mathrm{f}}_k, \quad P_{k|k-1} \triangleq P^{\mathrm{f}}_k, \tag{36}$$

$$\hat{x}_{k|k} \triangleq x^{\mathrm{a}}_k, \quad P_{k|k} \triangleq P^{\mathrm{a}}_k, \tag{37}$$

(27)–(30) can be written as

$$\hat{x}_{k+1|k} = A_k\hat{x}_{k|k} + B_k u_k, \tag{38}$$

$$P_{k+1|k} = A_k P_{k|k} A^{\mathrm{T}}_k + Q_k, \tag{39}$$

$$\begin{aligned} \hat{x}_{k+1|k+1} = {} & \hat{x}_{k+1|k} + P_{k+1|k}C^{\mathrm{T}}_{k+1}(C_{k+1}P_{k+1|k}C^{\mathrm{T}}_{k+1} + R_{k+1})^{-1} \\ & \cdot (y_{k+1} - C_{k+1}\hat{x}_{k+1|k}), \end{aligned} \tag{40}$$

$$\begin{aligned} P_{k+1|k+1} = {} & P_{k+1|k} - P_{k+1|k}C^{\mathrm{T}}_{k+1}(C_{k+1}P_{k+1|k}C^{\mathrm{T}}_{k+1} + R_{k+1})^{-1} \\ & \cdot C_{k+1}P_{k+1|k}, \end{aligned} \tag{41}$$

which is the standard notation for the Kalman filter [2]. Furthermore, the OTSF gain (31) can be written as

$$K^{\mathrm{f}}_{k+1} = P^{\mathrm{f}}_{k+1|k}C^{\mathrm{T}}_{k+1}\left(C_{k+1}P^{\mathrm{f}}_{k+1|k}C^{\mathrm{T}}_{k+1} + R_{k+1}\right)^{-1}, \tag{42}$$

and thus (40) and (41) can be written as

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K^{\mathrm{f}}_{k+1}(y_{k+1} - C_{k+1}\hat{x}_{k+1|k}), \tag{43}$$

$$P_{k+1|k+1} = (I - K^{\mathrm{f}}_{k+1}C_{k+1})P_{k+1|k}. \tag{44}$$

An alternative notation [1], [3], [4] is given by

$$\hat{x}^{+}_k \triangleq x^{\mathrm{f}}_k, \quad P^{+}_k \triangleq P^{\mathrm{f}}_k, \tag{45}$$

$$\hat{x}^{-}_k \triangleq x^{\mathrm{a}}_k, \quad P^{-}_k \triangleq P^{\mathrm{a}}_k. \tag{46}$$

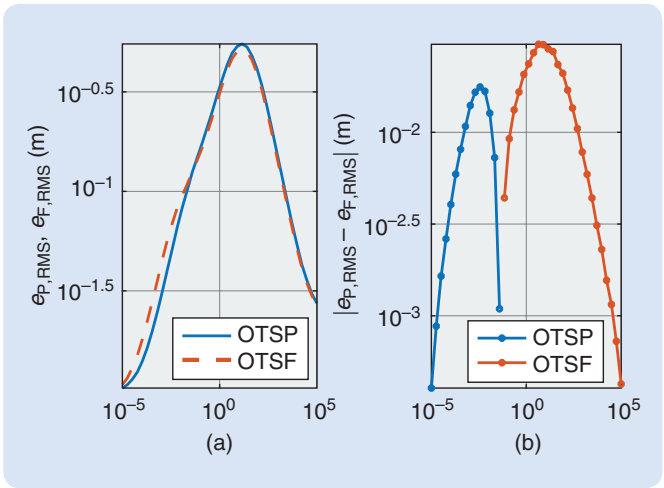


FIGURE 4 The root-mean-square (RMS) position-estimation errors for the optimal two-step predictor (OTSP) and optimal two-step filter (OTSF) versus the sensor-noise covariance α_{sn}. Part (a) shows the RMS position-estimation errors for the OTSP and OTSF, $e_{\mathrm{P,RMS}}$ and $e_{\mathrm{F,RMS}}$, respectively. Part (b) shows that the OTSP is more accurate than the OTSF for $\alpha_{\mathrm{sn}} < 10^{-2}$, whereas the OTSF is more accurate than the OTSP for $\alpha_{\mathrm{sn}} > 10^{-2}$.

Authorized licensed use limited to: University of Michigan Library. Downloaded on March 28,2020 at 16:38:43 UTC from IEEE Xplore. Restrictions apply.

OPTIMAL TWO-STEP PREDICTOR VERSUS OPTIMAL TWO-STEP FILTER

In this section, we investigate whether or not the OTSF estimates are more accurate than the OTSP estimates. Consider an undamped oscillator with a mass of 4 kg and stiffness of 2 N/m, sampled at 4 Hz using a zero-order hold input. The velocity is measured, and the position is to be estimated. The sampled-data system representing this arrangement has the dynamics

$$\begin{bmatrix} x_{k+1} \\ \dot{x}_{k+1} \end{bmatrix} = \begin{bmatrix} 0.9844 & 0.2487 \\ -0.1243 & 0.9844 \end{bmatrix} \begin{bmatrix} x_k \\ \dot{x}_k \end{bmatrix} + \begin{bmatrix} 0.0078 \\ 0.0622 \end{bmatrix} u_k + w_k, \tag{47}$$

$$y_k = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} x_k \\ \dot{x}_k \end{bmatrix} + v_k, \tag{48}$$

where $x_k \in \mathbb{R}$ is the position in meters, $y_k = \dot{x}_k \in \mathbb{R}$ is the velocity in meters/second, u_k is the applied force in Newtons, $w_k \in \mathbb{R}^2$ is the disturbance, and $v_k \in \mathbb{R}$ is the sensor noise.

To investigate the effect of the disturbance w_k, sensor noise v_k, and initial condition x_0 on the estimates of the position produced by the OTSP and OTSF, let

$$u_k \sim \mathcal{N}(0,1),\ w_k \sim \mathcal{N}(0, \alpha_{\mathrm{d}} I),\ v_k \sim \mathcal{N}(0, \alpha_{\mathrm{sn}}),\ x_0 \sim \mathcal{N}(0, \alpha_{\mathrm{ic}} I), \tag{49}$$

where $\alpha_{\mathrm{d}}, \alpha_{\mathrm{sn}}, \alpha_{\mathrm{ic}}$ are varied individually, with the remaining variables fixed. Specifically, we first vary $\alpha_{\mathrm{d}} \in [10^{-5}, 10^{5}]$, with $\alpha_{\mathrm{sn}} = \alpha_{\mathrm{ic}} = 10^{-5}$ fixed. Next, we vary $\alpha_{\mathrm{sn}} \in [10^{-5}, 10^{5}]$, with $\alpha_{\mathrm{d}} = \alpha_{\mathrm{ic}} = 10^{-5}$ fixed. Finally, we vary $\alpha_{\mathrm{ic}} \in [10^{-5}, 10]$, with $\alpha_{\mathrm{d}} = \alpha_{\mathrm{sn}} = 10^{-5}$ fixed. In all cases, the initial states of the OTSP and OTSF are set to zero and $P_0^{\mathrm{f}} = I_2$, $P_0^{\mathrm{a}} = I_2$.

In each case and for each choice of $\alpha_{\mathrm{d}}, \alpha_{\mathrm{sn}}, \alpha_{\mathrm{ic}}$, 10,000 simulations are run with randomly generated values of u_k, w_k, and v_k and the initial condition x_0. For each simulation, the root mean square (RMS) position-estimation errors for the OTSP and OTSF are computed for $0 \le k \le 80$. The RMS position-estimation errors for the OTSP and OTSF are then averaged across the 10,000 simulations. The RMS position-estimation errors for the OTSP and OTSF are defined to be $e_{\mathrm{P,RMS}}(\alpha_{\mathrm{d}}, \alpha_{\mathrm{sn}}, \alpha_{\mathrm{ic}})$ and $e_{\mathrm{F,RMS}}(\alpha_{\mathrm{d}}, \alpha_{\mathrm{sn}}, \alpha_{\mathrm{ic}})$, respectively. Specifically, we compute

$$e_{\mathrm{P,RMS}} = \frac{1}{10{,}000} \sum_{j=1}^{10{,}000} \sqrt{\frac{1}{80} \sum_{i=1}^{80} (x_{i,j}^{\mathrm{f}} - x_{i,j})^2}, \tag{50}$$

$$e_{\mathrm{F,RMS}} = \frac{1}{10{,}000} \sum_{j=1}^{10{,}000} \sqrt{\frac{1}{80} \sum_{i=1}^{80} (x_{i,j}^{\mathrm{a}} - x_{i,j})^2}, \tag{51}$$

where $x_{i,j}^{\mathrm{f}}$ is the position estimate produced by the OTSP at the ith step for the jth simulation, $x_{i,j}^{\mathrm{a}}$ is the position estimate produced by the OTSF at the ith step for the jth simulation,

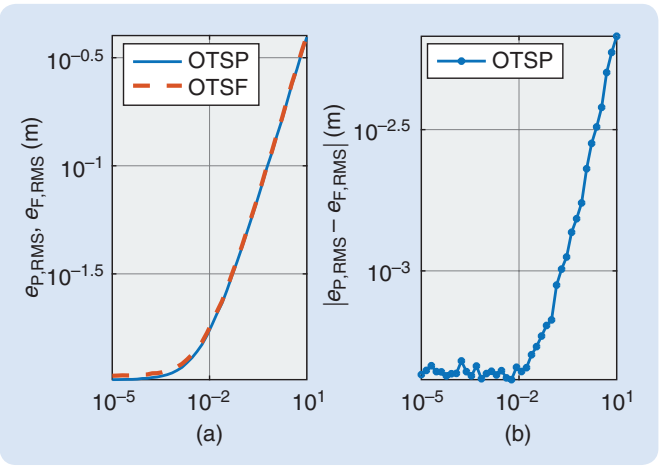


FIGURE 5 The root-mean-square (RMS) position-estimation errors for the optimal two-step predictor (OTSP) and optimal two-step filter (OTSF) versus the initial-condition-covariance scaling, α_{ic}. Part (a) shows the RMS position-estimation errors for the OTSP and OTSF, $e_{\mathrm{P,RMS}}$ and $e_{\mathrm{F,RMS}}$, respectively. Part (b) shows that the OTSP is uniformly more accurate than the OTSF.

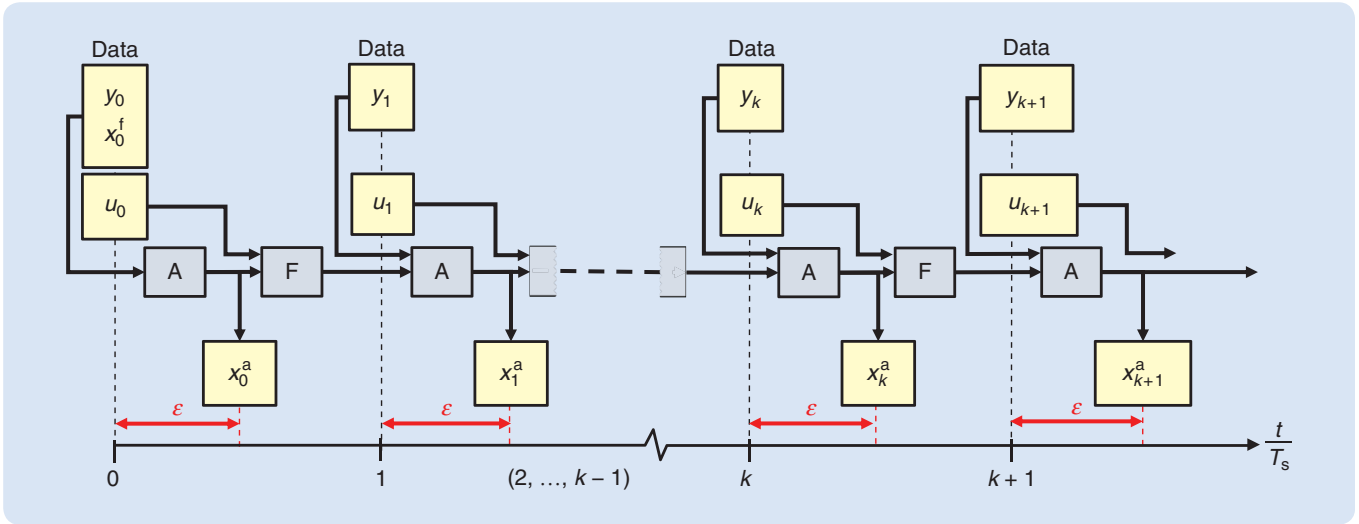


FIGURE 6 The timing diagram for the optimal two-step filter with start-up (OTSFSU), where A denotes an assimilation update and F denotes a forecast update. The measurement y_k and applied control u_k are available at step k for computation. For all $k \ge 0$, the OTSFSU produces an estimate of x_k with latency ε.

Authorized licensed use limited to: University of Michigan Library. Downloaded on March 28,2020 at 16:38:43 UTC from IEEE Xplore. Restrictions apply.

The filter can be realized only as a two-step algorithm, whereas the predictor can be realized as either a one- or two-step algorithm.

and $x_{i,j}$ is the true position at the ith step for the jth simulation. Furthermore, $|e_{P,RMS} - e_{F,RMS}|$ is plotted to show the accuracy of the OTSF position estimate relative to the OTSP position estimate in meters. Specifically, $e_{P,RMS} - e_{F,RMS}$ is positive in the case where the OTSF RMS position-estimation error is smaller than the OTSP RMS position-estimation error and vice versa.

For the three cases considered above, Figures 3(a), 4(a), and 5(a) show $e_{P,RMS}, e_{F,RMS}$ versus $\alpha_d, \alpha_{sn}, \alpha_{ic}$, respectively. Figures 3(b), 4(b), and 5(b) graph $|e_{P,RMS} - e_{F,RMS}|$ versus $\alpha_d, \alpha_{sn}, \alpha_{ic}$, respectively, where the values of $|e_{P,RMS} - e_{F,RMS}|$ are color coded based on the sign of $e_{P,RMS} - e_{F,RMS}$. Figure 5 shows that the OTSP position estimate is more accurate than the OTSF position estimate for all choices of α_{ic}. This situation is surprising since we expect the latency of the OTSF estimates to be offset by greater accuracy. We thus seek a variation of the OTSF that produces the expected improved accuracy in return for latency, as discussed in the next section.

TABLE 1 The initial estimate $\hat{x}_0$ and real-time data y and u used by the estimator to approximate x_k. The time at which the estimate of x_k becomes available is given in terms of the step k, sample time T_s, and latency ε.

Estimator	$\hat{x}_0$	y Data	u Data	$\hat{x}_k$	When Available?
OTSP	x_0^f	y_{k-1}	u_{k-1}	x_k^f	kT_s
OTSF	x_0^a	y_k	u_{k-1}	x_k^a	$kT_s + \varepsilon$
OTSFSU	x_0^f	y_k	u_{k-1}	x_k^a	$kT_s + \varepsilon$

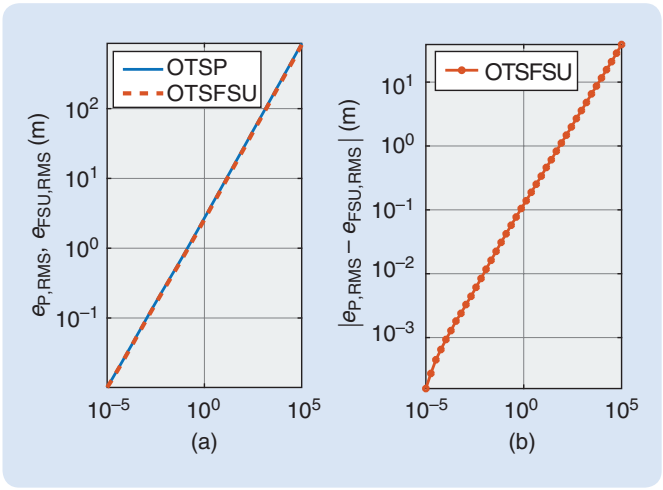


FIGURE 7 The root-mean-square (RMS) position-estimation errors for the optimal two-step predictor (OTSP) and optimal two-step filter with start-up (OTSFSU) versus the disturbance-covariance scaling, α_d. Part (a) shows the RMS position-estimation errors for the OTSP and OTSFSU $e_{P,RMS}$, and $e_{FSU,RMS}$, respectively. Part (b) shows that the OTSFSU is uniformly more accurate than the OTSP.

OPTIMAL TWO-STEP FILTER WITH START-UP

As shown in Figure 5, the OTSP may be more accurate than the OTSF. This is surprising because the OTSF produces latency estimates and uses more recent data, leading to the expectation of more accurate state estimates. This phenomenon can be traced to the fact that OTSF does not use the measurement y_0. In fact, at step $k = 0$ (20) uses y_0, whereas at step $k = 0$ (29) does not use y_0. This situation suggests the possibility of a variant of the OTSF that employs an additional assimilation update before the initial forecast update; otherwise, all subsequent updates of the OTSFSU are identical to the OTSF. Figure 6 presents a timing diagram for the real-time implementation of the OTSFSU. The next section investigates the accuracy of the OTSFSU compared to the OTSP to determine if the OTSFSU is more accurate, as compensation for its inherent latency.

REAL-TIME IMPLEMENTATION

To clarify the timing of the computation required for the OTSP, OTSF, and OTSFSU, note that for the OTSP the state estimate is x_k^f, whereas for the OTSF and OTSFSU the state estimate is x_k^a. Figures 1, 2, and 6 show timing diagrams for the real-time implementation of the OTSP, OTSF, and OTSFSU. Note that the OTSP is the only estimator that produces an estimate of x_k at step k. In contrast, the OTSF and OTSFSU produce estimates of x_k at time $kT_s + \varepsilon$, where T_s is the sample time of the digital implementation and $\varepsilon \in (0, T_s)$ is the latency (which is platform dependent). As pictured in Figure 2, the first state estimate produced by the OTSF is an estimate of x_1; this estimate is not available until $t = T_s + \varepsilon$. In addition, this estimate of x_1 does not use the data y_0, as can be seen by the absence of y_0 in Figure 2. Alternately, as displayed in Figure 6, the OTSFSU uses y_0 to produce an estimate of x_0 at $t = \varepsilon$. Table 1 lists the data used by each algorithm and the associated latency.

OPTIMAL TWO-STEP PREDICTOR VERSUS OPTIMAL TWO-STEP FILTER WITH START-UP

We now compare the OTSP and OTSFSU using the same procedure employed to compare the OTSP and OTSF. For the OTSFSU, define $e_{FSU,RMS}(\alpha_d, \alpha_{sn}, \alpha_{ic})$, and plot $|e_{P,RMS} - e_{FSU,RMS}|$. Specifically, compute

Authorized licensed use limited to: University of Michigan Library. Downloaded on March 28,2020 at 16:38:43 UTC from IEEE Xplore. Restrictions apply.

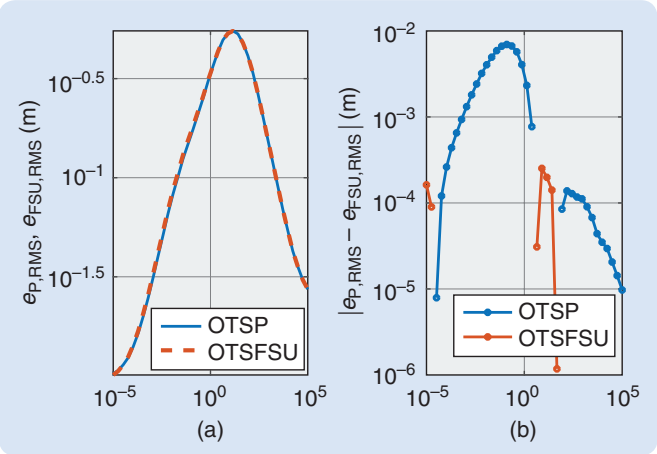


FIGURE 8 The root-mean-square (RMS) position-estimation errors for the optimal two-step predictor (OTSP) and optimal two-step filter with start-up (OTSFSU) versus the sensor-noise-covariance scaling, α_{sn}. Part (a) shows the RMS position-estimation errors for the OTSP and OTSFSU, $e_{P,RMS}$ and $e_{FSU,RMS}$, respectively. Part (b) shows that the OTSP is more accurate than the OTSFSU for $10^{-4.5} < \alpha_{sn} < 10^{-2}$ and $\alpha_{sn} > 10^{2}$.

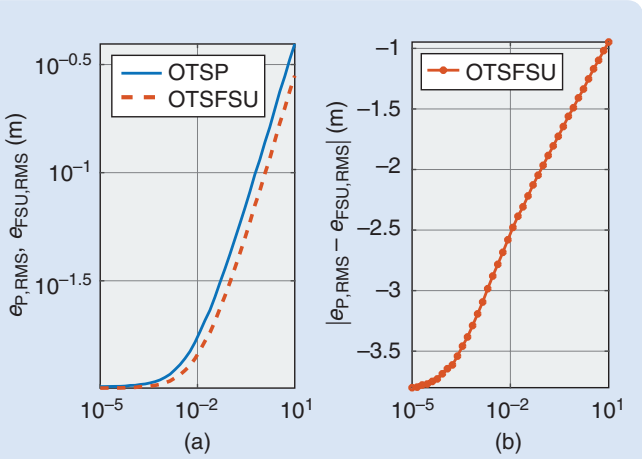


FIGURE 9 The root-mean-square (RMS) position-estimation errors for the optimal two-step predictor (OTSP) and optimal two-step filter with start-up (OTSFSU) versus the initial-condition-covariance scaling, α_{ic}. Part (a) shows the RMS position-estimation errors for the OTSP and OTSFSU, $e_{P,RMS}$ and $e_{FSU,RMS}$, respectively. Part (b) shows that the OTSFSU is uniformly more accurate than the OTSP.

$$e_{FSU,RMS} = \frac{1}{10,000} \sum_{j=1}^{10,000} \sqrt{\frac{1}{80} \sum_{i=1}^{80} (x_{i,j}^{a} - x_{i,j})^2}, \tag{52}$$

where $x_{i,j}^{a}$ is the position estimate produced by OTSFSU at the ith step for the jth simulation, and $x_{i,j}$ is the true position at the ith step for the jth simulation. Figures 7–9 show that the OTSFSU position estimate is more accurate than the OTSP position estimate for all choices of α_d and α_{ic}. Furthermore, as displayed in Figure 8, the OTSP shows no distinct advantage over the OTSFSU for variations of α_{sn}.

CONCLUSIONS

The goal of this article was to untangle the relationship between the OTSP and OTSF. A fundamental distinction between the filter and the predictor is that the filter can be realized only as a two-step algorithm, whereas the predictor can be realized as either a one- or two-step algorithm. An analysis by a timing diagram shows that the estimates produced by the OTSF take advantage of more recent data but have latency due to the time needed for computation. Partially contradicting the expected latency/accuracy tradeoff, a numerical example showed that the OTSF is not uniformly better than the OTSF. This discrepancy was traced to the nonuse of initial data by the OTSF, which suggested a variation of the OTSF that implements an additional assimilation update at start-up; otherwise the OTSF and the OTSFSU are identical. Revisiting the numerical example showed that the OTSFSU is uniformly more accurate than the OTSP. As with the OTSF, this additional accuracy comes at the price of latency.

ACKNOWLEDGMENTS

This research was partially supported by the Air Force Office of Scientific Research under Dynamic Data-Driven Applications Systems grant FA9550-16-1-0071 (http://www.1dddas.org/).

AUTHOR INFORMATION

Syed Aseem Ul Islam (aseemisl@umich.edu) received the B.S. degree in aerospace engineering from the Institute of Space Technology, Islamabad, Pakistan, and he is pursuing the Ph.D. degree in flight dynamics and control from the University of Michigan, Ann Arbor. His research interests include data-driven adaptive control for aerospace applications.

Ankit Goel received the B.E. degree in mechanical engineering from the Delhi College of Engineering, India, and the Ph.D. degree from the University of Michigan, Ann Arbor. His research interests include the data-driven estimation and control of high-dimensional complex systems.

Dennis S. Bernstein received the Sc.B. degree from Brown University, Providence, Rhode Island, and the Ph.D. degree from the University of Michigan, Ann Arbor, where he is a professor in the Aerospace Engineering Department. His research interests include the identification, estimation, and control for aerospace applications. He is the author of *Scalar, Vector, and Matrix Mathematics*, published by Princeton University Press.

REFERENCES

[1] A. Jazwinski, *Stochastic Processes and Filtering Theory*. New York: Elsevier, 1970.

[2] B. Anderson and J. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice Hall, 1979.

[3] J. Crassidis and J. Junkins, *Optimal Estimation of Dynamic Systems*. Boca Raton, FL: CRC Press, 2004.

[4] D. Simon, *Optimal State Estimation: Kalman, H-Infinity, and Nonlinear Approaches*. Hoboken, NJ: Wiley, 2006.

[5] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, 1960. doi: 10.1115/1.3662552.

[6] B. Teixeira, "Kalman filters," *IEEE Control Syst. Mag.*, vol. 28, no. 2, pp. 16–18, Apr. 2008. doi: 10.1109/MCS.2008.4472371.

Authorized licensed use limited to: University of Michigan Library. Downloaded on March 28,2020 at 16:38:43 UTC from IEEE Xplore. Restrictions apply.

The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/1477-7835.htm

Selection of sustainable transport system: a case study

Abhishek Singh
Department of Civil and Environmental Engineering, Delhi Technological University, Delhi, India
Amulya Gurtu
Cofrin School of Business, University of Wisconsin Green Bay, Green Bay, Wisconsin, USA, and
Rajesh Kumar Singh
Management Development Institute, Gurgaon, India

Received 30 March 2020
Revised 28 June 2020
Accepted 28 June 2020

Abstract

Purpose – Sustainable mobility will be the key to the survival of mankind in the 21st century. Cities with debt-ridden and poorly managed transport systems have to change to the ever-growing demands of the public transport system. The low cost of transport has been a key factor in sustainable development for any city. This study is trying to propose a framework for the selection of sustainable transport in context to an Indian case of NCR Delhi, India.

Design/methodology/approach – The present study has identified eight criteria for the selection of a sustainable transport system. Criteria for selection of sustainable transport are CO_2 emissions, cost of fuel, energy efficiency, cost of maintenance, number of accidents, congestion, number of injuries and road noise. Three alternatives of transport considered for this study are state-run bus, pooled car and *Shuttl* (App-based buses). The analytic hierarchy process (AHP) is used to prioritize the criteria. The study has further illustrated a framework for the selection of sustainable transport based on these criteria.

Findings – It is observed that CO_2 emissions are the most important criterion for a sustainable transport system. It is followed by a reduction in congestion and the number of injuries. Three alternatives of transport considered for this study are state-run bus, pooled car and *Shuttl* (a service provider). *Shuttl* has been found to be the most sustainable transport system.

Research limitations/implications – AHP is not able to capture the vagueness in decision-making. Therefore, fuzzy AHP can be considered for further detailed analysis as future scope of study in a different context of a sustainable transportation system. Major implications for policymakers and stakeholders are that development of public transport in cities should be done after considering different dimensions of sustainable operations.

Originality/value – The study has proposed a unique framework for the selection of a sustainable transport system by the public based on sustainability criteria. Findings will help policymakers in formulating strategies for developing sustainable transport system.

Keywords Transport system, City commute, Mobility, Sustainability, Analytic hierarchy process (AHP)

Paper type Research paper

1. Introduction

Transportation, besides a growing population and industrialization, increased the pollution significantly (Nakamura, 2004). Therefore, transportation systems need to be environmentally sustainable (Gupta and Singh, 2020). Even manufacturing organizations are trying to use the concept of reverse logistics for ensuring a sustainable transportation system (Agrawal and Singh, 2019). Sustainable transportation meets the requirements of the present without creating problems for future transportation systems (Black, 1996). Usually, a sustainable transport system works on the principle of the triple bottom line (e.g., World

The authors would like to thank reviewers and editor of the journal for valuable comments to improve the quality of this manuscript.

Management of Environmental Quality: An International Journal
© Emerald Publishing Limited
1477-7835
DOI 10.1108/MEQ-03-2020-0059

Bank, 1996; Loo, 2002; Schipper, 2003). The public transportation system plays an essential social function, and it needs to be designed responsibly for stopping global climate change (Haase and Nuissi, 2007). Traditionally, transportation systems focus on performance metrics such as vehicle mobility and congestion, and sustainability measures are not given due weightage (Zheng *et al.*, 2013). Without having an effective performance framework, the selection of a sustainable transportation system is a big challenge (Awasthi *et al.*, 2011). A sustainable transport system must provide mobility and accessibility to all urban residents in an environmentally friendly manner. This is a complex and challenging task when the needs and demands of people belonging to different income groups are not only different but also often conflicting. A mode of transportation is decided by individuals based on economic criteria, trip time involved, convenience, comfort and safety. According to May *et al.* (2001), sustainable transportation should ensure neat and clean streets, safety to the environment and support to a vibrant economy.

Zhao *et al.* (2020) have observed that sustainable transportation-related issues have emerged as an important topic of research. For a sustainable transportation system, long-term strategies need to be developed (Singh and Agrawal, 2018). Therefore, research in this area is highly significant for the economic growth of developing countries like India. Vehicles on roads create maximum urban air pollution. Air pollution is a major problem in metro cities. National Capital Region of Delhi, India (NCR), is considered one of the worse affected metro cities in the world due to vehicular pollution. It is being observed that after the lockdown in India due to COVID 19 pandemic in 2020, pollution level has drastically reduced in NCR Delhi and other metro cities. Therefore, controlling road transportation and the selection of optima alternatives by the public may play a vital role in controlling environmental pollution. The population and income growth in Indian cities over the last few decades have a marked effect on the mobility of its citizens. India's transport infrastructure demand has increased by almost eight times since 1980, which is higher than any other economy in Asia. This growth has been positive in many ways, including the development of thriving auto industry, growth of public infrastructure and allied economic growth. Fifteen of the top 20 polluted cities in the world are from India (Hindu, 2019).

Additionally, many cities in India are also ranked amongst the most congested cities in the world. The average speed for vehicles in some metro cities is reported as low as 17 km/h (10.6 m.p.h.). Congestion results in the reduction of productivity, fuel waste and pollution. A high-level estimate of the economic loss due to congestion, for the top four metros of India, is over USD 22 billion annually. Multiple efforts at the city, state and national levels have been taken to alleviate traffic congestion to improve productivity and to reduce pollution.

As an example, the Pradhan Mantri Gram Sadak Yojana (PMGSY) is one of the best examples to make the commute between rural and urban areas of India better. The emphasis on this scheme in the last three years has led to a record construction of all-weather roads at a rate of 130 km/day (81.3 miles/day). To address the twin challenges of pollution and congestion in urban India, several initiatives have been launched in various cities. Further, multiple national policy initiatives, such as the National Urban Transport Policy and the National Electric Mobility Mission Plan 2020, have been designed for improving mobility. While there are multiple examples of such initiatives, the magnitude and complexity of our mobility challenges require a comprehensive action. If India continues to grow with present carbon emission average rate of 6.8%, which was the trend between 2006 and 2015, it will supersede the European Union's total emissions by 2020 (EC-JRC report, 2016). Delhi, Gurgaon, Noida, Faridabad and Ghaziabad constitute the NCR of Delhi, India. NCR is a big and extended city with a population of about 46 million. Almost 10 million people commute to and from neighboring cities daily for different purposes, making it even more populated. The city roads are congested and do not serve the purpose of mobility for the increasing population. The distances between the office complexes and residential complexes are 20–

50 km (12–31 miles) in NCR, and most of the people prefer driving to work. Although Metro has reduced the problem to some extent, it does not provide last-mile connectivity. The government is trying to develop different alternatives of transport in metro cities, but sustainability is yet not the biggest concern for policymakers. Therefore, people use road transport choosing from the limited options like state-run buses, carpools or *Shuttl* (App-based office buses). The selection of sustainable transport by the public is a challenge, and it requires awareness also. This study identified the criteria for sustainable transport using analytical hierarchy process (AHP) and suggests a framework for the selection of the most sustainable transportation system in NCR Delhi, India. The AHP methodology and why it was used are explained in Section 3 (Research Methodology). The remainder of the paper has been organized as follows. Section 2 provides the literature review, Section 3 provides the research methodology, Section 4 shares the results and discusses them and, finally, Section 5 concludes the discussion with some limitations and possible future research.

2. Literature review

The primary objective of transportation is to move people, equipment and supplies safely and efficiently from one place to another (Gurtu *et al.*, 2019). The transport sector is the biggest contributor of the emissions in the USA and accounted for 29% of all emissions in 2017. Further, light-duty vehicles, which include passenger vehicles, caused 59% of emissions in the USA in 2017 (US EPA, 2020). Therefore, it is necessary to pay attention to sustainability in transportation. The concept of sustainable transport stemmed from sustainable development (Hall, 2002). Stefanieca *et al.* (2020) assessed the sustainability of inland transportation in China. Sustainable transport has two definitions. The first focuses on environmental problems and resource depletion, while the second encompasses social and economic welfare. The latter is preferred as it allows us to take a comprehensive view of all the effects of the transport sector (Litman and Burwell, 2006) and encourages people to seek integrated solutions to sustainable transport (Zhou, 2012). A sustainable transportation system is less costly to society. It should meet the transportation needs of all the sections of society. A sustainable transport system leads to sustainable development. According to Daly (1991), for sustainable development, the use of renewable and non-renewable resources should be less than its regeneration. "Another corollary to this definition is that the rate of generating pollution (to air, water, and land) must be kept the same or less than the time required to absorb and completely disintegrate that pollution in the environment" (Gurtu, 2014). In other words, emissions should be assimilated into the environment Mohan and Tiwari (1999). The transportation sector accounts for approximately 25% of worldwide carbon dioxide emissions (Van den Berg and De Langen, 2017).

In metro cities, most people use their own cars (Cosgrove *et al.*, 2009), and it causes traffic congestion (Arnott and Small, 1994). Transport sector contributes significantly to CO_2 emissions and climate change, but serious efforts have not been made by policymakers to control this problem (Marsden and Rye, 2010). Cruz and Katz-Gerro (2016) have tried to analyze few cases of transport companies and found that selection of green technologies and green practices by transport companies is influenced by government policies and economic constraints. Therefore, effective government policies and support for public transport may help in sustainable growth of cities. Vicente *et al.* (2020) have tried to analyze factors influencing passenger loyalty toward public transport services by using structural equation modeling approach. They have observed that public transport providers' commitment to environmental sustainability does have a direct positive effect on the passenger loyalty. Other researchers have also found in their studies that public prefer to choose environment-friendly solutions even if they cost more (European Commission, 2008; Kennell, 2016). According to Kennell (2016), investment in green operations may have more ROI if properly

communicated to public about its green initiatives. Currie *et al.* (2018) have tried to measure sustainability of public transport in terms of social, economic and environmental factors.

Energy is of utmost importance for all functions in supply chain management, logistics and transportation. Sustainable transportation uses energy efficiently. Road freight transport consumes energy significantly (Browne and Allen, 1998). According to Aßmann and Sieber (2005), worldwide transport consumes about 20% of global primary energy. However, according to a more recent estimate, the global transportation sector consumes 25% of all the energy (The Maritime Executive, 2020). Also, oil prices are dropping now. However, due to the high price of oil in the past, the cost of fuel was around 35% of the total costs of carriers (Gurtu, 2019; Thuermer, 2008). Higher fuel cost or lower energy efficiency also impact the costs of goods sold (Tanowitz and Rutchik, 2008).

Disturbance due to sound is known as noise pollution. It depends on the intensity of different sound sources (Müller-Wenk, 2004). Noise level above 70 decibels (dB) can cause many ailments such as high blood pressure, grouchiness and risk of cardiac infarctions. Noise from transport vehicles also causes vibrations to buildings and homes. Usually, vehicles cause two types of noise: engine noise and rolling noise. Proper maintenance of vehicles and the use of high-quality fuel help in reducing noise pollution.

Traffic accidents cause medical emergencies, casualties and emotional and financial impacts on individuals and their families. They impact family income, work productivity and losses to insurance companies (Elvik, 2000). Distracted driving such as texting and talking over the phone is an avoidable cause of road accidents (Henley *et al.*, 2018). Technological innovations in vehicles, compliance with traffic rules and following maintenance guidelines help in reducing traffic accidents.

Congestion index is the most prominent measure for the evaluation of a transportation system (Boarnet *et al.*, 1998; Vaziri, 2002). It tries to compare the time required for traveling during peak and off-peak hours. Due to congestion, traffic jams happen. People reach late to their offices. Work is not completed on time. It overall impacts the performance of all systems. It also causes more CO_2 emissions, and the fuel efficiency reduces. The cost of transportation also increases.

Emissions are increasing exponentially due to the growth of vehicles on the road. One of the reasons for the increase in the number of vehicles is the increase in the earnings of people. This problem can be tackled with the use of better technology and high quality of fuel. Alternative fuels may also be explored. Oil is the most commonly used source of road transport. In place of oil, transportation systems may explore alternative sources of energy such as biofuels, solar energy, electricity and hydrogen (Gurtu *et al.*, 2015, 2016, 2019). Research needs to be done for comparison of emissions from different sources. It will also help in the selection of sustainable fuel for road transport.

Travel behavior and the choice of the mode of transportation have a significant influence on the sustainability of transport systems (Schneider, 2013), but the influence is mediated by various factors, such as transport policies, car ownership, household income, age (Mattioli, 2014), culture, travel distance (Cattaneo *et al.*, 2018) and weather conditions (Kim and Macdonald, 2016). Song *et al.* (2017) attempted to assess the expected environmental costs of accessibility; Colvile *et al.* (2004) attempted to evaluate the road transport sustainability regarding the exposure to air pollution from traffic through modeling the movement of air, vehicles and exhaust emissions; and Salehi *et al.* (2017) developed a method to minimize total carbon emissions taking into consideration speed control. The transportation system should be sustainable in terms of social, economic and environmental measures. For environmental sustainability, fuel efficiency should be high. Emissions should be less harmful to society. For economic sustainability, the cost of transportation should be minimized by optimizing speed, the number of persons in vehicles and the time required for traveling. For social sustainability, the transportation system should ensure safety, social interaction and

availability (Shi and Zhou, 2012). Any sustainable system must also be energy-efficient and must create minimum waste (Reddy *et al.*, 2000; Zheng *et al.*, 2013).

From the above literature review, it can be inferred that a reduction in transport emissions is crucial for environmental sustainability. This will also help to reduce many health-related, particularly respiratory, ailments in areas with heavy traffic. This understanding was also confirmed by the subject-matter experts. Based on this understanding and many rounds of discussions with experts, a total of eight indicators for sustainable transport have been finalized. These are CO_2 emissions, congestion, number of injuries, cost of fuel, number of accidents, cost of maintenance, energy efficiency and road noise.

3. Research methodology

AHP is used in this study for prioritization of criteria of sustainable transport and for evaluating the most sustainable transport. AHP was developed by Saaty (1980) to solve problems, which are complex, and a solution depends on many factors or criteria. Managers use AHP in various strategic and tactical decision-making situations. Singh (2013) has used AHP for prioritization of factors for a coordinated supply chain. Chhabra *et al.* (2017) have used AHP for analyzing alternatives for green logistics. Bhandari *et al.* (2018) have used AHP for justification of advanced manufacturing technologies for small and medium-sized enterprises.

Analytic Hierarchy Process (AHP): AHP helps decision-makers to analyze the relative importance of criteria for effective decision-making (Saaty, 1980). These criteria are compared on the Saaty's (1980) nine-point rating scale (Table 1). The steps used to find the importance of rating of criteria using AHP are discussed as follows:

Step 1. To prepare a pairwise comparison matrix, each criterion is compared with respect to another on the 9-point rating scale.

Step 2. To produce the normalized matrix, each element of each column is divided by the sum of all the elements of the corresponding column.

Step 3. Average value of each row of matrix C' (i.e. the matrix produced in step-2) is calculated.

Step 4. The consistency of the decision matrix is checked by the consistency ratio. The value of maximum eigenvalue (λ_{max}) is required for finding the consistency index (CI) as follows.

(1) Assume that the pairwise comparison matrix is represented by P1 and the principal matrix is represented by P2

Intensity of importance	Definition
1	Equal importance
3	Weak importance of one over another
5	Essential or strong important
7	Demonstrated importance
9	Absolute importance
2, 4, 6, 8	Intermediate values between the two adjacent judgment
Reciprocals of above non-zero	If activity *i* has one of the above non-zero numbers assigned to it when compared with activity *j*, then *j* has the reciprocal value when compared with *i*

Table 1. Tomas Saaty's (1994) nine-point scale

(2) P3 may be defined as the product of P1 and P2, while P4 is defined as P3/P2

(3) let λmax be the average of the elements of P4

Step 5. After finding the maximum eigenvalue of the matrix, the value of CI is calculated as shown in Equation (1),

$$\mathrm{CI} = \frac{\lambda_{\max} - n}{n - 1} \tag{1}$$

Step 6. Finally, the consistency ratio is calculated, as shown in Equation (2).

$$\mathrm{CR} = \frac{\mathrm{CI}}{\mathrm{RI}(n)} \tag{2}$$

where RI (n) is a random index, which varies with the value of n (Table 2). If the value of CR becomes greater than 0.1, then go to step 1and check the consistency of the decisions taken for pairwise comparison. If it is less than 0.1, it means decisions are consistent.

4. Results and discussion

AHP is used to select the most sustainable alternatives for the transportation system in NCR Delhi, India. This section presents the analysis and development of a framework for selecting environment-friendly alternatives.

4.1 Phase 1 – AHP model for analysis

In this case, the goal is to evaluate alternatives in sustainable mobility transportation for NCR Delhi, India. The criteria for the AHP model have been selected based on a literature review (Castillo *et al.*, 2010) and discussion held with industry experts from the urban transportation and environment department. The criteria for selection of alternative are CO_2 emissions, energy efficiency, cost of operation, road noise, number of accidents, number of injuries, congestion and cost of fuel. For this study, authors have considered three transport alternatives. These are state-run buses, car-pool (App-based transport) and *Shuttl*. The formulation of AHP model (Figure 1) consists of a goal, main factors and result (alternatives). The framework has three stages which include the problem at the first stage, criteria as the second stage and the three alternatives – state-run buses, car-pool and *Shuttl* practices – as the third stage. Figure 1 shows the framework for selection of sustainable transport using AHP.

4.2 Phase 2 – Collection and measurement of primary data

After the model of AHP for hierarchy has been formulated, the next phase is about measuring and collecting the primary data. This was done by making a team of five experts and assigning the pairwise comparison to the main factors used in the AHP hierarchy. These experts were chosen and interviewed from different domains of the transportation industry.

Table 2. List of random index values

n	2	3	4	5	6	7	8	9
RI(n)	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45

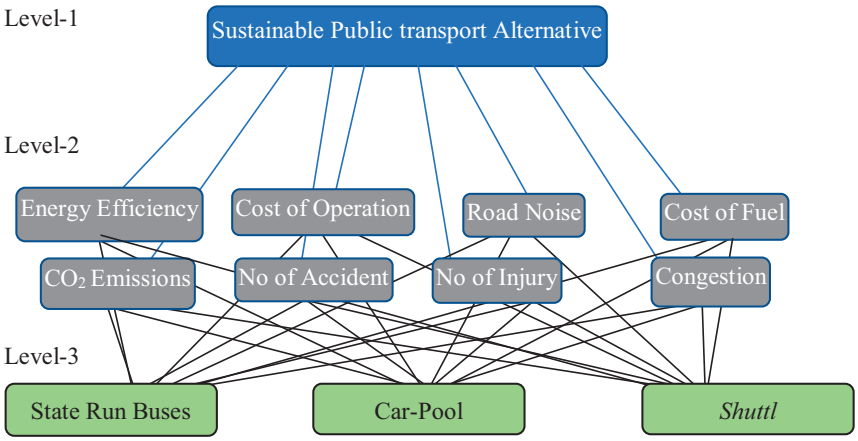


Figure 1.
AHP framework for the selection of sustainable transport alternative

Two of them were fleet managers, and three of the experts were strategists for sustainable transport. The nine-point scale developed by Saaty (1994) for AHP was used to assign relative scores to a pairwise comparison between main factors. The experts were asked to assign a relative score to each comparison using a scale mentioned in Table 2, and the process was continued for all levels of hierarchy. For collecting the primary data from experts, a questionnaire was prepared, which comprised all main factors and was used to obtain the pairwise comparison judgment from the experts. In the case of no consensus, importance was given to the decision which had a majority.

4.3 Phase 3 – formulation of pairwise comparison matrices

In this step, pairwise comparison matrices are constructed for each level of attributes. The governing element for a lower-level element is the higher-level element. Comparison is made between the elements in the lower level to each other. This is done on the basis of their effect on the governing element above it. This gives a square matrix of judgments. The pairwise comparison is made in terms of domination of an element over another element to express the judgments as integers. If element 1 dominates over element 2, then the whole integer is entered in row 1 column 2, and its reciprocal is entered in row 2 column 1. If the compared elements are the same, value 1 is assigned to both positions. Following the above method, pairwise comparison for all criteria shown in level 2 is made, and relative weights for all criteria are found as shown in Table 3. It is observed that CO_2 emission is the most important sustainability criteria, with a relative weight of 0.3258. It is followed by congestion (0.2084), the number of injuries (0.1511), cost of fuel (0.1072), number of accidents (0.0839), cost of maintenance (0.0571), energy efficiency (0.0392) and road noise (0.0273).

Criteria	Criteria weights (priority vector)
CO_2 emissions	0.3258
Cost of fuel	0.1072
Energy efficiency	0.0392
Cost of maintenance	0.0571
Number of accidents	0.0839
Congestion	0.2084
Number of injuries	0.1511
Road noise	0.0273

Table 3.
Weights of different criteria for the selection of sustainable transport

The next phase of the study is to find a solution for the selection of sustainable transport on the basis of the above-mentioned sustainability criteria. First, with respect to each criterion, a pairwise comparison matrix for three selected alternatives is made. The results are shown in Table 4. Values of priority vector in the last column show the relative weight. The importance of different alternatives differs from criteria to criteria. For example, in terms of CO_2 emissions, *Shuttl* is the most important alternative, with a relative weight of 0.5571. In terms of congestion, state-run bus is the most important alternative, with a weight of 0.6333.

Similarly, in terms of other criteria, the most important alternatives have been identified. In Table 5, the relative weights of each criterion and weight for all alternatives are summarized. After getting the relative weights of all alternatives against each sustainability criteria, global weights for each alternative are calculated. For each choice, a set of global priority weight was obtained by multiplying local weights of lower factors with the weights of all higher parent factors and then adding all the multiplied products. The results are given in Table 6. The sum of global weights of an alternative against all criteria gives the global

Table 4.
Pairwise comparison judgment matrices

Normalized pairwise matrix of alternatives for carbon emissions				
Alternatives	State-run bus	Pooled car	*Shuttl*	Priority vector (P.V.)
State-run bus	0.3000	0.3750	0.2857	0.3202
Pooled car	0.1000	0.1250	0.1429	0.1226
Shuttl	0.6000	0.5000	0.5714	0.5571
Normalized pairwise matrix of alternatives for congestion				
State-run bus	0.6522	0.5556	0.6923	0.6333
Pooled car	0.1304	0.1111	0.0769	0.1062
Shuttl	0.2174	0.3333	0.2308	0.2605
Normalized pairwise matrix of alternatives for number of injury				
State-run bus	0.2308	0.2727	0.2258	0.2431
Pooled car	0.0769	0.0909	0.0968	0.0882
Shuttl	0.6923	0.6364	0.6774	0.6687
Normalized pairwise matrix of alternatives for cost of fuel consumed				
State-run bus	0.1667	0.0769	0.2727	0.1721
Pooled car	0.5000	0.2308	0.1818	0.3042
Shuttl	0.3333	0.6923	0.5455	0.5237
Normalized pairwise matrix of alternatives for number of accidents				
State-run bus	0.2308	0.2727	0.2258	0.2431
Pooled car	0.0769	0.0909	0.0968	0.0882
Shuttl	0.6923	0.6364	0.6774	0.6687
Normalized pairwise matrix of alternatives for cost of maintenance				
State-run bus	0.1250	0.0769	0.1818	0.1279
Pooled car	0.5000	0.3077	0.2727	0.3601
Shuttl	0.3750	0.6154	0.5455	0.5119
Normalized pairwise matrix of alternatives for energy efficiency				
State-run bus	0.1429	0.1000	0.1818	0.1416
Pooled car	0.4286	0.3000	0.2727	0.3338
Shuttl	0.4286	0.6000	0.5455	0.5247
Normalized pairwise matrix of alternatives for road noise				
State-run bus	0.2308	0.3750	0.2105	0.2721
Pooled car	0.0769	0.1250	0.1579	0.1199
Shuttl	0.6923	0.5000	0.6316	0.6080

desirability index of the concerned alternative. Global desirability indexes for all alternatives are summarized in Table 7. It is observed that the global desirability index of *Shuttl* is highest, that is, 0.5155. It is followed by state-run buses, with a global desirability index of 0.3322. The last alternative is pooled cars, with a global desirability index of 0.1523.

Therefore, it can be concluded that, overall, *Shuttl* is the most suitable transportation system in NCR considering the criteria of sustainability. It is followed by state-run buses, and lastly by pooled car system. *Shuttl* is most preferred also due to its last-mile connectivity and comfortability to passengers. The *Shuttl* company is located in NCR and specializes in providing safe, comfortable and hassle-free travel to daily commuters. The company has aggregated over 1500 AC buses and 500 cabs across the country. Electric vehicles have also been inducted in the fleet. With the experience gained in the past three and half years, the company is poised to take a quantum leap in its endeavor to benchmark itself against the very best in the industry; it has worked with different governments and municipalities to provide the best solution in public private partnership (PPP) model. Due to the rigorous quality controls and client-oriented approach, the company has acquired its client list with some very reputed and renowned names. For the past few years, *Shuttl* company has started putting efforts toward the implementation of green practices in various activities. Therefore, it is felt that other alternatives should also adopt green practices to make their transport system sustainable.

Table 5. Weights of criteria for alternatives

Criteria for selection of alternative	Level 2 weights	Priority vector state-run bus	Priority vector pooled car	Priority vector *Shuttl*
CO_2 emissions	0.3258	0.3202	0.1226	0.5571
Cost of fuel	0.1072	0.1721	0.3042	0.5237
Energy efficiency	0.0392	0.1416	0.3338	0.5247
Cost of maintenance	0.0571	0.1279	0.3601	0.5119
Number of accidents	0.0839	0.2431	0.0882	0.6687
Congestion	0.2084	0.6333	0.1062	0.2605
Number of injuries	0.1511	0.2431	0.0882	0.6687
Road noise	0.0273	0.2721	0.1199	0.6080

Table 6. Global weights of alternatives

	State-run bus	Pooled car	*Shuttl*
CO_2 emissions	0.1043	0.0399	0.1815
Cost of fuel	0.0184	0.0326	0.0561
Energy efficiency	0.0056	0.0131	0.0206
Cost of maintenance	0.0073	0.0206	0.0293
Number of accidents	0.0204	0.0074	0.0561
Congestion	0.1320	0.0221	0.0543
Number of injury	0.0367	0.0133	0.1010
Road noise	0.0074	0.0033	0.0166
Total global weight	*0.3322*	*0.1523*	*0.5155*

Table 7. Global desirability index of alternatives

Global desirability for state-run bus	0.3322
Global desirability for pooled car	0.1523
Global desirability for *Shuttl*	0.5155

5. Conclusions, limitations, future scope and managerial implications

The purpose of this study was to identify the criteria for sustainable transport system in metro cities and suggests a framework for the selection of the most sustainable transportation system in NCR Delhi, India. We have used AHP for illustrating the proposed framework. This study has been done for policymakers in designing sustainable transportation in an existing metro which has severe traffic congestion. In the present situation, the development of sustainable transport system is a challenge. In this study, three alternatives of transport are considered in context to NCR Delhi, India. These are state-run buses, pooled cars and *Shuttl*. This study is trying to illustrate a framework for the selection of sustainable transport.

In the present era of globalization, ensuring sustainable transport is a big challenge. Green practices in transportation can achieve both economic and environmental objectives. In the era of environmentalism, cities and municipalities having sustainable transport options will surely benefit. The benefits will ultimately lead to performance improvements, increased public comfort, higher profits and overall sustainable growth of cities. We have considered eight criteria for the selection of a sustainable transport system. Criteria for selection of sustainable transport are CO_2 emissions, cost of fuel, energy efficiency, cost of maintenance, number of accidents, congestion, number of injuries and road noise. Based on the pairwise comparison of different criteria through AHP, it is observed that CO_2 emission is the most important criterion, followed by a reduction in congestion and number of injuries. After finding the weights of different criteria, the global desirability index of all three alternatives is calculated. It is found that *Shuttl* has the highest global desirability index. It means that in terms of these criteria, *Shuttl* is the most sustainable alternative.

Findings will be very useful primarily for policymakers and for city planners in developing sustainable transport systems in metro cities. These findings will also have a major impact on the selection of sustainable transport mode by public in cities. Usually public use their own vehicles or other means of transport due to unavailability of sustainable transport. It will further help in terms of improving energy efficiency and reducing fuel consumption, reduction of carbon footprint, maintaining the environment, among others. The results of this research will help future researchers to prioritize various factors according to their benefits and to have a holistic approach toward green logistics. Once the decisions are made at higher levels, it will motivate the people at the mid-level and lower level to work toward sustainable activities. This ultimately will improve the society and the quality of life.

There are certain limitations to the present research. First limitation is that we have considered limited number of sustainability criteria; these criteria may be increased to develop a more comprehensive framework. Framework based on triple bottom line approach can be also used as a part of future studies. Apart from it, ranking of these criteria may also vary from city to city. Second limitation is that AHP is not able to capture the vagueness in decision-making. Therefore, fuzzy AHP can be considered for further detailed analysis as future scope of study in a different context of a sustainable transportation system. Findings may be further validated through case studies for other metro cities with heavy traffic congestions. Further, a similar study can be conducted in another country, and results can be compared to analyze the similarities and differences in the preferences. We can also study about the behavior of public for sustainable transport selection through empirical studies.

References

Agrawal, S. and Singh, R.K. (2019), "Forecasting product returns and reverse logistics performance: structural equation modelling", *Management of Environmental Quality: An International Journal*, doi: 10.1108/MEQ-05-2019-0109.

Arnott, R. and Small, K. (1994), "The economics of traffic congestion", *American Scientist*, Vol. 82 No. 5, pp. 446-455.

Awasthi, A., Chauhan, S.S. and Omrani, H. (2011), "Application of fuzzy TOPSIS in evaluating sustainable transportation systems", *Expert Systems with Applications*, Vol. 38, pp. 12270-12280.

Aßmann, D. and Sieber, N. (2005), "Transport in developing countries: renewable energy versus energy reduction", *Transport Reviews*, Vol. 25 No. 6, pp. 719-38.

Bhandari, D., Singh, R.K. and Garg, S.K. (2018), "Justification of advanced manufacturing technologies for small and medium enterprises from auto component sector: AHP approach", *International Journal of Productivity and Quality Management*, Vol. 23 No. 4, pp. 473-491.

Black, W.R. (1996), "Sustainable transportation: a US perspective", *Journal of Transport Geography*, Vol. 4, pp. 151-159.

Boarnet, M.G., Kim, E.J. and Parkany, E. (1998), *Measuring Traffic Congestion*.

Browne, M. and Allen, J. (1998), "Strategies to reduce the use of energy by road freight transport in cities", *Transport Logistics*, Vol. 1 No. 3, pp. 195-209.

Castillo, N.H. and Pitfield, D.E. (2010), "ELASTIC _ a methodological framework for identifying and electing sustainable transport indicators", *Transportation Research Part D: Transport and Environment*, Vol. 15 No. 4, pp. 179-188.

Cattaneo, M., Malighetti, P., Morlotti, C. and Paleari, S. (2018), "Students' mobility attitudes and sustainable transport mode choice", *International Journal of Sustainability in Higher Education*, Vol. 19 No. 5, pp. 942-962.

Chhabra, D., Garg, S.K. and Singh, R.K. (2017), "Analysing alternatives for Green Logistics in an Indian automotive organisation: a case study", *Journal of Cleaner Production*, Vol. 167 No. 20, pp. 962-969.

Colvile, R.N., Kaur, S., Britter, R., Robins, A., Bell, M.C. and Shallcross, D. (2004), "Sustainable development of urban transport systems and human exposure to air pollution", *The Science of the Total Environment*, Vol. 334, pp. 481-e487.

Cosgrove, D., Gargett, D. and Mitchell, D. (2009), *Urban Passenger Transport: How People Move about in Australian Cities - Information Sheet 31*, Bureau of Infrastructure, Transport and Regional Economics, Canberra, Australia, ISSN 1440–9593.

Cruz, I.S. and Katz-Gerro, T. (2016), "Urban public transport companies and strategies to promote sustainable consumption practices", *Journal of Cleaner Production*, Vol. 123, pp. 28-33.

Currie, G., Truong, L. and Gruyter, C.D. (2018), "Regulatory structures and their impact on the sustainability performance of public transport in world cities", *Research in Transportation Economics*, Vol. 69, pp. 494-500.

Daly, H.E. (1991), *Steady State Economics*, Island Press, Washington, DC.

EC-JRC report (2016), *PBL Netherlands Environmental Assessment Agency European Commission, Joint Research Centre. Trends in Global CO2 Emissions: 2016 Report*, PBL Publishers, Netherlands, pp. 14/30-32.

Elvik, R. (2000), "How much do road accidents cost the national economy?", *Accident Analysis and Prevention*, Vol. 32 No. 6, pp. 849-851, doi: 10.1016/s0001-4575(00)00015-4.

European Commission (2008), "Special eurobarometer 295: attitudes of European citizens towards the environment, brussels, European commission", available at: http://ec.europa.eu/public_opinion/archives/ebs/ebs_295_en.pdf.

Gupta, A. and Singh, R.K. (2020), "Managing operations by a logistics company for sustainable service quality: Indian perspective", *Management of Environmental Quality International Journal*, doi: 10.1108/MEQ-11-2019-0246.

Gurtu, A. (2014), *A Framework for Reducing Greenhouse Gas (GHG) Emissions through Carbon Pricing for Offshore Outsourcing*, Ph.D., Ryerson University, Toronto, available at: http://digital.library.ryerson.ca/islandora/object/RULA%3A3388 (accessed 28 January 2015).

Gurtu (2019), "A pioneering approach to reducing fuel cost and carbon emissions from transportation", *Transportation Journal*, Vol. 58 No. 4, doi: 10.5325/transportationj.58.4.0309.

Gurtu, A., Jaber, M.Y. and Searcy, C. (2015), "Impact of fuel price and emissions on inventory policies", *Applied Mathematical Modelling*, Vol. 39 Nos 3-4, pp. 1202-1216, doi: 10.1016/j.apm.2014.08.001.

Gurtu, A., Searcy, C. and Jaber, M.Y. (2016), "A framework for reducing global manufacturing emissions", *The Journal of Environment and Development*, Vol. 25 No. 2, pp. 159-190, doi: 10.1177/1070496515623821.

Gurtu, A., Searcy, C. and Jaber, M.Y. (2019), "Transportation and sustainable supply chains", in Sarkis, J. (Ed.), *Handbook on the Sustainable Supply Chain*, 1 ed., Edward Elgar Publishing, pp. 410-428.

Haase, D. and Nuissi, H. (2007), "Does urban sprawl drive changes in the water balance and policy? The case of Leipzig (Germany), 1870–2003", *Landscape and Urban Planning*, Vol. 80 Nos 1–2, pp. 1-13.

Hall, R.P. (2002), *Introducing the Concept of Sustainable Transportation to the U.S. DOT through the Reauthorization of TEA-21*, Massachusetts Institute of Technology, Cambridge, MA.

Henley, W.H., Ranganathan, S.K. and Gurtu, A. (2018), "Improving effectiveness of public service advertisements to prevent texting and driving of American youth", *International Journal of Nonprofit and Voluntary Sector Marketing*, Vol. 23 No. 4, pp. 1-9, doi: 10.1002/nvsm.1626.

Hindu (2019), *Fifteen of the 20 Most Polluted Cities in the World Are in Indi*, 5th March 2019.

Kennell, B. (2016), "Environmental concern empowers the people", available at: http://www.huffingtonpost.com/brian-kennell/environmental-concern-emp_b_8105580.html (accessed 18 October 2016).

Kim, H. and Macdonald, E. (2016), "Does wind discourage sustainable transportation mode choice? Findings from San Francisco, California, USA", *Sustainability*, Vol. 8 No. 3, p. 257.

Litman, T. and Burwell, D. (2006), "Issues in sustainable transportation", *International Journal of Global Environmental Issues*, Vol. 6 No. 4, p. 331e347.

Loo, B.P.Y. (2002), "Role of stated preference methods in planning for sustainable urban transportation: state of practice and future prospects", *Journal of Urban Planning and Development*, Vol. 128 No. 4, pp. 210-224.

Marsden, G. and Rye, T. (2010), "The governance of transport and climate change", *Journal of Transport Geography*, Vol. 18, pp. 669-678.

Mattioli, G. (2014), "Where sustainable transport and social exclusion meet: households without cars and car dependence in great britaBritainin", *Journal of Environmental Policy and Planning*, Vol. 16 No. 3, pp. 379-400.

May, A.D., Page, M. and Hull, A. (2001), "Developing a set of decision-support tools for sustainable urban transport in the UK", *Transport Policy*, Vol. 15, pp. 328-340.

Mohan, D. and Tiwari, G. (1999), "Sustainable transport systems: linkages between environmental issues, public transport, non-motorised transport and safety", *Economic and Political Weekly*, Vol. 34 No. 25, pp. 1589-1596.

Müller-Wenk, R. (2004), "A method to include in lca road traffic noise and its health effects", *International Journal of Life Cycle Assessment*, Vol. 9 No. 2, pp. 76-85, doi: 10.1007/bf02978566.

Nakamura, H., Hayashi, Y. and May, A.D. (Eds) (2004), "Urban transport and the environment—an international perspective", *World Conference on Transportation Research Society and Institute for Transport Policy Studies*, Elsevier, Amsterdam.

Reddy, A.K., Anand, Y.P. and D'Sa, A. (2000), "Energy for a sustainable road/rail transport system in India", *Energy for Sustainable Development*, Vol. 4 No. 1, pp. 29-44.

Saaty, T.L. (1980), *The Analytic Hierarchy Process*, McGraw-Hill, New York, NY.

Saaty, T.L. (1994), "How to make a decision: the analytic hierarchy process", *Interface*, Vol. 24 No. 6, pp. 19-43.

Salehi, M., Jalalian, M. and Siar, M.M.V. (2017), "Green transportation scheduling with speed control: trade-off between total transportation cost and carbon emission", *Computers and Industrial Engineering*, Vol. 113, pp. 392-e404.

Schipper, L. (2003), "Sustainable urban transport in the 21st century: a new agenda", *Proceedings of Conference: Transportation, Energy, and Environmental Policy: Managing Transitions, held in Monterey, CA, 9/11/2001–9/12/01*, Transportation Research Board, Washington, DC.

Schneider, R.J. (2013), "Theory of routine mode choice decisions: an operational framework to increase sustainable transportation", *Transport Policy*, Vol. 25, pp. 128-e137.

Shi, J. and Zhou, N. (2012), "A quantitative transportation project investment evaluation approach with both equity and efficiency aspects", *Research in Transportation Economics*, Vol. 36 No. 1, pp. 93-100, doi: 10.1016/j.retrec.2012.03.00.

Singh, R.K. (2013), "Prioritising the factors for coordinated supply chain using analytic hierarchy process (AHP)", *Measuring Business Excellence*, Vol. 17 No. 1, pp. 80-98, Emerald.

Singh, R.K. and Agrawal, S. (2018), "Analyzing disposition strategies in reverse supply chains: fuzzy TOPSIS approach", *Management of Environmental Quality: An International Journal*, Vol. 29 No. 3, pp. 427-443.

Song, Y., Miller, H.J., Stempihar, J. and Zhou, X.S. (2017), "Green accessibility: estimating the environmental costs of network-time prisms for sustainable transportation planning", *Journal of Transport Geography*, Vol. 64, pp. 109-e119.

Stefanieca, A., Hosseinib, K., Xiec, J. and Lia, Y. (2020), "Sustainability assessment of inland transportation in China: a triple bottom line-based network DEA approach", *Transportation Research Part D*, Vol. 80 No. 2020, p. 102258.

Tanowitz, M. and Rutchik, D. (2008), *Squeezing Opportunity Out of Higher Fuel Costs*, Supply Chain Management Review, October, pp. 34-40.

The Maritime Executive (2020), "Transport uses 25 percent of world energy", available at: https://www.maritime-executive.com/article/transport-uses-25-percent-of-world-energy (accessed 28 March 2020).

Thuermer, K.E. (2008), "Air Cargo braces for a slowdown", *Supply Chain Management Review*, December.

US EPA (2020), US Environment Protection Agency Report, 2020.

Van den Berg, R. and De Langen, P.W. (2017), "Environmental sustainability in container transport: the attitudes of shippers and forwarders", *International Journal of Logistics Research and Applications*, Vol. 20 No. 2, pp. 146-162.

Vaziri, M. (2002), "Development of highway congestion index with fuzzy set models", *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1802 No. 1, pp. 16-22, doi: 10.3141/1802-03.

Vicente, P., Sampaio, A. and Reis, E. (2020), "Factors influencing passenger loyalty towards public transport services: does public transport providers' commitment to environmental sustainability matter?", *Case Studies on Transport Policy*, Vol. 8, pp. 627-638.

World Bank (1996), *Sustainable Transport: Priorities for Policy Reform*, The International Bank for Reconstruction and Development/ The World Bank, Washington, DC.

Zhao, X., Ke, Y., Zuo, J., Xiong, W. and Wu, P. (2020), "Evaluation of sustainable transport research in 2000-2019", *Journal of Cleaner Production*, Vol. 256, p. 120404.

Zheng, J., Garrick, N.W., Atkinson-Palombo, C., McCahill, C. and Marshall, W. (2013), "Guidelines on developing performance metrics for evaluating transportation sustainability", *Research in Transportation Business and Management*, Vol. 7, pp. 4-13, doi: 10.1016/j.rtbm.2013.02.001.

Zhou, J. (2012), "Sustainable transportation in the US: a review of proposals, policies, and programs since 2000", *Frontiers of architectural research*, Vol. 1 No. 2, pp. 150-e165.

About the authors

Abhishek Singh is doing his graduation in environment engineering from Delhi Technological University, Delhi, India. He has keen interest in environment-related research issues.

Amulya Gurtu is an Associate Professor of Supply Chain Management at the Cofrin School of Business, University of Wisconsin, Green Bay, WI, USA. His research interests are in global manufacturing and supply chains and their impact on the environment and society. He obtained his PhD from Ryerson University, Canada, and an MBA from Western University, Canada. His research has been published in the *Journal of Business and Economic Perspectives,* the *International Journal of Physical Distribution and Logistics Management, Transportation Journal, the International Journal of Nonprofit and Voluntary Sector Marketing*, *Operations and Supply Chain Management*, *Services Marketing Quarterly*, the *International Journal of Inventory Research*, the *Journal of Cleaner Production*, the *Journal of Environment and Development, Management Research Review, Sustainable Production and Consumption* and *Applied Mathematical Modelling*. He served as the Editor-in-Chief of the *International Journal of Sustainable Entrepreneurship and Corporate Social Responsibility* (IJSECSR) published by IGI Global, and he is currently serving as an associate editor of the *International Journal of Systems Science: Operations and Logistics* published by Taylor and Francis.

Dr. Rajesh Kumar Singh is professor in operations management area at Management Development Institute (MDI), Gurgaon, India. His areas of interest include supply chain management, Industry 4.0 and circular economy. He has about 170 research papers published in reputed international/national journals and conferences. He has published papers in journals such as *Annals of Operations Research, Resources, Conservation and Recycling, Journal of Cleaner Production, Industrial Management and Data Systems, Production planning and Control, International Journal of Production Research, Singapore Management Review, International Journal of Productivity and Performance Management, Benchmarking: An International Journal, Journal of Modelling in Management, Competitiveness Review: An International Business Journal, Global Journal of Flexible Systems and Management, International Journals of Productivity and Quality Management, International Journal of Services and Operations Management, International Journal of Automotive Industry and Management, International Journal of Logistic Systems Management, IIMB Management Review and Productivity Promotion*. Presently, he serving as an associate editor of *International Journal of Consumer Studies* published by Wiley. He has also worked as a guest editor for many reputed journals. Rajesh Kumar Singh is the corresponding author and can be contacted at: rksdce@yahoo.com

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**

International Journal of Mechanical and Production Engineering Research and Development (IJMPERD)
ISSN (P): 2249–6890; ISSN (E): 2249–8001
Vol. 10, Issue 3, Jun 2020, 3397-3406
© TJPR Pvt. Ltd.

TRANS STELLAR
•Journal Publications •Research Consultancy

SIX SIGMA APPLICATION IN KETCHUP INDUSTRY FOR DAMAGED GLASS BOTTLES: A CASE STUDY

YATHARTH SHARMA, SARMAD HASSAN & RAJAT KUMAR

Delhi Technological University, New Delhi, India

ABSTRACT

The aim of this paper is to examine the problem of loss incurred due to damage of glass bottles in the tomato sauce/ketchup industry which not only results in depreciating economy to the manufacturer but also degrades the reputation of a company and degrades its image. This paper uses the DMAIC methodology which is a part of Six Sigma for evaluating the problem. The identification of the root causes was done using Kaoru Ishikawa Diagram (Fish Bone Diagram). The weight of individual root cause was calculated using Analytical Hierarchy Process (AHP). Pareto Chart was established to figure out the root causes of the greatest consequence, which were having a very significant impact on the deliverables. Subsequently, brainstorming was done with the experts of this field including the department of Quality Control and logistics was organized which facilitated in figuring out the solutions by reviewing the process map. This paper has been the first attempt to find and review the root causes of the problem of Glass Bottles damage in this particular industry and suggest solutions so that there is a decline in the quantity of damages using DMAIC approach given under Six Sigma methodology.

KEYWORDS: AHP, DMAIC, Industrial Management, Lean Management, Quality Control & Six Sigma

Received: Jun 20, 2020; **Accepted**: Jul 01, 2020; **Published**: Jun 14, 2020; **Paper Id.**: IJMPERDJUN2020323

Original Article

1. INTRODUCTION

Six Sigma is a useful methodology for problem-solving. It gives a valuable measurement approach. Six Sigma has a mathematical base, which, with proper utilization of methodologies, can help in refining the quality of both the process and the product. The basic goal of this methodology is the application of a strategy based on measurement focused on improving the process and reducing variation by applying specific projects. Managers in the food industry have to thus focus on maintaining profitability in a reducing market, along with providing efficient service and good quality products to increasingly sophisticated buyers. In order to accomplish this apparently unattainable goal, companies can choose two strategies. First, they can improve their system's operational efficiency. Second, they can take proper steps to improve its operational quality.

2. CASE STUDY

This case study is performed at Company XYZ which is a leading ketchup bottle manufacturer in India. The DMAIC methodology is applied below. To the best of our knowledge, no research work is done till date to resolve the problem of damage of Glass Bottles in Ketchup industry.

2.1 Define Phase

2.1.1 Problem Statement

To reduce the quantity of damaged glass bottles from 9,830 in the preceding year to 4,200 in the next year.

2.1.2 Business Case

Company XYZ has experienced 9,830 damaged glass bottle units in the preceding year worth of INR 1,100,000/-. This has resulted in the followingdirect cash losses to the company. These losses are mentioned in table 1.

Table 1

Particular	Loss Amount (INR)
Loss of Glass bottle cost	70337 /-
Loss of cost of extra engagement of labor	44322 /-
Loss in revenue due to devaluation of the product	126632 /-
Total Loss	**241,291 /-**

Therefore, loss from each bag damaged= INR 241,291 / 9830 =INR 24.54/-

2.1.3 Process Map

The sequence in which the movement of bag takes place and its sub process map is shown in figure. 1

2.2 Measure Phase

After this tool of data analysis, we use AHP to find the weightage of each root cause followed by use of Pareto Chart to prioritize our resources and direct attention to vital root causes.

The cause & effect relationship shown in figure 2 below can occur at the vendors point, in transit between vendors point and main store, at the main store, in transit between main store and filling station and at the filling station. Whereas the causes & effects relationship shown in figure. 3 can occur at the filling station, in transit between filling station and warehouse, in transit between various warehouse positions, in transit between warehouse and platform, at platform, in transit between platform and truck, at truck during transit between plant and point of unloading and at the point of unloading.

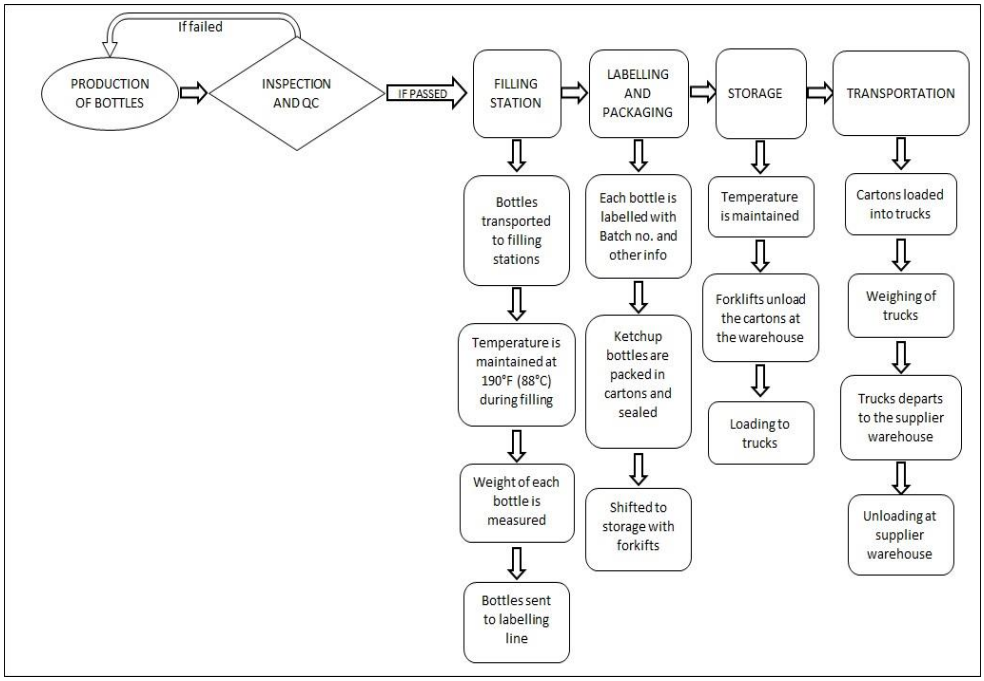


Figure 1: Process Map and Sub-Process Map (Glass Bottle).

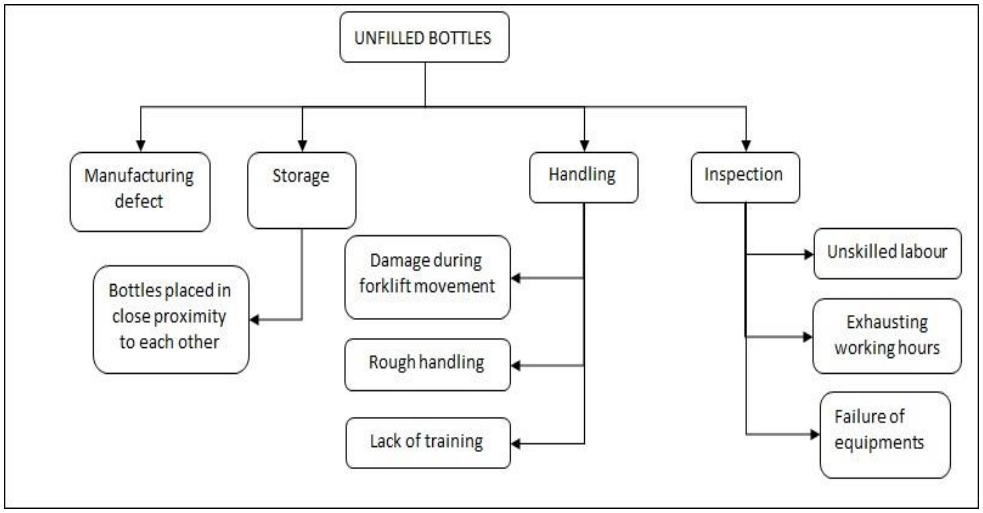


Figure 2: Fish Bone Diagram (Unfilled Bottles).

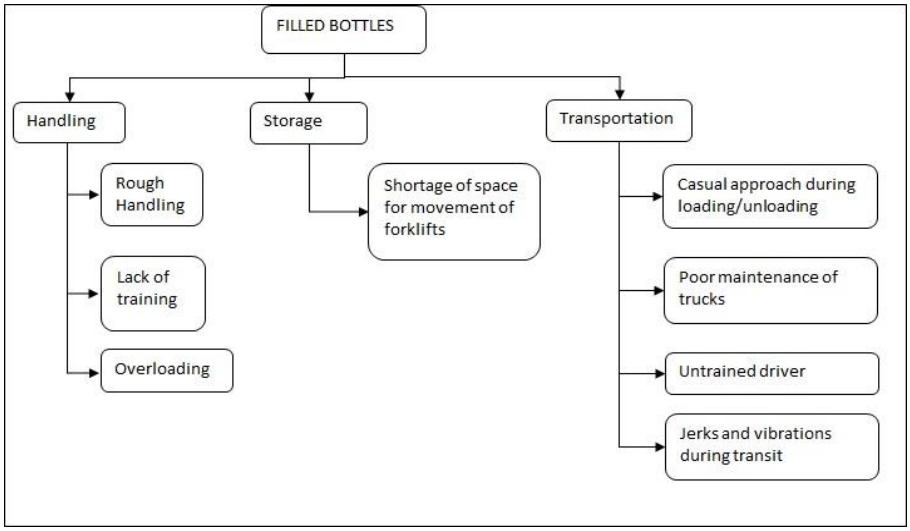


Figure 3: Fish Bone Diagram (Filled Bottle).

2.3 Analysis Phase

Due to similarity between the root causes, they are combined together under six main heads, which from here onwards will be identified as the root causes. First root cause is Storage defects (ST) which includes improper management of space which may cause defects during movement of forklifts. Second root cause is Transportation (T) which is very important to consider since the products are fragile and can break if not taken care properly. It includes improper loading/unloading practice, poor maintenance of trucks, too much jerk and vibration during transit, etc. Third root cause is Production (P) where defects can arrive due to unskilled labor, improper production techniques, etc. Fourth root cause is Packaging (PG). Fifth root cause is lack of Attention and Supervision (S) at multiple stages like filling station, loading/unloading. Sixth root cause is issues during Inspection (IN), this may be due to non-maintenance of Inspecting machines, false positive inspection, etc.

2.3.1 Weight Age Calculation using ahp

Weightage of every individual root cause is calculated using AHP (Analytic Hierarchy Process). AHP is designed tosolve

complex multi criteria problems. Many quantitative approaches have been taken into observation to evaluate new technology implementation. Table 2 depicts the pairwise combination of each of the root cause. The fraction obtained in table 2 is converted into decimals as presented in table 3. On the basis of these pairwise combinations, iterations are performed till we obtain consistency ratio of less than 0.1 (If the consistency ratio is more than 0.1, then the data provided is not reliable). These iterations is shown in table 4.

Table 2: Pairwise Combinations of Each Root Cause

	ST	T	PR	PG	S	IN
ST	1	1/4	1/3	1/2	1/3	1/5
T	4	1	2	5	3	4
PR	3	1/2	1	2	1	1/3
PG	2	1/5	1/2	1	1/5	1/2
S	3	1/3	1	5	1	3
IN	5	1/4	3	2	1/4	1
SUM	18	2.533	7.833	15.500	5.783	9.033

Table 3: Pairwise Combination of Root Cause In Decimals

	ST	T	PR	PG	S	IN
ST	0.055	0.098	0.042	0.032	0.057	0.022
T	0.222	0.394	0.255	0.322	0.518	0.442
PR	0.166	0.197	0.127	0.129	0.172	0.036
PG	0.111	0.078	0.063	0.064	0.034	0.055
S	0.166	0.131	0.127	0.322	0.172	0.332
IN	0.277	0.098	0.382	0.129	0.043	0.110

Table 4: Checking For Consistency

	ST	T	PR	PG	S	IN	Weighted Sum Average	Criteria Weight	λ
ST	0.051	0.090	0.046	0.034	0.070	0.035	0.326	0.051	6.331
T	0.206	0.359	0.277	0.340	0.627	0.695	2.504	0.359	6.967
PR	0.154	0.180	0.138	0.136	0.209	0.058	0.875	0.138	6.325
PG	0.103	0.072	0.069	0.068	0.042	0.087	0.441	0.068	6.476
S	0.154	0.120	0.138	0.340	0.209	0.521	1.483	0.209	7.099
IN	0.257	0.090	0.415	0.136	0.052	0.174	1.125	0.174	6.473

After taking the average value ofλ, we get $\lambda(avg)$=6.611

$$Consistency\ Index\ (CI) = \frac{\lambda(\text{avg})-n}{n-1} = 0.122366$$

$$Consistency\ Ratio = \frac{CI}{RI} = 0.098682$$

Where RI is Random Index and is fixed as 1.24 for matrix size (Factors) of 6. Since the consistency ratio is less than 0.10, we can say that our matrix is reasonably consistent.

ST	0.05
T	0.36
PR	0.14
PG	0.07
S	0.21
IN	0.17

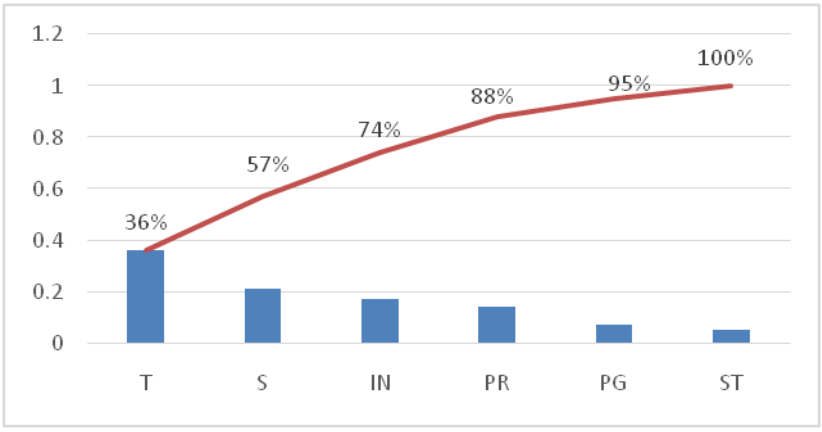


Table 5: Values of Pareto Chart Figure 4: Pareto Chart (Weightage of Each Root Cause).

2.3.2 Analysis of Weight Age of each Root Cause

Figure 4 shows analysis of root causes through their weightage obtained from AHP, as shown in table 5

It can be understood from the pare to chart that Transportation (T) and Storage (S) combined accounts for more than half of the loss due to damaged bottles. Other major root cause being Inventory (IN). Remaining root causes (PG & ST) account for just 20% of the loss. Giving due priority to T, S, IN can cover upto 3/4th of the total loss due to damaged bottles. CE being the only factor on which the company will have least control because the product is FMCG and it spent less time in company warehouse.

2.3.3 Analysis of Loss that is Associated with each Root Cause

Through the weightage calculated from AHP, we now deduce the loss related with each root cause by multiplying their weightage with the total no. of bags damaged bottles. This gives us the no. of bottles damaged due to each root cause or the loss related with each root cause in INR, since, we know the cost of each bottle damaged (As calculated in the business case).Table 8 shows this analysis. Furthermore, table 8 is represented in the form of a Pareto chart shown in figure. 6 which high lights the importance of each root cause in monetary terms. From figure 6 it is clearly visible that our target of reducing loss due to bottle damage by 43% is easily achievable by primarily focusing on 3 main root causes, namely, T, S and IN.

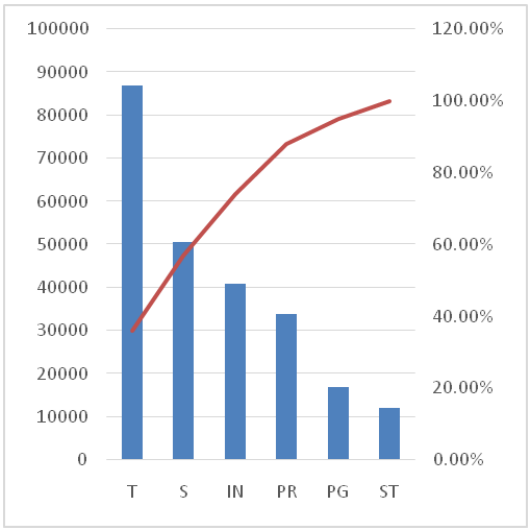


Figure 5: Pareto Chart (Losses Linked With Each Root Cause).

Table 6: Values of Pareto Chart

Root Cause	Weight Age	No. of Bottles Damaged	Loss Amount (INR)
ST	0.05	492	12064.55
T	0.36	3539	86864.76
PR	0.14	1376	33780.74
PG	0.07	688	16890.37
S	0.21	2064	50671.11
IN	0.17	1671	41019.47

Since cost of each glass bottle damaged = INR 24.54 /- (Illustrated in the Business Case)

2.3.4 Date wise Analysis of 12-Months Data

It is important to analyze the point of time during the month at which glass bottle damage takes place. This helps us in verifying the practicality of our interpretations of figure 4 and figure 5. Table 8 gives a 12 months data of summation of glass bottle damage on each of the date. Table 8 is represented in the form of a paretochart in Figure. 6 which is not following the decreasing trend since we need to analyze the trend across the month. It is clearly visible that in the second half of the month, due to excess movement of the glass bottles and lower rate of attention by staff on each bottle, more bottles get damaged. This verifies our claim that over majority of the losses due to bag damage can be prevented if three root causes T, S and IN are properly focused on, along with brief attention to the issue of maintenance.

2.4 Improve Phase

In order to reduce loss due to bags getting damaged, following solutions are suggested in table 7.

Table 7: Suggestions from the Improve Phase

S. No.`	Solutions	Target Cause
1	Proper loading &unloading procedures must be followed	T
2	Training to the workforce and regularly briefing by supervisor in order for cheering up morale.	S
3	Trucks must be serviced properly so that its suspension system does not becomes stiff	T
4	More number of testing must be done,	IN
4	Proper training must be provided to the driver	T
5	Solving the points of bottlenecks in order to become more efficient, automated monitoring of the production processes	PR
6	reducing false positive reports of the empty bottles	IN
7	Gentle handling of the bottles, prevention of collision of bottles from each other	PG
8	Warehouse and production plant must be organized properly so that space management can be done	ST
9	Prevent overloading of the trucks and forklift	T/ST/S
10	Give proper guidelines for proper handling of the bottles	S
11	Give workforce adequate breaks and manage proper work-life balance	S

2.5 Control Phase

The aim of this project was to bring back the focus of the management to surge in losses due to surge in the amount of glass bottles getting damaged. Thus, this project has identified not only the root causes but also direct financial loss due to each of the root cause of this problem. Solutions based on interpretations of the pareto chart have been accepted and implemented. For sustenance of results, the new process was standardized, documented, and distributed at every department to execute on a daily basis. Additionally, the ongoing performance was reviewed till it became part of an everyday daily process.

Table 8: Pareto Chart (Date wise Analysis of 12-Months Data

Date	Damaged Bottles	Date	Damaged Bottles	Date	Damaged Bottles	Date	Damaged Bottles
1	146	9	235	17	378	25	376
2	204	10	137	18	598	26	312
3	116	11	278	19	478	27	289
4	118	12	248	20	446	28	256
5	198	13	221	21	364	29	221
6	273	14	464	22	398	30	309
7	489	15	443	23	287	31	451
8	263	16	489	24	345		

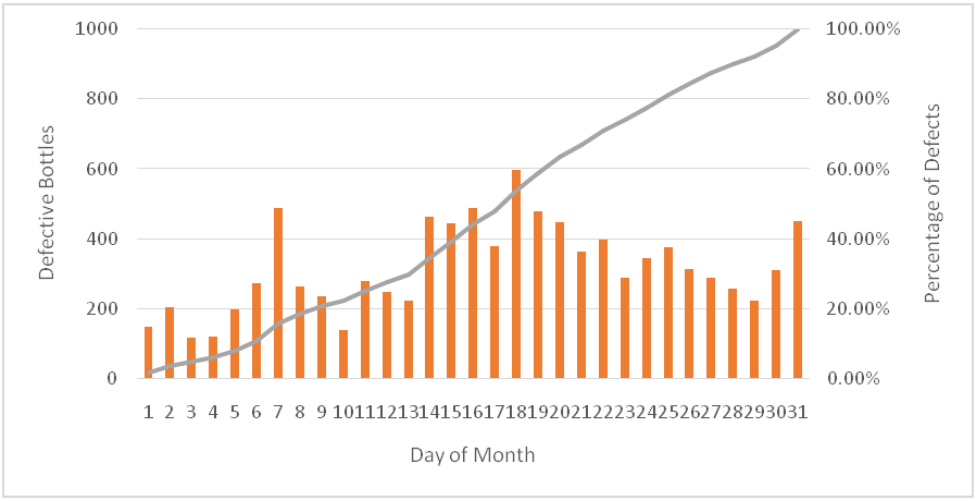


Figure 6: Pareto Chart (Date wise Analysis of 12 Month Data).

3. CONCLUSIONS

This study is performed to classify and examine the main causes of the problem of Glass Bottle damage in Ketchup/Tomato Sauce industry and suggest solutions to reduce damages.

In order to recognize opportunities for improvement, a process map was documented. Root causes were recognized with the support of fish-bone diagram. Weightage of each root cause was calculated using analytical hierarchy process and a pare to chart aided to focus attention on the most pressing root causes.

Accordingly, brainstorming was performed with the production, warehouse, logistics and quality control section was done which helped in framing the solutions by examining the process map.

Although this is one of the first attempts to focus on the problem of Glass Bottle damage, academics and practitioner could use this study for future study in other industries where Glass Bottle are used. Future research can be conducted by calculating the weightage of each root cause using paired comparison method or balanced score card method. A follow up on the level of lean implementation and cultural change in the companies addressed in this study is advised. Also, specifically concerning other causes, future work may focus on more accurately quantifying the impact of machine design changes and the impact in inventory reduction. Finally, another interesting research avenue may be comparing Lean Management implementation results among companies of different sectors. This would allow further understanding the

potential and importance of process innovation in different industrial contexts.

REFERENCES

1. *Antony J (2008). Can Six Sigma be effectively deployed in SMEs? Int. J.Prod. Perform. Manage., 57(5): 420-423.*
2. *Antony J, Banuelas R (2001). Six Sigma: A business strategy for manufacturing organizations. Manuf. Eng., 80(3): 119-121.*
3. *Antony J, Banuelas R (2002). Key ingredients for the effective implementation of Six Sigma programmes. Measuring Bus. Excellence., 6(4): 20-28.*
4. *AHN, Kook Chan. "Impact Response of the Film Mechanical Properties of the Coated Glass Plate."International Journal of Mechanical and Production Engineering Research and Development (IJMPERD) 9. 2, Apr 2019, 499-506*
5. *Bergman B, Kroslid D (2000). Six Sigma : a revival of the Profound Knowledge of variation. Proceedings of the 3rd International Conference on Building People and Organizational Excellence (Aarhus, Denmark), pp. 260- 266.*
6. *Brewer PC, CPA, Eighme JE (2005). Using six sigma to improve the finance function: Here are some tips for success. Strat. Finance., 85(5): 27-33.*
7. *Kang, H. D., and Kook Chan Ahn. "A Study on the Dynamic Behaviour of the Coating Tempered Glass Plate Under Impact." Int. J. of Mechanical and Production Engineering Research and Development (IJMPERD), 8 (6) (2018): 193-200.*
8. *Breyfogle FW (1999). Implementing Six Sigma: Smarter Solutions using Statistical Methods, New York: John Wiley.*
9. *Chen SC, Chen KS, Hsia TC (2005). Promoting customer satisfactions by applying six sigma: An example from the automobile industry. Qual. Manage. J., 12(4): 21-33.*
10. *Mohan, N. A. N. J. A. N. G. U. D., Sathyashankara Sharma, and R. I. T. E. S. H. Bhat. "A comprehensive study of glass fibre reinforced Polymer (gfrp) drilling." International Journal of Mechanical and Production Engineering Research and Development 9.1 (2019): 1-10.*
11. *Kumar M, Antony J (2008). Comparing the quality management practices in UK SMEs. Ind. Manage. Data Syst., 108(9): 1153-1166.*
12. *Hellsten U, KlefsjoÈ B (2000). TQM as a management system consisting of values, techniques and tools. The TQM Magazine., 12: 238- 244.*
13. *Ahirrao, S. P., Kashinath N. Borse, and SonaliBagrecha. "Eco friendly pavement blocks of waste glass, fly ash, and dust." International Journal of Civil, Structural, Environmental and Infrastructure Engineering Research and Development 3.5 (2013): 77-80.*
14. *Starbird D (2002). Business excellence: Six Sigma as a management system: A DMAIC approach to improving Six Sigma management processes. Annual Quality Congress Proceedings, Milwaukee, pp. 47-55.*
15. *Heymans, B. (2015). Lean manufacturing and the food industry. Retrieved June 5, 2015, from http://www.flowmakers.com/articles/Articlefoodindustryandkaizen.pdf*
16. *Hirano, H. (2009). JIT implementation manual – The complete guide to just-in-time manufacturing: Volume 2 – Waste and the 5S's. New York: Productivity Press.*
17. *George, M. L. (2003). Lean six sigma for service: How to use lean speed and six sigma quality to improve services and transactions. McGraw-Hill.*
18. *Kennedy, I., Plunkett, A., &Haider, J. (2013). Implementation of lean principles in a food manufacturing company. In A.*

Azevedo (Ed.)

19. *Advances in Sustainable and Competitive Manufacturing Systems, Lecture Notes in Mechanical Engineering (pp. 1579-1590). Springer. doi: 10.1007/978-3-319-00557-7_127*
20. *Ozden, B., "Use of AHP in decision making for flexible manufacturing systems", Journal of Manufacturing Technology Management, Vol. 16 No.7, pp. 808-819, 2005*
21. *Sanders, D. Hild, R., "Six Sigma on business processes: Common organizational issues" Quality Engineering, Vol.12 No.4, pp.603–610, 2000*
22. *Henry, J. R. (2012). Achieving lean changeover: Putting SMED to work. Boca Raton, FL: Productivity Press*

AUTHOR'S PROFILE



Yatharth Sharma received the B.Tech Degree in Mechanical Engineering from Delhi Technological University, in 2019. He was a former member of the team DTU AUV, SAE and ISHRAE. His interests include playing badminton, counselling on career paths and Industrial Engineering applications. He worked with Hilti India as GET and in Mercedes-Benz, EIL, Air India and Pawan Hans as intern and developed skills related to the defence industry equipment manufacturing.



Rajat Kumar received the B.Tech Degree in Mechanical Engineering from Delhi Technological University, in 2019. He was a former BAJA member of the team DelTech BAJA. His interests include designing the transmission system in BAJA vehicles, application of Input-Output Model for End-of-Life-Vehicles in India, thermodynamics and Industrial Engineering applications. He worked with Bharat Electronics Limited as intern and developed skills related to the defence industry equipment manufacturing.



Sarmad Hassan received the B.Tech Degree in Mechanical Engineering from Delhi Technological University, in 2019. He was a former member of ISHRAE. His interests include designing UAV and Robotic systems and Industrial

Engineering applications. He worked with Omnipresent Robot Tech as intern and now as a design engineer and developed skills related to the defence industry equipment manufacturing.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/342754860

Sodium hydroxide effect on the mechanical properties of flyash-slag based geopolymer concrete

Article *in* Structural Concrete · July 2020

DOI: 10.1002/suco.202000068

CITATIONS
0

READS
29

2 authors, including:



Manvendra Verma
Delhi Technological University
4 PUBLICATIONS **1** CITATION

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

behavior of Geopolymer concrete View project

All content following this page was uploaded by Manvendra Verma on 09 July 2020.

The user has requested enhancement of the downloaded file.

SPi	Journal Code	Article ID	Dispatch: 28-JUN-20	CE:
	SUCO	202000068	No. of Pages: 12	ME:

Received: 31 January 2020 | Revised: 30 April 2020 | Accepted: 19 May 2020

DOI: 10.1002/suco.202000068

TECHNICAL PAPER

fib WILEY

Sodium hydroxide effect on the mechanical properties of flyash-slag based geopolymer concrete

Manvendra Verma | Nirendra Dev

Department of Civil Engineering, Delhi Technological University, Delhi, India

Correspondence
Manvendra Verma, Department of Civil Engineering, Delhi Technological University, Delhi 110042, India.
Email: manvendraverma_2k16phdce05@dtu.ac.in

Funding information
Delhi Government

Abstract

Geopolymer concrete is an innovative, sustainable, cementless, and eco-friendly concrete that directly reduces the carbon footprints due to the total replacement of the cement from the concrete. A very excessive amount of CO_2 produces in the production of cement. In the experimental investigation, analysis of the mechanical properties or engineering properties of the GPC of the different molarity of NaOH (8–16 M), and also different Na_2SiO_3/NaOH ratio (0.5–3.0) analyses in the destructive testing of the GPC. Examine the curing temperature effect on the engineering properties of the GPC. After the experimental investigation, oven-cured specimens got a higher engineering strength compared to the ambient-cured specimens of the same mix designs. The 14 M mix design got the optimum point for engineering strength among the various molar concentrated mix designs. In the case of the alkaline ratio, the 2.5 got the optimum point of the engineering strength among all ratios of the sodium silicate to sodium hydroxide. The highest compressive strength, splitting tensile, and flexural strength in the all mix designs got are 35.7 N/mm^2, 5.2 N/mm^2, and 5.6 N/mm^2 respectively at 56 days after oven-curing. Based on results, proposed the correlation equation between the splitting strength and compressive strength and equation between the flexural strength and compressive strength.

KEYWORDS

compressive strength, curing condition, flexural strength, geopolymer concrete, molarity of sodium hydroxide, splitting tensile

1 | INTRODUCTION

Concrete is the second most abundant material used in the world after the water. Conventionally concrete used the Portland cement as a binding material in the concrete. India produced cement around 502 million tons in 2018. Cement is not eco-friendly, so many environmental issues associated with the production of cement in the industry. Cement production contributes around 8% of carbon dioxide in greenhouse gas emission. Cement production reduces the natural resource and also polluted

Abbreviations: AA, alkali activated; AAFS, alkali activated flyash slag; ASTM, American Society for testing and materials; EDS, energy-dispersive X-ray spectroscopy; GGBFS, ground granulated blast furnace slag; GPC, geopolymer concrete; OPC, ordinary Portland cement; SEM, scanning electron microscope; XRD, X-ray diffraction.

Discussion on this paper must be submitted within two months of the print publication. The discussion will then be published in print, along with the authors' closure, if any, approximately nine months after the print publication.

 © 2020 *fib*. International Federation for Structural Concrete

the environment that is very hazardous for our future. There are limited natural resources available on the earth, so we must go for sustainable development.

On the other hand, abundance of industrial solid waste like as flyash, slag and rice husk ash available on the earth. From the industrial production, around 200 million tons of flyash produced annually in the thermal power plant and 20 million tons of slag produced annually in steel manufacturing plants in India. Flyash and slag are wastes of the industry, so they dumped the industrial solid wastes in the ground using cultivated land and also polluted the environment due to the limited mode of use of the industrial solid wastes like as fly ash and slag.

GPC made by the flyash got more density as compare to made by the metakaolin due to its spherical and porous structure of the flyash particle. It also has very high fineness and increases the workability of the GPC. So, they got a higher solid to liquid ratio. For increasing the workability of the Geopolymer mortar increase the dosage of water or superplasticiser but after an optimum point dosage of superplasticiser create the adverse effect and reduce the strength of the mortar cubes. Moreover, increase the dosage of sodium silicate, and NaOH concentrations degrade the workability of mortar. In the AAFS mortar, autogenous shrinkage generates due to the self-desiccation after the hardening but reducing the volume due to the chemical shrinkage after made end product of the reaction in small size. Increasing the dosage of sodium silicate or slag content, reduced the total porosity and also reduced the mesopores volume, resulting in increasing the density of the mortar. The Flyash-Slag based mortar shows the higher drying shrinkage than the OPC samples, due to the higher mesopores volume of Flyash-Slag based mortar compared to OPC mortar.[1]

AAFS mortar showed higher flexural strength and compressive strength as compared to the OPC mortars. The lesser amount of water absorption as compared to OPC mortar and higher the drying shrinkage as compared to OPC mortar, increasing the dosage of flyash degrade the drying shrinkage of AAFS mortar. The end products of the AAFs mortar are semi-crystalline silicate hydrate and amorphous aluminosilicate gel.[2] The Flow of AA mortar rises with the increment of the water to binder ratio and degrades of the fine aggregate to binder ratio. If the fine aggregate to binder ratio higher than 2.5, the workability of the AA mortar sharply reduced. Generally, the slag based AA mortar got a higher strength as compared to the flyash based AA mortar when both type of mortar binder activated by the same NaOH and Na_2SiO_3. Increment of the water content directly reduced the strength of the AA mortar.[3,4] The compressive strength of Geopolymer mortar with the larger quantity of aggregate can be enlarged by the optimizing the amount of alkali because the increment of the proportion of the aggregate in the geopolymer mortar degrade the amount of the geopolymerization but not directly effect on the compressive strength.[5] The thickness of the geopolymer paste has affected the bonding between the aggregate and Geopolymer paste in which the thickness of the paste higher destroy by the vapor evaporation under the elevated temperature, but the paste is too thin paste cannot build the bond correctly. So, with the increase of aggregate size required to change the thickness of the paste.[6] High calcium flyash used as binder resulted in the high strength mortar up to 86 MPa, excellent volumetric stability with the higher flowability.[1] The molarity of the NaOH in the mix design of GPC directly increases the strength and also reduce the workability of the plastic concrete.[7,8] The molarity of the NaOH increase in the mix reduces the setting time of the GPC.[9] The NaOH and sodium silicate mix ratio of 1:1 use to activate the binder. It can generally achieve the strength up to the M50.[10] Alkali activated slag concrete has a higher capability in term of the corrosion resistivity to the OPC concrete. It also reduces the total embodied energy of the construction.[11]

2 | EXPERIMENTAL PROGRAM

A total of 10 numbers mix designs of the GPC cast and tested for the various properties. All the mix samples cured in two conditions ambient and oven, to analyze the effect of the curing condition on the mix designs. GPC made by the Flyash and GGBFs as a binder in 75/25 ratio by weight respectively, binders activated by the NaOH and Na_2SiO_3 at various ratio and with the different molarity of the Sodium hydroxide.

2.1 | Materials specifications

Sodium hydroxide present in the white delinquent flakes forms with the minimum purity level of 97.0% in the material sample purchased from the Fisher Scientific Delhi. Sodium silicate present in compact, tacky/slightly cloudy liquid form with the minimum assay of Na_2O titrimetric is 10.0% and SiO_2, gravimetric is 25.5–28.5% in the solution of the sample purchased from the central drug house, Delhi. Locally available m-sand or stone dust used as a fine aggregate in the mix design of the tests. Stone dust is a well-graded medium sand type of zone II. Its water absorption, specific gravity, silt content and bulk density are 1.21%, 2.62, 6%, and 1,610 kg/m^3, respectively and locally crushed 10 mm and 20 mm coarse aggregate

Uncorrected Proofs

TABLE 1 Chemical composition of flyash and GGBFS

Characteristics	SiO_2	Al_2O_3	CaO	Fe_2O_3	MgO	SO_3	LOI
Flyash (%)	45.8	21.4	13.7	12.6	1.3	1.9	0.1
GGBFS (%)	34.52	20.66	32.43	0.57	10.09	0.77	0.3

as per ASTM standards.[12–18] In the mix design, use the high performance super plasticizing admixture conplast SP430 manufactured by the Fosroc constructive solutions. Conplast SP 430 conforms with BS 5075, BS: EN 934-2 and with ASTM C494 as type A and type F, depending on the dosage used.[19] GGBFS generated by the isolation of molten slag with the help of high-pressure water jets. Separation of the molten slag prevents crystallization, so the end products found in the form of glassy and granular aggregates. That end product crushed, pulverized and separated for the multipurpose of the applications. GGBFS used in the project taken from the Bhilai steel plant, Durg, Chhattisgarh in India.[20]

Table 1 presents the chemical compositions present in the sample materials and Figure 1 shows the scanning electron microscope picture in which that shows the porous undefined shape of particles of GGBFS. Flyash is produced in the thermal power stations by the combustion of coals for the electricity generation. Flyash collects from the Gurunanak Dev thermal plant, Bhatinda, Punjab in India. Flyash is the type of class f by the classification as per the ASTM C618-33512.[21]

Table 1 presents the chemical compositions present in the sample materials and Figure 2 shows the scanning electron microscope picture in which that shows the porous spherical particles of flyash. Figures 3 and 4 show the EDS graph shows the elemental graph of the flyash and GGBFS, respectively in which shows the mineral elements present in the flyash and GGBFS samples.

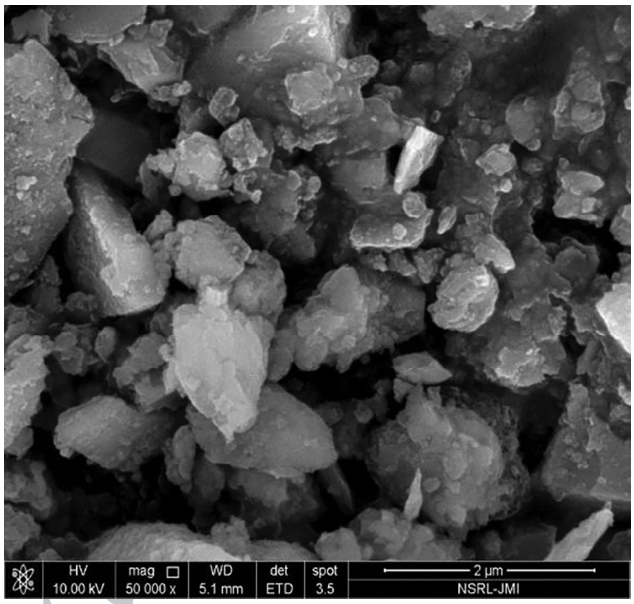


FIGURE 1 SEM image of GGBFS

2.2 | Specimens preparation

All the raw material mix in the 40-l capacity mixture for one time of the batch. The Alkaline chemical made 24 hr before the mixing, sodium hydroxide, sodium silicate and superplasticiser mix before 24 hr of casting. At the time of mixing first pour the coarse aggregate in the mixture then after pouring the sand or fine aggregate and mix properly then after pouring the chemical in the mix and at last pour the extra water as per requirement during the mixing, mix the concrete 3–5 min in the pan mixture.[22] The proper mixed concrete plastic paste cast in the molds are as cube, cylinder and beam shape. As per the Indian standards cubes dimension is 150 mm*150 mm*150 mm, cylinders size is 150 mm*300 mm as per diameter*height and prism in size of 100 mm*100 mm*500 mm as per width*height*length for the testing the mechanical properties of the GPC.[23] The specimens demold after the 3 days of casting then after cured in the oven for the 24 hr and ambient cured. Table 2 describes the all mix composition of the raw materials for the GPC.

The mix design from 1 to 5 shows the variation of the molarity of NaOH adds in the mix design. M1 shows 8 M NaOH dosage and is increases continuously in increasing order with the mix design. The optimum point in term of the strength got from the destructive test at the 14 M dosage in the mix design. After the molarity analysis in the mix design, check the alkaline ratio from 0.5 to 3.0, whereas the sodium silicate to sodium hydroxide called an alkaline ratio. The mix design from M6 to M10 shows the variation in the alkaline ratio from 0.5 to 3.0 in the increasing order.

3 | RESULTS AND DISCUSSION

In the experimental investigation analysis, the compressive strength, splitting tensile, flexural tensile strength and modulus of elasticity of the test results of the specimens of the mix designs. The specimens of all mixes tested at 7 days, 14 days, 28 days, 42 days, and 56 days after the casting.

Uncorrected

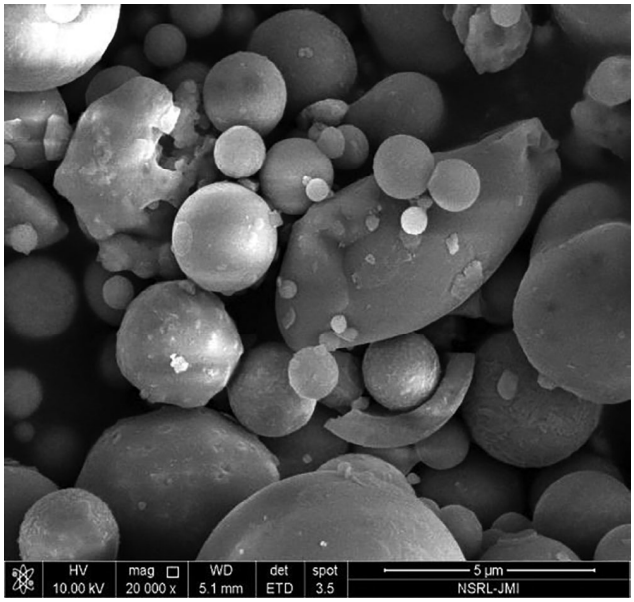


FIGURE 2 SEM image of flyash

3.1 | Effect of sodium hydroxide molar concentration

Sodium hydroxide works as a binder (flyash and GGBFS) activation in the mix designs of the GPC; the changes in the molar concentration of NaOH directly affect the mix design of the GPC. Analysis, the effect of the molarity of NaOH in mix designs of the GPC by varying 8–16 M and examine the test results of the specimens of all design mixes by the destructive strength test.

3.1.1 | Compressive strength

The compressive strength of the GPC varies with the varying molarity of the NaOH and curing condition. The compressive strength rises with the increment in the molarity of NaOH in the mix design but beyond an optimum point decreases the compressive strength in the oven-cured specimens. The highest compressive strength of oven cured 14 M mix is 34.2 N/mm^2 at 56 days but in the ambient cured specimen got at 16 M mix is 25 N/mm^2 at 56 days. Figure 5a shows the compressive strength of ambient cured specimens and Figure 5b shows the compressive strength of oven cured specimens with the variation of the molarity of NaOH in the design mix of GPC. The highest compressive strength of design mixes of 8 M, 10 M, 12 M, 14 M and 16 M are 13.9 N/mm^2, 15.5 N/mm^2, 21.8 N/mm^2, 24.0 N/mm^2, and 25.0 N/mm^2, respectively at 56 days after ambient-curing but after oven-curing compressive strength of 8 M, 10 M, 12 M, 14 M, and 16 M are 23.2 N/mm^2, 25.0 N/mm^2, 27.9 N/mm^2, 34.2 N/mm^2, and 31.0 N/mm^2, respectively at 56 days.

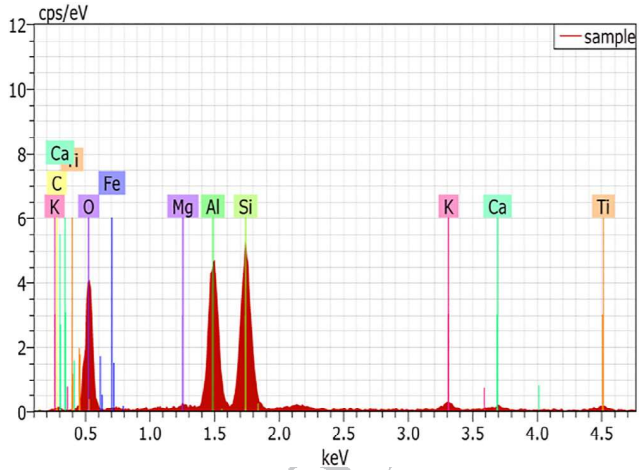


FIGURE 3 EDS graph of GGBFS

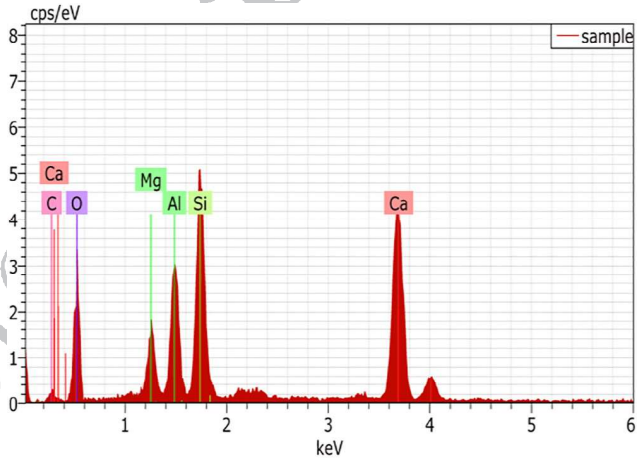


FIGURE 4 Compressive strength of (a) ambient-cured GPC and (b) oven-cured GPC

3.1.2 | Splitting strength

The splitting tensile strength test used to find the indirect tensile strength of the design mix of the GPC. Figure 6a shows the splitting strength of ambient-cured specimens of mix designs and Figure 6b shows the splitting strength of oven-cured samples of the mix designs, in which varies the molarity of NaOH in the mix design of GPC from 8 M to 16 M. In the ambient cured specimens 16 M got the highest splitting tensile strength is 3.6 N/mm^2 at 56 days, but in oven-cured samples 14 M got the most upper splitting tensile strength is 5.2 N/mm^2 at 56 days. The highest splitting tensile strength of 8 M, 10 M, 12 M, 14 M and

Color Figure - Print and Online
Color Figure - Print and Online
Uncorrected Proof

TABLE 2 Mix design

Mix	Fly ash (kg/m^3)	GGBFS (kg/m^3)	Coarse aggregate (kg/m^3)	Fine aggregate (kg/m^3)	NaOH solution (kg/m^3)	Sodium silicate (kg/m^3)	Super plasticizer (kg/m^3)	Extra water (kg/m^3)
M1	303.75	101.25	1,269	683	81.0 (8 M)	81.0	4.05	20.25
M2	303.75	101.25	1,269	683	81.0 (10 M)	81.0	4.05	20.25
M3	303.75	101.25	1,269	683	81.0 (12 M)	81.0	4.05	20.25
M4	303.75	101.25	1,269	683	81.0 (14 M)	81.0	4.05	20.25
M5	303.75	101.25	1,269	683	81.0 (16 M)	81.0	4.05	20.25
M6	303.75	101.25	1,269	683	108.0 (14 M)	54.0	4.05	20.25
M7	303.75	101.25	1,269	683	64.8 (14 M)	97.2	4.05	20.25
M8	303.75	101.25	1,269	683	54.0 (14 M)	108.0	4.05	20.25
M9	303.75	101.25	1,269	683	46.28 (14 M)	115.72	4.05	20.25
M10	303.75	101.25	1,269	683	40.5 (14 M)	121.5	4.05	20.25

Color Figure - Print and Online

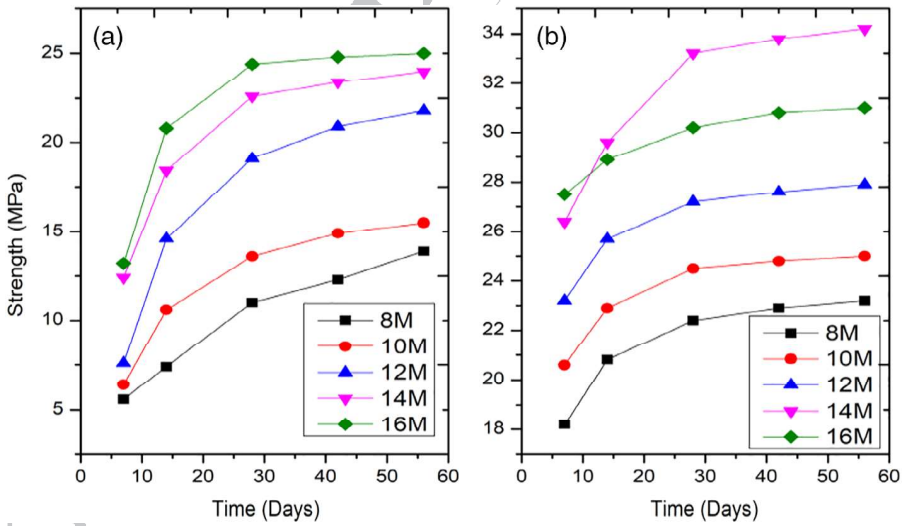


FIGURE 5 Splitting strength of (a) ambient cured GPC and (b) oven cured GPC

Color Figure - Print and Online

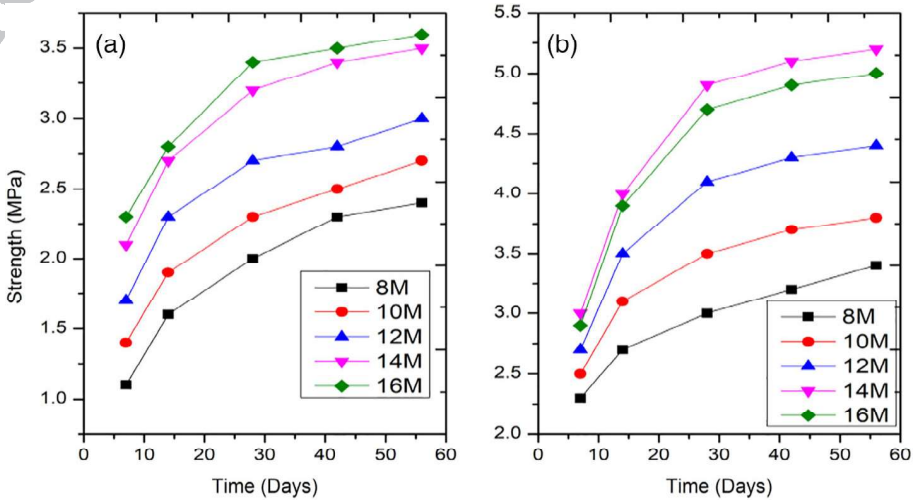


FIGURE 6 Flexural strength of (a) ambient-cured GPC and (b) oven-cured GPC

Uncorrected Proofs

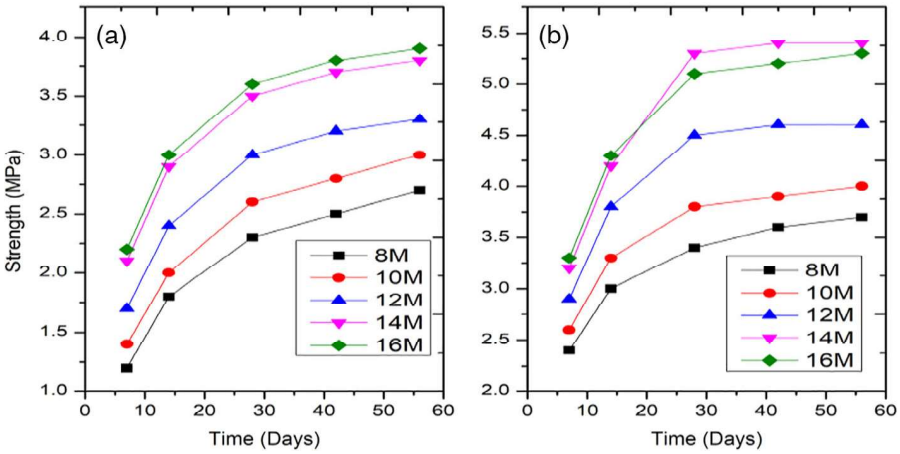


FIGURE 7 Graph between density and molarity of NaOH of GPC

16 M are 2.4 N/mm^2, 2.7 N/mm^2, 3.0 N/mm^2, 3.5 N/mm^2 and 3.6 N/mm^2 respectively at 56 days after ambient-curing but after oven-curing splitting tensile strength of 8 M, 10 M, 12 M, 14 M and 16 M are 3.4 N/mm^2, 3.8 N/mm^2, 4.4 N/mm^2, 5.2 N/mm^2 and 5.0 N/mm^2 respectively at 56 days.

3.1.3 | Flexural strength

Flexural tensile strength to find bending tension of the beam specimen. Figure 7 shows the flexural tensile strength of mixes ambient cured samples and oven cured samples, respectively. Similarly, the oven-cured samples got more flexural tensile strength as compared to the ambient-cured specimens, the highest strength of oven-cured specimens of 14 M mix is 5.4 N/mm^2 at 56 days, but in ambient-cured specimens, the most upper strength of 16 M mix is 3.9 N/mm^2. In both curing conditions, the flexural tensile strength increases with the increment of the molarity of NaOH in mix design but in the oven-cured strength decreases beyond the 14 M NaOH in mixes. The highest flexural tensile strength of 8 M, 10 M, 12 M, 14 M, and 16 M are 2.7 N/mm^2, 3.0 N/mm^2, 3.3 N/mm^2, 3.8 N/mm^2, and 3.9 N/mm^2, respectively at 56 days after ambient-curing but after oven-curing flexural tensile strength of 8 M, 10 M, 12 M, 14 M, and 16 M are 3.7 N/mm^2, 4.0 N/mm^2, 4.6 N/mm^2, 5.4 N/mm^2, and 5.3 N/mm^2, respectively at 56 days.

3.1.4 | Density, Poisson ratio, and modulus of elasticity

Figure 8 presented the density of the mix designs to the analysis of the effect of the molarity of NaOH in mix designs of GPC specimens. All the specimens of various mix designs with the variation of molarity 8–16 M cured

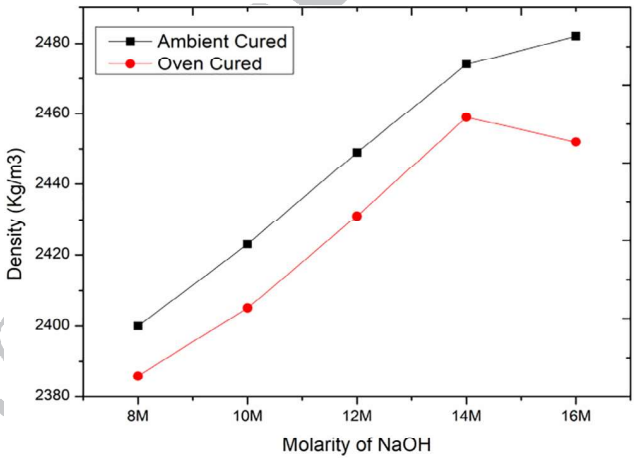


FIGURE 8 Graph between the modulus of elasticity and molarity of NaOH of GPC

under both conditions ambient and oven-curing. Analysis of the variation in results of the Poisson ratio, modulus of elasticity and density drawn out from the samples testing. Density, Poisson ratio and modulus of elasticity tested after the 28 days of the casting. Density finds out through the weight of the cube specimens of the various mixes. The Poisson ratio finds through the testing of the cylindrical specimens tested in the UTM machine by applying statistical compression loading along the longitudinal direction and got the vertical and horizontal strain. The Elastic modulus finds through the stress–strain curve made by the cyclic loading on the cylindrical specimen about one-third of the fracture load statically. The ambient cured sample has a higher density than the oven-cured GPC specimens of different mix designs. A mix of molarity of NaOH in mixes rise in the density of the design mix specimens, but after the increase beyond the 14 M slightly decreases the density of the mix designs specimens. 16 M mix designs samples got the highest average density of 2,482 kg/m^3 at 28 days of ambient

Color Figure - Print and Online
Uncorrected Proofs

cured specimens, whereas the oven-cured samples got the average density of 2,474 kg/m^3. The Poisson ratio of the ambient-cured GPC specimens are slightly excessive than oven-cured specimens, but the modulus of elasticity of the GPC mix designs higher for the oven-cured samples.

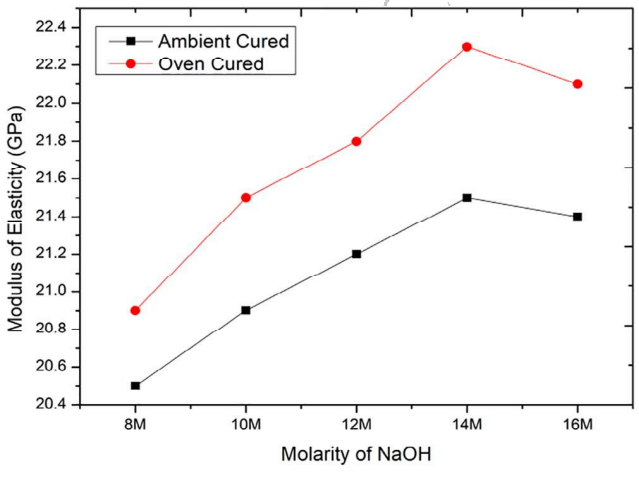
Figure 9 shows the maximum modulus of elasticity of the 14 M mix design of oven cured specimens is 22.3 GPa, whereas the ambient cured specimens modulus of elasticity is 21.5 GPa of the 14 M mix design of GPC.

3.2 | Effect of sodium silicate to sodium hydroxide (Na_2SiO_3/NaOH) ratio

Sodium silicate works as a binder (flyash and GGBFS) activation with the sodium hydroxide in the mix designs of the GPC, the changes in the ratio of Na_2SiO_3/NaOH directly affect the properties of the mix design of the GPC. Analysis, the effect of the rate of Na_2SiO_3/NaOH in mix designs of the GPC by varying the ratio 0.5–3.0 and examine the test results of the specimens of all design mixes by the destructive strength test.

3.2.1 | Compressive strength

The compressive strength of the GPC varies with the ratio of Na_2SiO_3/NaOH and curing condition. The compressive strength increases with the increment in the ratio in the mix design but beyond a point decreases the compressive strength in both curing condition samples. The highest compressive strength of oven cured 14 M mix is 35.7 N/mm^2 at 56 days and in the ambient cured specimen got 25.8 N/mm^2 at 56 days. Figure 10a shows the compressive strength of ambient cured specimens and in Figure 10b shows the compressive strength of oven cured specimens with the ratio of Na_2SiO_3/NaOH in design mix of GPC. The highest compressive strength of design mixes of ratio 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0 are 20.3 MPa, 24.0 MPa, 25.0 MPa, 25.4 MPa, 25.8 MPa, and 23.4 MPa respectively at 56 days after ambient-curing but after oven-curing compressive strength of the ratio 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0 are 24.6 MPa, 34.2 MPa, 34.7 MPa, 35.1 MPa, 35.7 MPa, and 32.5 MPa, respectively at 56 days.



Color Figure – Print and Online

FIGURE 9 Compressive strength of (a) ambient-cured GPC and (b) oven-cured GPC

3.2.2 | Splitting strength

The splitting tensile strength test to find the indirect tensile strength of the design mix of the GPC. Figure 11 shows the splitting tensile strength of ambient-cured specimens and oven cured samples of mix designs, in which varies the ratio of Na_2SiO_3/NaOH in the mix design of GPC from 0.5 to 3.0. In the both curing conditions specimens got the highest splitting tensile strength at the ratio of 2.5 and the uppermost strength of ambient cured and oven cured specimens are 4.0 MPa, 5.2 MPa respectively at 56 days in the mixes of GPC. The highest splitting tensile strength of design mixes of ratio 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0 are 2.7 MPa, 3.5 MPa, 3.6 MPa, 3.8 MPa, 4.0 MPa, and 3.6 MPa, respectively at 56 days after ambient-curing but after oven-curing compressive strength of the ratio 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0 are 3.5 MPa, 5.2 MPa, 5.0 MPa, 5.1 MPa, 5.2 MPa, and 5.0 MPa, respectively at 56 days.

3.2.3 | Flexural strength

Flexural tensile strength to find bending tension of the beam specimen. Figure 12 shows the flexural tensile strength of mixes ambient cured samples and oven cured samples, respectively. Similarly, the oven-cured samples got more flexural tensile strength as compared to the ambient-cured specimens, the highest strength of oven-cured specimens of the ratio of 2.5 in the mix is 5.6 MPa at 56 days, but in ambient cured samples the most upper strength of the ratio 2.5 in the mix is 4.0 MPa. In both curing conditions, the flexural tensile strength rises with the increment of the rate of sodium silicate to sodium hydroxide in the mix design but in the oven-cured strength decreases beyond the ratio of 2.5 in design mixes. The highest flexural tensile strength of design mixes of ratio 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0 are 3.0 MPa, 3.8 MPa, 4.0 MPa, 4.0 MPa, 4.0 MPa, and 3.9 MPa, respectively at 56 days after ambient-curing but after

Uncorrected Proofs

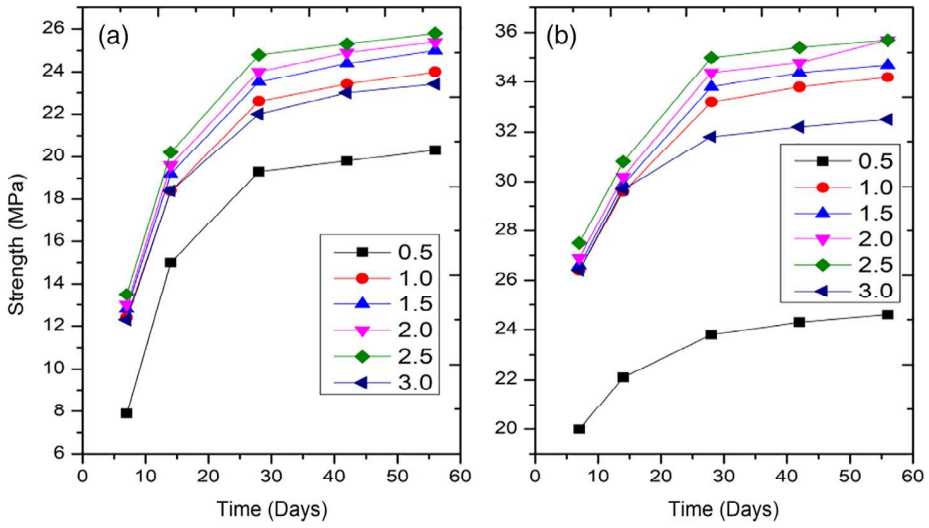


FIGURE 10 Splitting strength of (a) ambient cured GPC and (b) oven-cured GPC

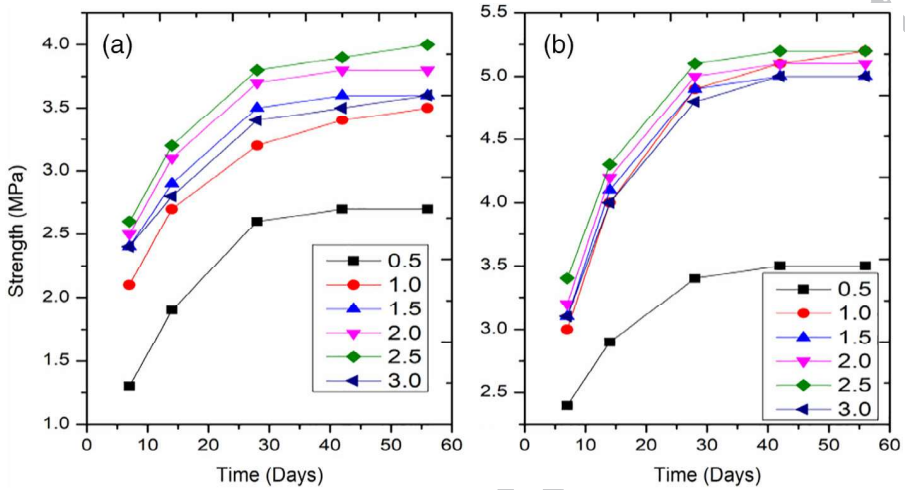


FIGURE 11 Flexural strength of (a) ambient-cured GPC and (b) oven-cured GPC

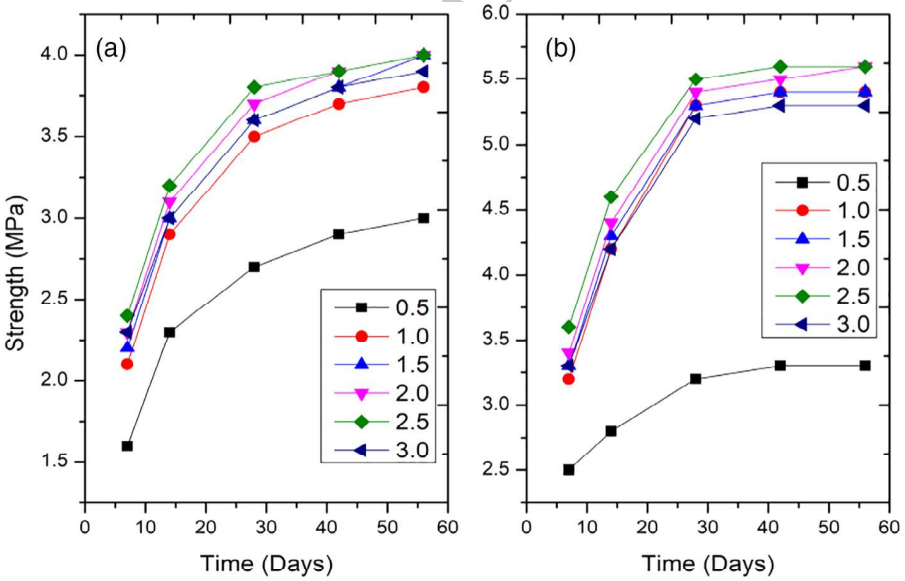
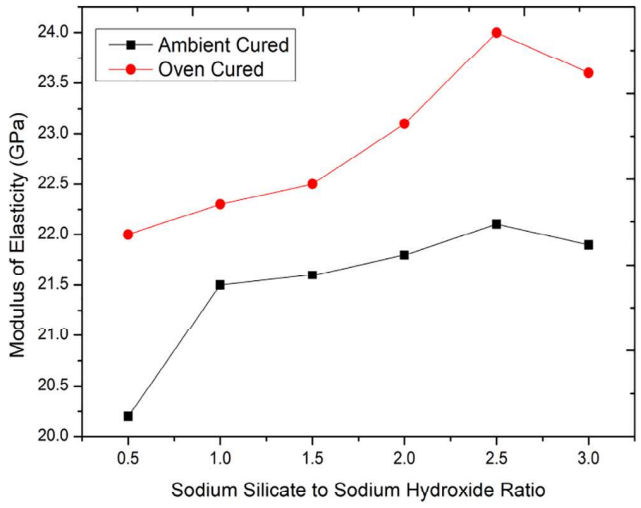


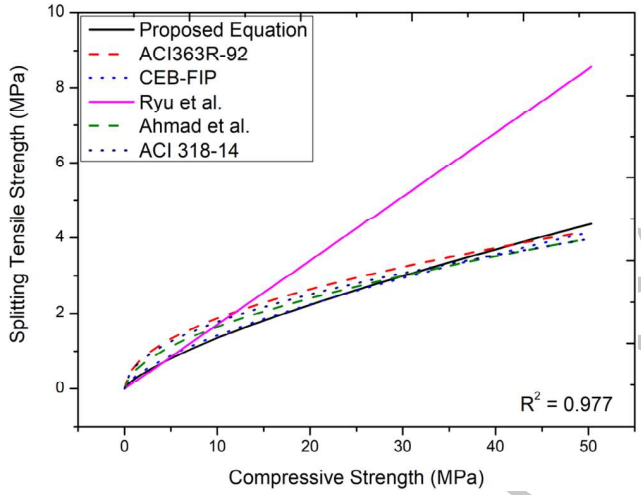
FIGURE 12 Graph between the density and Na_2SiO_3/NaOH ratio in GPC

Color Figure - Print and Online
Color Figure - Print and Online
Color Figure - Print and Online
Proofs



Color Figure - Print and Online

FIGURE 13 Graph between the modulus of elasticity and Na_2SiO_3/NaOH ratio in GPC



Color Figure - Print and Online

FIGURE 14 Relationship between splitting tensile strength and compressive strength

oven-curing compressive strength of the ratio 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0 are 5.4 MPa, 5.4 MPa, 5.6 MPa, 5.6 MPa, 5.3 MPa, and 5.0 MPa, respectively at 56 days.

3.2.4 | Density, Poisson ratio, and modulus of elasticity

Figure 13 represent the effect of the ratio of Na_2SiO_3/NaOH on the density of mix design of GPC; Ambient cured specimens have higher density compared to oven-cured samples and got the highest density are 2,494 kg/m^3 at the ratio of 2.5 of ambient cured specimens. Figure 14 represents the effect of the ratio of Na_2SiO_3/NaOH on the modulus of elasticity of mix design of GPC. Oven cured samples have an excessive modulus of elasticity as compared to the ambient cured samples and got the highest modulus of elasticity at the ratio of 2.5 are 24 GPa of oven cured samples.

4 | CORRELATION BETWEEN THE MECHANICAL PROPERTIES

After the long-term experimental investigation, GPC mechanical properties show the correlation to each other. Based on the results out from the destructive experiment test of various mixes, proposed the correlation equation among the compressive strength, splitting strength and flexural strength.

4.1 | Relationship between splitting tensile strength and compressive strength

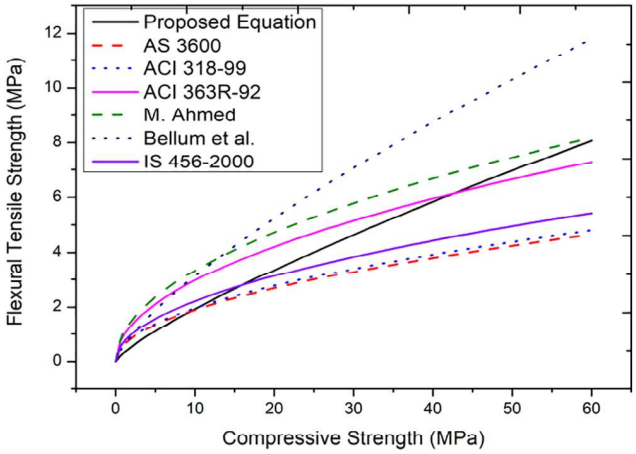
The splitting tensile strength of the GPC measured through the cylindrical specimen of 150 mm dia × 300 mm height by a static load on the transverse side on the specimens and compressive strength from the 150 mm × 150 mm size cube test in the UTM machine in the laboratory. From the results, a correlation Equation (1) proposed between the splitting and compressive strength and compared with the other correlation equations given by the code or other authors. The ACI committee gives Equation (2) in the state of the art of high strength concrete[24] and a committee of euro code international gives Equation (3) in the model code 1990.[25] Equation (4) introduced by an ACI standard and report for building code requirement for structural concrete[26] and Ahmed et al. proposed Equation (5) in the international journal of structural engineering in 2014.[27] Ryu et al. introduce Equation (6) of the correlation of flyash based GPC in the journal construction and building material in 2013.[10] All the correlation are very similar and shows that splitting strength are around to12-13% of the compressive strength of the mixes specimens. Figure 15 graph shows the various equations and compared with the proposed equation of empirical correlation.

$$\text{Proposed Eq.}: f_{st} = 0.228 f_c^{0.82} \quad (1)$$

$$\text{ACI363R} - 92: f_{st} = 0.59\sqrt{f_c} \quad (2)$$

$$\text{CEB} - \text{FIP}: f_{st} = 0.301 f_c^{0.67} \quad (3)$$

Uncorrected Proofs



Color Figure : Print and Online

FIGURE 15 Relationship between flexural strength and compressive strength

$$\text{ACI } 318-14: f_{st} = 0.56 f_c^{0.5} \quad (4)$$

$$\text{Ahmed et al.}: f_{st} = 0.462 f_c^{0.55} \quad (5)$$

$$\text{Ryu et al.}: f_{st} = 0.17 f_c^{0.75} \quad (6)$$

4.2 | Relationship between flexural strength and compressive strength

Flexural strength of the concrete is fundamental in structural engineering because it shows the tensile strength of concrete and bending behavior of the specimens. From the experimental results, proposed the correlation Equation (7) between the Flexural strength and compressive strength of the mixes and compared to the other correlation Equations (8)–(13) given by the codes and authors. ACI committee introduces Equation (8) in the publication of the state of the art of high strength concrete.[24] A standard report presented on the code for the design of concrete in the structure introduces Equation (9) of correlation.[26] Australian code for the reinforced concrete design presented the correlation Equation (10), and the Indian code introduced Equation (11) between the flexural strength and compressive strength.[28,29] Ahmed et al. proposed Equation (12) in the international journal of structural engineering and Bellum et al. also proposed Equation (13) between the correlation of flexural strength and compressive strength.[27,30] Figure 16 shows the proposed equation shows that the flexural strength is around 15–16% of the compressive strength of the same mix of GPC.

$$\text{Proposed Equation}: f_{fs} = 0.3066 f_c^{0.799} \quad (7)$$

$$\text{ACI363R} - 92: f_{fs} = 0.94\sqrt{f_c} \quad (8)$$

$$\text{ACI } 318-99: f_{fs} = 0.62\sqrt{f_c} \quad (9)$$

$$\text{AS } 3600: f_{fs} = 0.6\sqrt{f_c} \quad (10)$$

$$\text{IS } 456-2000: f_{fs} = 0.7\sqrt{f_c} \quad (11)$$

$$\text{M.Ahmed}: f_{fs} = 1.055\sqrt{f_c} \quad (12)$$

$$\text{Bellum et al.}: f_{fs} = 0.57 f_c^{0.74} \quad (13)$$

where f_c are characteristic strength in MPa, f_{fs} are flexural strength in MPa.

5 | CONCLUSIONS

The conclusion drained through the testing of the various mix designs specimens in the form of cube cylinder and prism molds by the destructive and nondestructive testing of ambient-cured and oven-cured specimens, in which the molarity of NaOH and Na_2SiO_3/NaOH varying in the design mix of GPC.

- The compressive strength rises with the enlargement in molarity of NaOH in the mix design but beyond a point decreases the compressive strength in the oven-cured specimens. The highest compressive strength of oven cured 14 M mix is 34.2 N/mm^2 at 56 days but in the ambient-cured specimen got at 16 M mix is 25 N/mm^2 at 56 days.
- In the ambient cured specimens 16 M got the highest splitting tensile strength is 3.6 N/mm^2 at 56 days, but in oven-cured samples 14 M got the most upper splitting tensile strength is 5.2 N/mm^2 at 56 days.
- Similarly, the oven-cured specimens got more flexural tensile strength as compared to the ambient-cured specimens, the highest strength of oven-cured specimens of 14 M mix are 5.4 N/mm^2 at 56 days, but in ambient-cured specimens, the most upper strength of 16 M mix is 3.9 N/mm^2.
- The compressive strength enlarges with the increment in the ratio in the mix design but beyond a point decreases the compressive strength in both curing condition samples. The highest compressive strength of oven cured 14 M mix is 35.7 N/mm^2 at 56 days and in

Uncorrected Proofs

the ambient cured specimen got 25.8 N/mm^2 at 56 days.

- In the both curing conditions specimens got the highest splitting tensile strength at the ratio of 2.5 and the uppermost strength of ambient cured and oven cured specimens are 4.0 MPa, 5.2 MPa respectively at 56 days in the mixes of GPC.
- In both curing conditions, the flexural tensile strength rises with the increment of the rate of sodium silicate to sodium hydroxide in mix design but in the oven-cured strength decreases beyond the ratio of 2.5 in design mixes.
- Based on results, proposed the correlation Equations (1) between the splitting strength and compressive strength are $f_{st} = 0.228 f_c^{0.82}$ and Equation (7) between the flexural strength and compressive strength are $f_{fs} = 0.3066 f_c^{0.799}$.

ACKNOWLEDGMENTS

This work supported by the Civil Engineering Department, Delhi Technological University funded by the Delhi Government. Moreover, the help of Sandeep Mishra for XRD analysis and Ankesh Kumar and Jawed Alam for the another in the work that is related to the central instrumentation facility of the Delhi Technological University.

ORCID

Manvendra Verma https://orcid.org/0000-0002-0840-841X

REFERENCES

1. Chindaprasirt P, Chareerat T, Sirivivatnanon V. Workability and strength of coarse high calcium fly ash geopolymer. Cem Concr Compos. 2007;29:224–229.
2. Nath P. School of Civil and Mechanical Engineering Study of Fly Ash Based Geopolymer Concrete Cured in Ambient Condition; 2014.
3. Yang K-H, Song J-K, Lee K-S, Ashour AF. Flow and compressive strength of alkali-activated mortars. ACI Mater J. 2009;106 (1):50–58.
4. Yang K-H, Song J-K. Workability loss and compressive strength development of cementless mortars activated by combination of sodium silicate and sodium hydroxide. J Mater Civ Eng. 2009;21:119–127.
5. Temuujin J, Van Riessen A, Mackenzie KJD. Preparation and characterisation of fly ash-based geopolymer mortars. Construct Build Mater. 2010;24(10):1906–1910.
6. Shu-guang HU, Jing WU, Wen Y, Yong-Jia HE. Preparation and properties of geopolymer-lightweight aggregate refractory concrete. J Cent South Univ Technol. 2009;16:914–918.
7. Patankar SV, Ghugal YM, Jamkar SS. Mix design of fly ash-based geopolymer concrete. Advances in structural engineering: Materials, volume three, 2015.
8. Patankar SV, Ghugal YM, Jamkar SS. Effect of concentration of sodium hydroxide and degree of heat curing on fly ash-based geopolymer mortar effect of concentration of sodium hydroxide and degree of heat curing on fly ash-based geopolymer mortar. Indian J Mater Sci. 2014;2014:938789.
9. Lee NK, Lee HK. Setting and mechanical properties of alkali-activated fly ash/slag concrete manufactured at room temperature. Construct Build Mater. 2013;47:1201–1209.
10. Ryu GS, Lee YB, Koh KT, Chung YS. The mechanical properties of fly ash-based geopolymer concrete with alkaline activators. Construct Build Mater. 2013;47:409–418.
11. Awoyera P, Adesina A. A critical review on the application of alkali-activated slag as a sustainable composite binder. Case Stud Constr Mater. 2019;11:e00268.
12. B. Statements. Standard Test Method for Surface Moisture in Fine Aggregate 1. pp. 1–3, 2019.
13. C. Ag, B. Statements. Standard Test Method for Relative Density (Specific Gravity) and Absorption of Coarse. pp. 1–5, 2019.
14. B. Statements. Standard Test Method for Materials Finer than 75-μm (No. 200) Sieve in Mineral Aggregates by Washing 1, no. 200, pp. 7–10, 2019.
15. C. Ag, B. Statements, T. A. Evaluation. Standard Test Method for Bulk Density ('Unit Weight') and Voids in Aggregate 1. pp. 1–5, 2019.
16. B. Statements, C. Ag. Standard Test Method for Sieve Analysis of Fine and Coarse Aggregates 1. pp. 1–5, 2019.
17. S. Container. Standard Test Method for Moisture in Mineral Aggregate Used on Built-Up Roofs 1, Vol. 89, no. Reapproved, 2017, pp. 1–2, 2019.
18. C. Ag, B. Statements, W. Pycnometer. Standard Test Method for Relative Density (Specific Gravity) and Absorption of Fine, Vol. I, pp. 1–6, 2019.
19. C. Aggregates. Standard Specification for Chemical Admixtures for Concrete 1. pp. 1–10, 2019.
20. P.-I. I. Metals. Yearbook, 2015, Vol. 2015, no. 0712, pp. 1–10, 2016.
21. T. Fly, C. Ag. Standard Specification for Coal Fly Ash and Raw or Calcined Natural Pozzolan for Use, pp. 1–5, 2019.
22. C. Ag, T. C. Vertically, M. Cabinets, M. Rooms, M. H. Concrete, T. A. Evaluation, Standard Practice for, pp. 1–8, 2019.
23. I. Standard. Specification for Moulds for Use in Tests of Cement and Concrete, no. 1952, 2008.
24. Russell HG, Anderson AR, Banning JO, et al. State-of-the-Art Report on High-Strength Concrete Reported by ACI Committee 363, vol. 92, no. Reapproved, 1997.
25. ceb-fip-model-code-1990-design-code.pdf, 1990.
26. A. C. I. Committee. Building Code Requirements for Structural Concrete (ACI 318-14), 2014.
27. Ahmed M, El Hadi KM, Mallick J, Hasan MA, Ahmed A. Evaluating the co-relationship between concrete flexural tensile strength and compressive strength. Int J Struct Eng. 2014;5(2): 115–131.
28. C. Concrete. Reinforced Concrete Design in accordance with AS 3600—2009, 2009.
29. O. F. I. Standards. Plain and Reinforced Concrete – Code of Practice, no. July 2000.
30. Bellum RR, Muniraj K, Rama S, Madduru C. Empirical relationships on mechanical properties of class-F fly ash and GGBS based geopolymer concrete. Ann Chim – Sci Matér. 2019;43(3): 189–197.

Uncorrected Proofs

AUTHOR BIOGRAPHIES



Manvendra Verma,
Department of Civil Engineering,
Delhi Technological University,
Delhi, India.
Email: manvendraverma_2k16phdc-e05@dtu.ac.in



Nirendra Dev,
Department of Civil Engineering,
Delhi Technological University,
Delhi, India.
Email: nirendradev@dce.ac.in

How to cite this article: Verma M, Dev N. Sodium hydroxide effect on the mechanical properties of flyash-slag based geopolymer concrete. *Structural Concrete*. 2020;1–12. https://doi.org/10.1002/suco.202000068

Uncorrected Proofs

geosciences

MDPI

Article

Spatial Landslide Risk Assessment at Phuentsholing, Bhutan

Abhirup Dikshit [1], Raju Sarkar [2,3,*], Biswajeet Pradhan [1,4], Saroj Acharya [2] and Abdullah M. Alamri [5]

1 Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), University of Technology Sydney, Sydney, NSW 2007, Australia; abhirupdikshit@gmail.com (A.D.); biswajeet.pradhan@uts.edu.au (B.P.)
2 Center for Disaster Risk Reduction and Community Development Studies, Royal University of Bhutan, Phuentsholing 21101, Bhutan; sarojzacharya12@gmail.com
3 Department of Civil Engineering, Delhi Technological University, Bawana Road, Delhi 110042, India
4 Department of Energy and Mineral Resources Engineering, Sejong University, Choongmu-gwan, 209 Neungdong-ro, Gwangjin-gu, Seoul 05006, Korea
5 Department of Geology & Geophysics, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia; alamri.geo@gmail.com
* Correspondence: rajusarkar@dce.ac.in

Received: 13 February 2020; Accepted: 6 April 2020; Published: 7 April 2020



Abstract: Landslides are one of the most destructive and most recurring natural calamities in the Himalayan region. Their occurrence leads to immense damage to infrastructure and loss of land, human lives, and livestock. One of the most affected regions is the Bhutan Himalayas, where the majority of the landslides are rainfall-induced. The present study aims to determine the hazard and risk associated with rainfall-induced landslides for the Phuentsholing region located in the southwestern part of the Bhutan Himalayas. The work involves developing a landslide risk map using hazard and vulnerability maps utilizing landslide records from 2004 to 2014. The landslide hazard map was generated by determining spatial and temporal probabilities for the study region. The spatial probability was computed by analyzing the landslide contributing factors like geology, slope, elevation, rainfall, and vegetation based on comprehensive field study and expertise about the area. The contributing factors were divided into various classes and the percentage of landslide occurrence under each class was calculated to understand its contributing significance. Thereafter, a weighted linear combination approach was used in a GIS environment to develop the spatial probability map which was multiplied with temporal probabilities based on regional rainfall thresholds already determined for the region. Consequently, vulnerability assessment was conducted using key elements at risk (population, land use/land cover, proximity to road, proximity to stream) and the weights were provided based on expert judgment and comprehensive field study. Finally, risk was determined and the various regions in the study area were categorized as high, medium, and low risk. Such a study is necessary for low-economic countries like Bhutan which suffers from unavailability of extensive data and research. The study is conducted for a specific region but can be extended to other areas around the investigated area. The tool can serve as an indicator for the civil authorities to analyze the risk posed by landslides due to the rapid infrastructure development in the region.

Keywords: landslide risk; GIS; hazard; Bhutan

1. Introduction

Landslides are considered one of the most prominent and devastative natural disasters leading to immense loss of human lives and infrastructure [1]. The triggering factor for the majority of the

landslides is either rainfall or earthquake of which rainfall-triggered landslides can be categorized as very high and intense. The Himalayan region experiences very high rainfall and contributes over 70% of global fatal landslides [2]. Landslides in the Himalayas occur mainly due to monsoonal precipitation leading to huge financial loss and often disrupting the only livelihood of people [3,4]. The global fatal landslide database developed by [2] shows a total of 64 deaths in Bhutan due to nonseismic landslide events from 2004 to 2017, but the actual number of deaths is expected to be very high. The menace is expected to escalate because of the cutting of hills and deforestation emphasizing infrastructure development linked to a rise in population.

The studies on rainfall-induced landslides are quite varied and focus on several aspects like threshold modeling [5,6], monitoring systems [4,7], and computational techniques [8], along with GIS approach. Apart from landslide forecasting, landslide mitigation is important to understand the landslide hazard and the associated risk. Several works have been carried out to develop risk maps using the topography of the area susceptible to landslides [9–11]. Others have used ground conditions like the combined effect of rainfall and topography of the region to develop risk maps (South Italy [12], San Francisco Bay area [13]). Ho et al. [14] and Wong [15] used population density along with the natural contributing landslide factors to develop risk maps. Hadmoko et al. [16] utilized a vector-based GIS method to analyze several factors affecting landslides to develop hazard maps of the Central Java region. Pradhan and Abdulwahid [17] developed landslide risk maps using logistic regression and multihazard cases developed by considering precipitation data. Althuwaynee and Pradhan [11] developed a semi-quantitative landslide risk assessment using GIS-based analysis for Malaysia and were able to successfully show the predictive capability of the model. Liu and Miao [18] determined the landslide hazard, vulnerability, and risk in China by using certainty factor and a logistic regression model. Goorchi et al. [19] developed a new technique to analyze the risk linked to a rock slope, including the probability of failure and vulnerability rates. Saldivar-Sali and Einstein [20] proposed a regional landslide risk rating system in the Philippines. Aljeano et al. [21] developed a similar approach to analyze rock fall risk in quarries and the results depicted close correlation to the experimental work conducted for unstable slopes typically used for road construction. The effect of soil properties on slope instability has also been studied by several researchers using reliability-based optimization techniques, wherein the spatial variability related to shear parameters of soil is analyzed [22,23]. Cardarilli et al. [24] conducted an investigation to understand the territorial resilience towards landslide in the Southern Italy region by analyzing the correlations among the causative factors with an aim to reduce loss of lives and economic damage.

All these aforementioned studies have one thing in common, that is, the availability of sufficient data and prior research work. The present study has been performed for the Bhutan Himalayas, which lacks in both aspects. The research conducted in the past is so minimal that a Scopus search using the keywords “landslide*” and “Bhutan” resulted in only 36 documents. The publication of the latest landslide global disaster database has led several researchers to focus on this area and in the recent years, various research works have been conducted to better understand the landslides in the region. The majority of the landslide incidences are triggered due to rainfall, therefore we have focused solely on rainfall-induced landslides. The previous research in Bhutan has been towards forecasting, which led to the development of several thresholds using various statistical [3,25,26] and probabilistic [27] approaches. However, landslide mitigation and understanding the elements at risk are yet to be conducted for any part of Bhutan. Therefore, the present study is a preliminary step towards landslide risk assessment with an aim to perform a complete and comprehensive risk assessment in the future. The study was focused only on the Phuentsholing region, primarily due to two main reasons, the first being its significance as a major trade route and the second being the availability of sufficient data to conduct risk analysis. The work presented reflects a very simple but effective approach to determine landslide risk by computing hazard and vulnerability. Hazard comprises three different components (spatial, temporal, and magnitude) and these are multiplied directly, as an independent relationship is assumed between the components [28,29]. However, the unavailability of a detailed landslide inventory

map results in the exclusion of the magnitude component. In the present study, the spatial component was calculated using a weighted sum analysis based on landslide contributing factors which was multiplied with the temporal probabilities determined based on regional rainfall thresholds [26] to develop the landslide hazard map. The predisposing factors for spatial aspect were identified based on a comprehensive field study conducted by the authors and examining the historical landslide locations in the study area. Landslide vulnerability was computed using weighted overlay analysis on key elements at risk by providing weights based on expert opinion. Finally, a landslide risk map was developed by combining both the hazard and vulnerability maps. Such an approach is useful for a country like Bhutan which lacks a proper landslide mitigation taskforce along with an operational early warning landslide system [30]. The present study would give an opportunity to develop risk assessment tools to analyze the slide activity and provide recommendations for mitigation purposes.

2. Study Area

The study area is Phuentsholing, situated in southwestern Bhutan, and is located from 26°50′ N to 26°59′ N latitude and 89°15′ E to 89°28′ E longitude covering an area of around 131.7 km^2. The area is further divided into six subdivisions (known as chiwog), which are Lingdaen, Pachhu, Dzongkhag Tromde, Dophugchen Wangduegatshel, Deling Marpji, andChong Geykha Dophulakha (Figure 1).

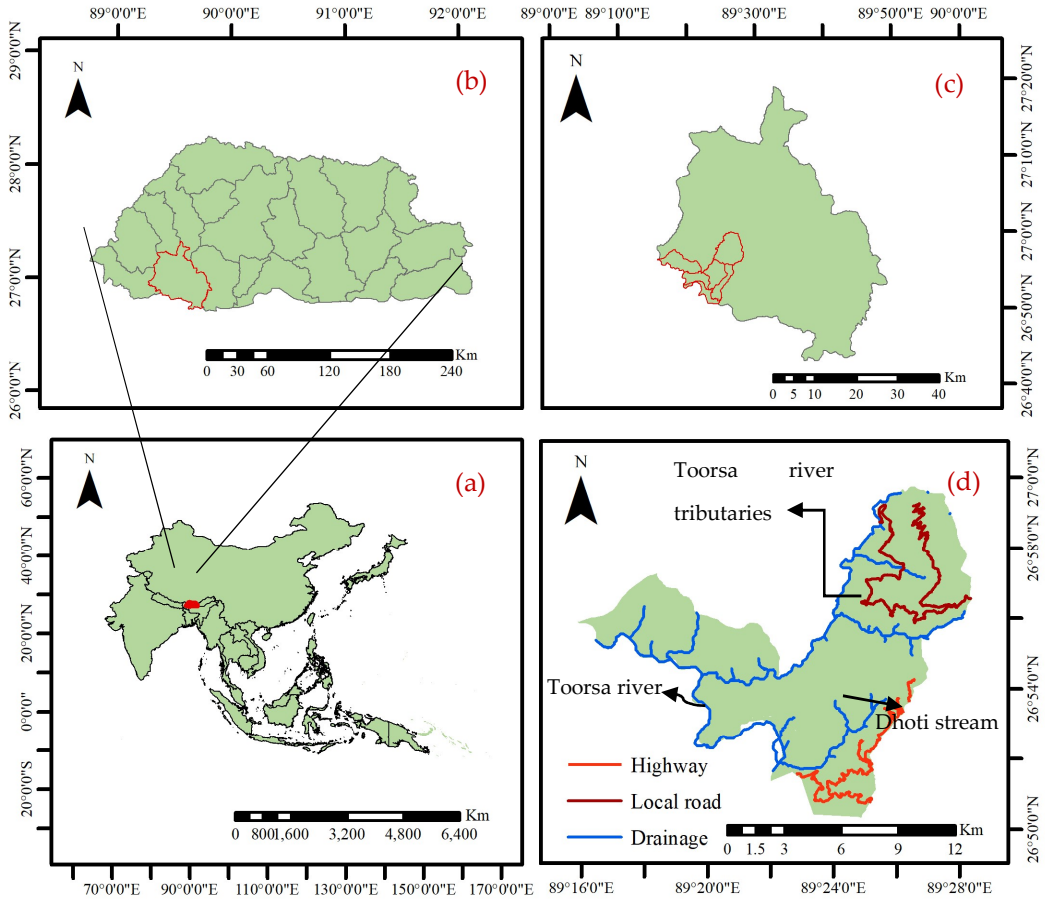


Figure 1. Location of the study region showing (**a**) Bhutan in Southeast Asia, (**b**) Bhutan highlighting Chukha Dzongkhag, (**c**) Chukha Dzongkhag highlighting Phuentsholing region, and (**d**) territorial boundaries of the study region.

Landslides in the study region are mainly due to intense monsoon rainfall, weak geology, and anthropogenic activities. The landslide typology in the study area can be characterized as rock fall, rockslide, debris flow, debris slide, and earth slide [24,25]. Gansser [31] described the regional

geological setting of the Bhutan Himalayas and characterized the majority of the country as part of the Himalayan Crystalline Complex. The areas that are most prone to landslides are heavily fractured and weathered rocks of phyllite, slate, and schist that contain high amounts of clay [26,27]. The study region has been identified as geologically weak with the presence of several shear zones, including weathered and folded phyllite rocks turning to clay [25].

The Phuentsholing-Thimphu highway, which is a part of the study region, is a vital lifeline as the majority of the import and export goods of the country are carried out through this route. The landslides lead to blockage of roads at several sections leading to a considerable amount of economic losses in repair and maintenance every year. The highway was the first motorable road of the country, constructed in 1961, and it crosses through areas susceptible to landslide, subsidence, erosion, and rockfalls. The instability in certain sections of the road has been as long as 20–50 years old. A 2005 report on the landslide investigations along the highway showed that a landslide at a 360 m section of the highway between the 17 km and 18 km point, known as Sorchen landslide, was so severe that the road alignment, which had four bends, eventually was left with only one bend over time. Figure 2 depicts two damage pictures taken during landslide events in Chukha Dzongkhag along the Phuentsholing-Thimphu Highway after the 2017 monsoon.



Figure 2. Landslide damage along Phuentsholing-Thimphu Highway (after 2017 monsoon).

3. Data Used

The understanding of various causative factors and preparation of thematic layers are necessary to develop hazard and vulnerability maps. In the present study, a risk map was developed by analyzing the hazard contributory factors (geology, slope, elevation, vegetation, rainfall) with vulnerability factors (land use, population, proximity to road and stream). Hazard has been defined as a probability of a specific danger likely to occur in a given time period whereas risk refers to hazard multiplied by the potential loss of life [20].

Geology is one of the most significant contributory factors to landslide initiation [32]. The region has four significant geological groups, which can be subcategorized into a total of seven types, which are: (i) Baxa Group (Phuentsholing Formation, Pangsari Formation), (ii) Daling–Shumar Group (Daling Formation, Orthogneiss unit, Shumar Formation), (iii) Lesser Himalayan Zone (Jaishidanda Formation), and iv) structurally lower Greater Himalayan section (lower metasedimentary unit) (Figure 3). The Baxa Group can be characterized as heavily fractured and weathered phyllites, slates, and schists [25]. The Daling–Shumar Group comprises tan weathering, cliff-forming quartzite and muscovite-biotite schist and phyllite [33]. The majority of the settlement area is dominated by the Baxa formation group wherein the thrust fault, strike, and dip of bedding and foliation are more concentrated. Similarly, such features were also delineated in the structurally lower Greater Himalayan Zone.

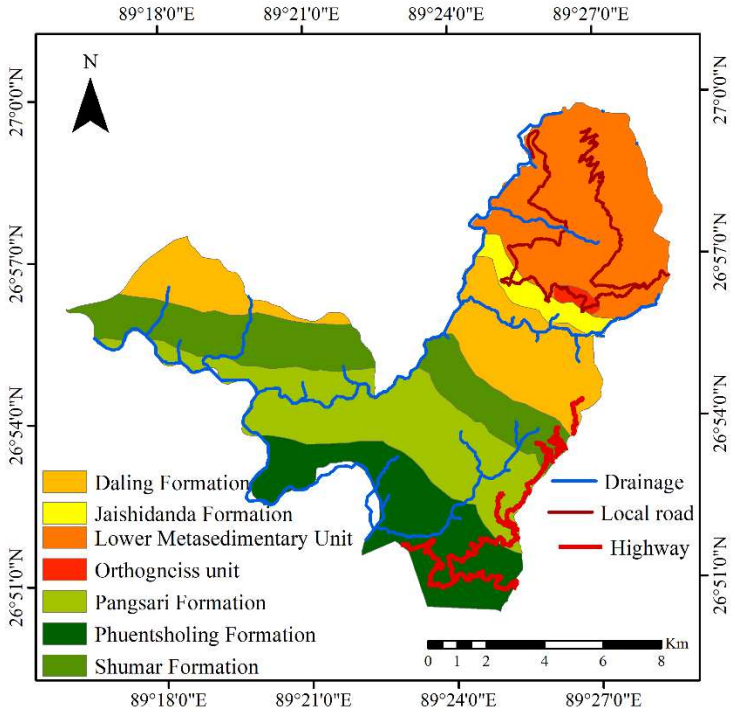


Figure 3. Geological map of the study area.

The other important factors are slope and elevation, which vary from 0°–75° and 183–2500 m, respectively. The majority of the study region (>80%) falls under 45° suggesting that the area is situated on a relatively flat plateau. More than half of the region has an elevation of less than 600 m and only 5% of the region is under high elevation zone (1500–2500 m).

The vegetation map was derived using the Normalized Difference Vegetation Index (NDVI) from the Landsat image (2017) and was classified into vegetated (86.76%) and nonvegetated areas (13.24%) (Figure 4c). NDVI always generates a value between −1 and +1 with high positive NDVI values (green) meaning high vegetation, while water usually has negative NDVI values (yellow and red), with urban features usually having values near zero. The maps were developed using the image analysis toolbar in ArcGIS 10.4.

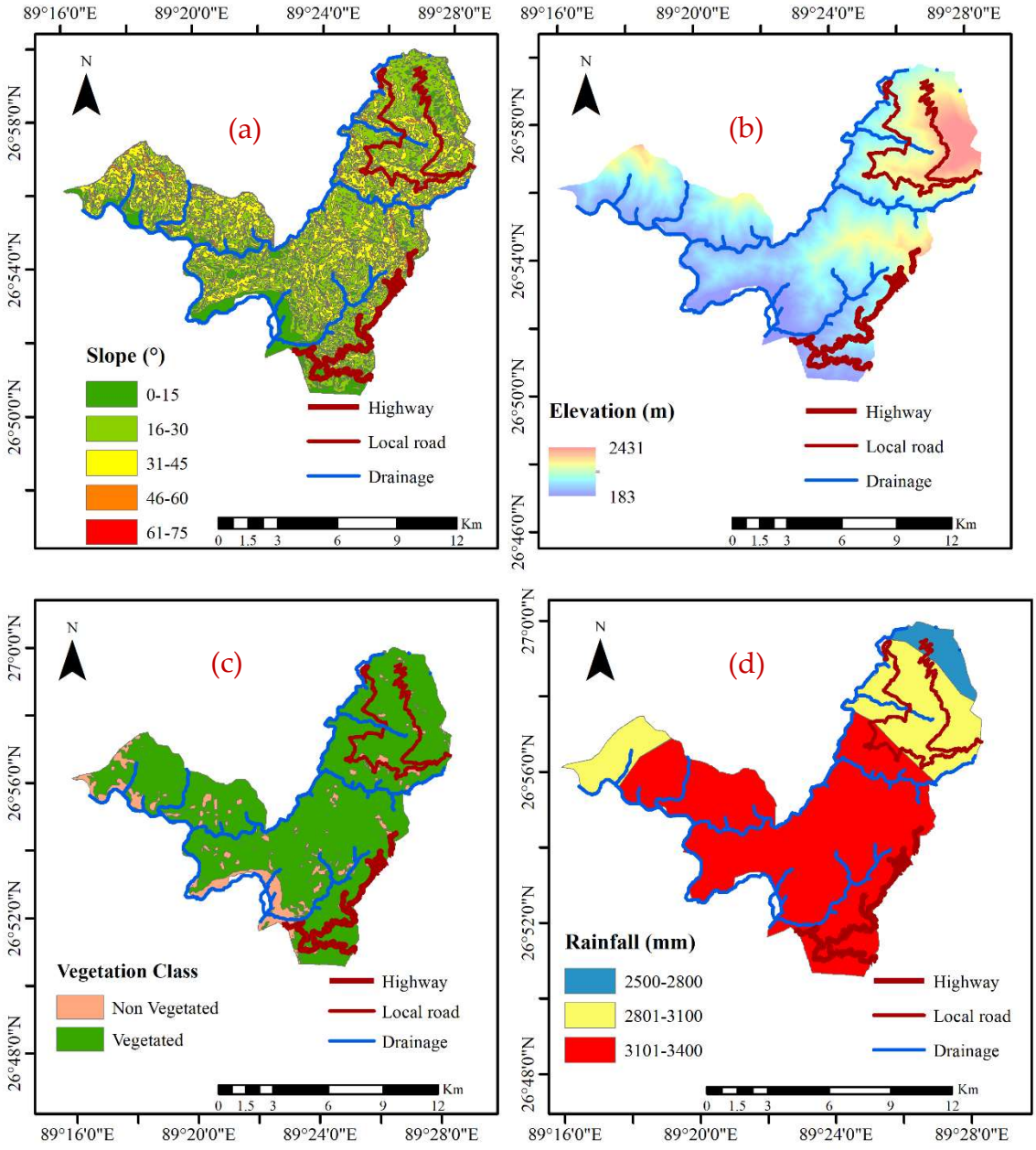


Figure 4. (**a**) Slope, (**b**) elevation, (**c**) vegetation, and (**d**) rainfall map of the study area.

The study area is located in the subtropical and temperate climatic zone, which receives very high annual rainfall up to 6000 mm. As per the Köppen climate classification [34], the climate is subtropical monsoonal receiving 10 times more rainfall during the wettest summer month as compared to dry winter months. The contribution of monsoonal rainfall (June–September) is 80% of the annual rainfall. However, pre- and post-monsoonal months can also suffer from heavy rainfall due to local convection and tropical cyclone activity over the Bay of Bengal [35], with January and February being relatively drier [25]. Figure 4d shows the rainfall map based on the annual rainfall measurements from 2004 to 2014 collected from the National Center for Hydrology and Meteorology of the Royal Government of Bhutan.

A total of 64 landslides occurred from 2004 to 2014, which was collected from Project Dantak, the Border Roads Organization, Government of India. The data included the coordinates of the landslide events which were collected from the scarp along with the dates of occurrences. The number of landslide events was counted as a single landslide event if multiple events occurred on the same

day and coordinates were selected where maximum damage or loss of land prevailed. Table 1 shows the number of landslides in each class of landslide contributing factors.

Table 1. Number of landslides under various factors.

Factor		Area (km^2)	Number of Landslides
Geology	Pangsari Formation	28.4	20
	Shumar Group	22.7	20
	Phuentsholing Formation	21.0	24
	Daling Formation	23.9	-
	Jaishidanda Formation	4.0	-
	Orthogneiss unit	0.9	-
	Lower metasedimentary unit	30.8	-
Slope (in °)	0–15	19.8	14
	16–30	62.4	31
	31–45	44.8	18
	46–60	4.6	1
	61–75	0.1	
Elevation (in m)	187–587	64.4	26
	588–994	45.5	38
	995–1499	17.7	-
	1500–2437	4.1	-
Rainfall (in mm)	2500–2800	5.8	-
	2801–3100	33.0	-
	3101–3400	92.9	64
Land Use/Land Cover Type	Forest	104.7	37
	Built up	1.4	7
	Agriculture	19.4	14
	Miscellaneous	6.2	6

4. Method

The method revolves around the determination of some important landslide risk factors (i.e., hazard, vulnerability, elements at risk, and potential damage). Landslide risk is defined as the maximum potential loss due to a landslide event for a definite area and during a certain period [18]. Hence, landslide risk can be calculated as:

$$\text{Risk} = \text{Hazard} \times \text{Vulnerability} \times \text{Elements at risk} \quad (1)$$

A complete quantitative risk analysis includes explicitly the economical exposure of elements at risk like buildings, infrastructures, and assets. However, due to the unavailability of such factors, population is used as a proxy to obtain a relative estimation of exposure and the final output is a map of qualitative risk classes.

The first factor is "hazard" which has been often erroneously used synonymously with "susceptibility" [36], but there are major differences between them. Landslide hazard is the probability of a landslide event of a certain magnitude within spatial and temporal limits [11], whereas landslide susceptibility is the likelihood of landslide occurrence depending on local conditions [37].

The present study determines the landslide hazard by considering spatial (P_s) and temporal (P_T) components [33,34]. The spatial aspect was extracted by analyzing all the landslide contributing factors and applying a weighted linear combination (WLC) method. WLC is quite often used by landslide researchers, but an accurate method of determining weights is not defined and is based on expert opinion [38]. It is based on determining weighted averages of landslide causative parameters pertinent to the study area and available data. The parameters are classified and multiplied with the provided weight in a GIS overlay environment, and the final output is determined by adding weighted averages.

The temporal probability was calculated using the rainfall threshold determined by Dikshit et al. [26] using the Poisson probability model. Readers are referred to [3,39,40] for the detailed explanation on determining the temporal landslide probability.

$$\text{Hazard (H)} = P_s \times P_T, \quad (2)$$

Landslide vulnerability is the degree of loss of a particular element at risk due to the occurrence of a certain magnitude of the event [41,42]. Glade and Crozier [43] highlighted that there is no fixed technique for quantitative vulnerability assessment of elements at risk with respect to different landslide magnitudes and types. Therefore, vulnerability is usually determined by either assigning a fixed value to elements at risk (0 for no damage and 1 for complete damage), or assigning values based on expert opinion and knowledge about the area [44]. Thereafter, the potential damage is calculated by combining the vulnerability and elements at risk. In the present study, the potential damage map was developed using the WLC method by providing the exposure of the weights to causative factors expressing both elements of vulnerability and exposure. The factors considered for vulnerability were land use/land cover (LULC), proximity to road, and proximity to stream, and for exposure was population.

The classification of derived thematic layers can be conducted using several techniques like equal intervals, natural breaks, and quantile and statistical techniques which are default GIS processes [45]. In the present study, the natural breaks method was used to classify values as it is quite popular among the researchers [45–47].

5. Results and Discussions

5.1. Landslide Hazard Assesment

As described in the methodology, the spatial component (susceptibility) was based on the WLC method and the temporal probability was determined based on the thresholds obtained using an empirical-based approach. The susceptibility map was based on landslide causative factors and the weights were provided as per Equation (3).

$$\text{LS} = \text{Rainfall} \times 0.3 + \text{Geology} \times 0.3 + \text{Elevation} \times 0.2 + \text{Slope} \times 0.1 + \text{Vegetation} \times 0.1 \quad (3)$$

where LS is landslide susceptibility. The weights are assigned using GIS overlay capability, based on its influence on landslide in comparison to other factors and performing a weighted sum to combine the factors [38]. The fault data were not added as the majority of the landslides are rainfall-triggered and can be attributed as a low-seismic region [48]. The temporal probability is calculated by multiplying annual exceedance probability and the probability of landslide occurrence. The annual exceedance probability is the probability of threshold being exceeded in a specific year, which is determined using the Poisson probability model. The maps derived using both the components resulted in the hazard map. Figure 5 represents the landslide hazard map, which has been divided into five classes: very low (0–0.33), low (0.34–0.44), moderate (0.45–0.58), high (0.59–0.72), and very high (0.73–1).

The results show that the central-eastern parts of the area are more highly hazardous than the 4 km radius of central Phuentsholing, which comes under high hazard zone. The eastern and western parts of the region fall under very low to medium hazard zone. This could prove to be very helpful in landslide prevention mapping and could be used for land use planning and in inhibitory works by governmental agencies.

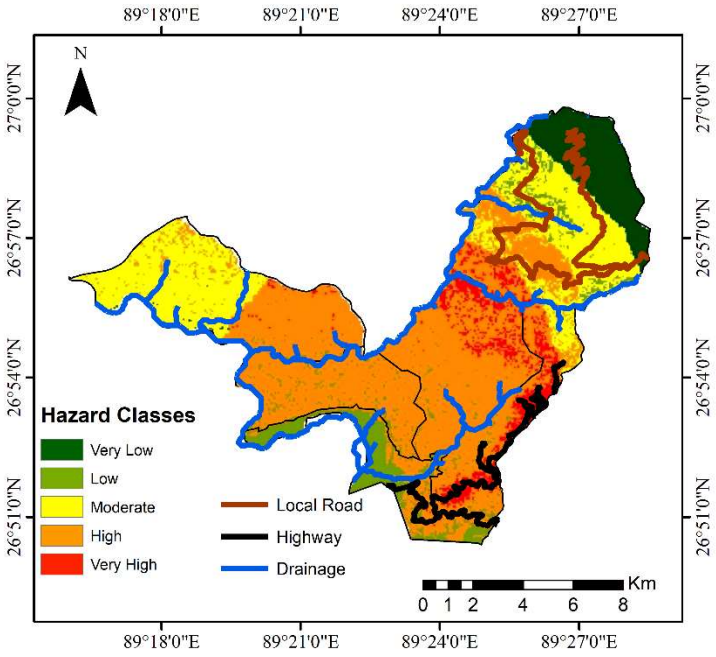


Figure 5. Landslide hazard map of the study region.

5.2. Landslide Vulnerability Assessment

The vulnerability assessment is defined as the expected degree of loss for elements at risk due to a certain event, with values ranging between 0 (no damage) and 1 (complete damage). Its assessment is based on estimation of vulnerability of life and property. A detailed risk assessment would involve considering several elements at risk, like population density, residential and commercial buildings, and their relationship with the type of landslides. However, due to the unavailability of such detailed data, the component of elements at risk could not be considered. The potential damage map was developed using four layers of elements related to exposure and vulnerability, which were LULC, population, proximity to road, and proximity to stream Figure 6a–d and a similar approach to susceptibility mapping was used. Apart from expert opinion, the weights (Equation (4)) provided indicate a logical progression. For example, the built-up area would have more loss due to a higher population, thereby providing them with a higher multiplier. Such regions are specifically used for either commercial or residential purposes or both. These areas can be dangerous on the hills as most of the houses are built on slopes and the roads are constructed by hill cutting. Agriculture covers a significant portion of the study area as it is a means of livelihood for the people staying there. Miscellaneous region includes water bodies and mining sites, and any effect on such sites due to landslides can lead to economic and ecological impact. Similarly, the road network is a lifeline to the people and to the economy of the country, therefore a higher multiplier was provided. This makes the proximity to road factor critical as road cuts are quite common in this region and the risk decreases as distance increases away from the road network. The relevance to the proximity to stream factor is due to the presence of stream channels which could suffer from erosion and hill cutting especially during heavy monsoons. The cutting of hills increases slope instability, making the area more vulnerable to landslides, and it usually increases in regions of higher drainage density and higher elevation. The total population in the study area is 33,444 with 18,043 males and 15,401 females as per census in Bhutan, 2017 [49].

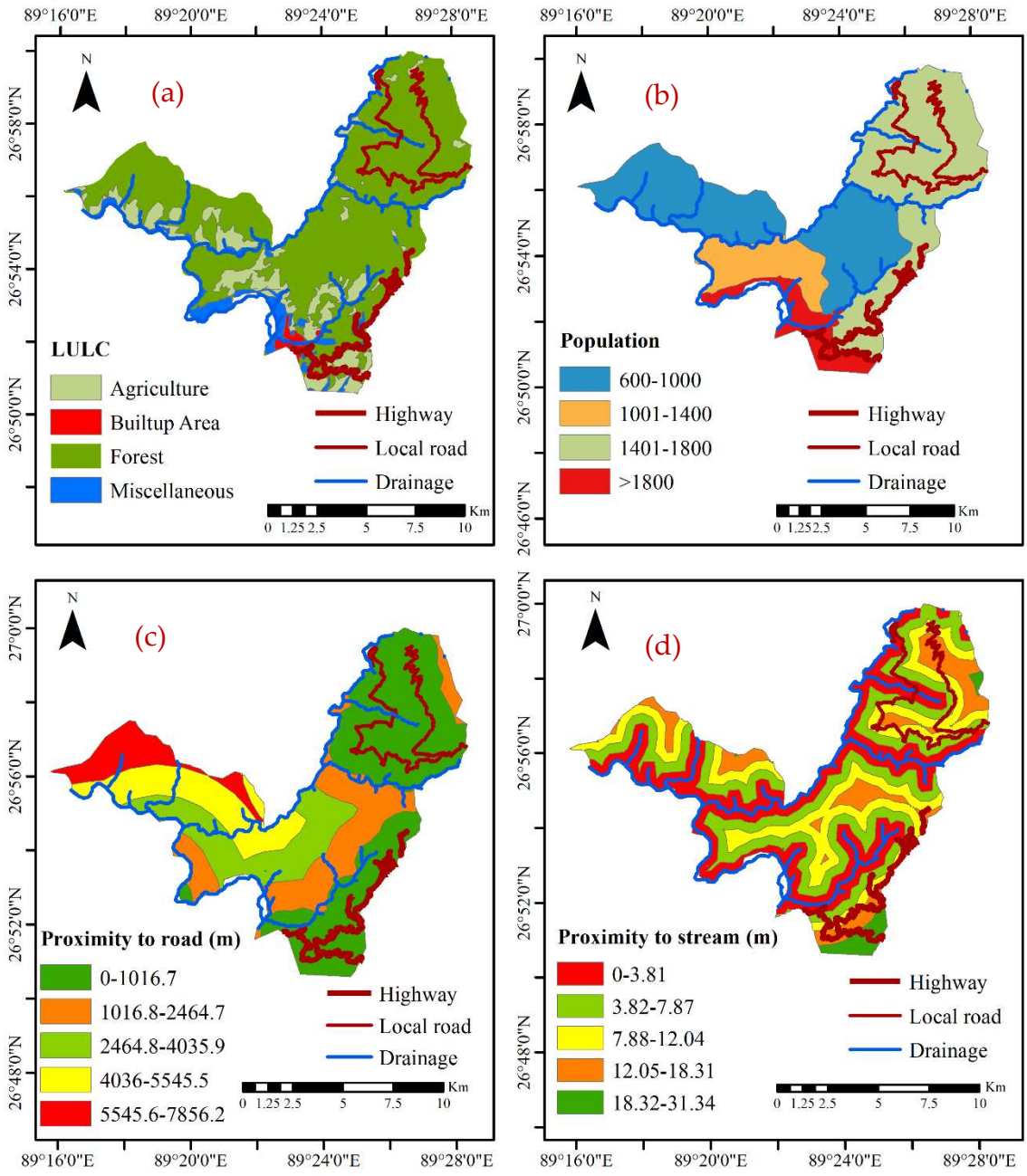


Figure 6. Landslide vulnerability factor maps (**a**) LULC, (**b**) population, (**c**) proximity to road, (**d**) proximity to stream.

The weights to the layers were provided based on Equation (4).

$$LV = \text{Population} \times 0.3 + \text{LULC} \times 0.2 + \text{Proximity to road} \times 0.3 + \text{Proximity to stream} \times 0.2 \quad (4)$$

where LV is landslide vulnerability map. The vulnerability map was developed and classified into five classes based on the natural breaks classification technique. Figure 7 represents the landslide vulnerability map and has been divided into five categories, namely, very low (0–0.36), low (0.37–0.48), moderate (0.49–0.64), high (0.65–0.76), and very high (0.77–1). The map depicts that very high and high vulnerable regions can be found in the NE and SW part of the location covering the populated areas. The developed vulnerability map can serve as a guideline for local disaster management agencies for development activities.

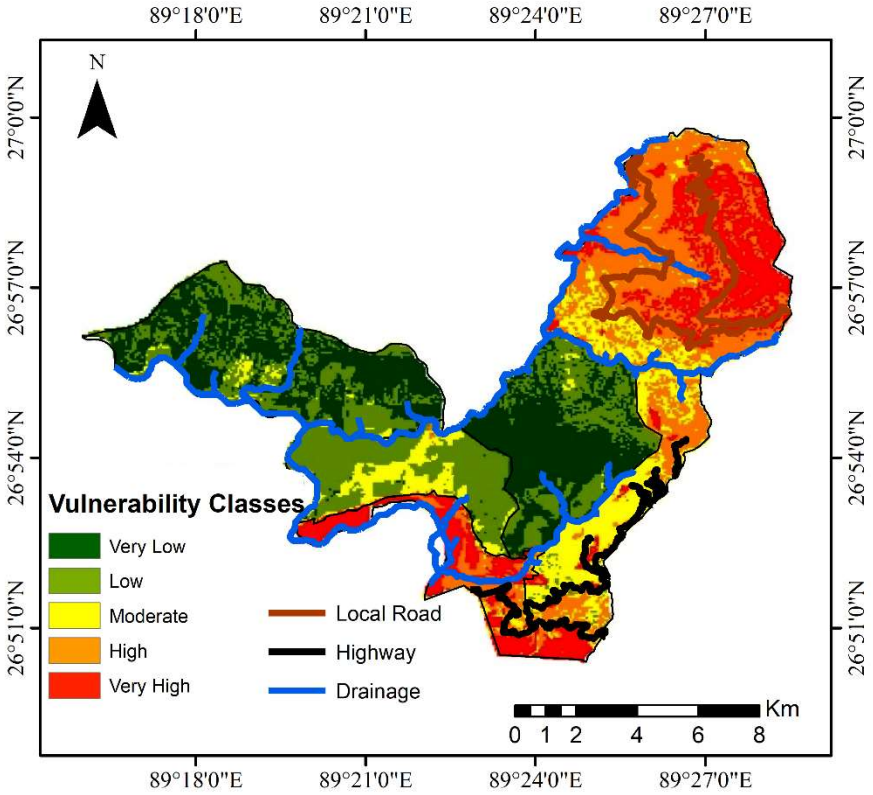


Figure 7. Landslide vulnerability map for Phuentsholing.

5.3. Landslide Risk Assessment

The type of damage linked to landslide risk in the present study has been confined to loss of life and/or property along with a road network. The risk map was developed by multiplying the landslide hazard and vulnerability map. Such a study would also help in estimating the vulnerable area and the impact from any landslide event in the foreseeable future. In order to understand the vulnerability to human lives, the risk to population was estimated and the high-risk zones were identified (Figure 8). The obtained map was classified into five classes (very low (0–0.12), low (0.13–0.21), moderate (0.22–0.37), high (0.38–0.55), and very high (0.56–1)) based on the natural break classification technique [47]. Table 2 depicts the area and affected population under various risk zones whereas Table 3 determines the various landslide risk zones for all six subdivisions (chiwogs). It can be inferred that 20.2% of the population with an area of 15.13 km^2 lives in high to high-risk landslide zones.

Table 2. Area and population distribution with respect to risk zones.

Risk Zone	Area (%)	Area (km^2)	Population
Very Low	44.70	58.7	19,148
Low	27.75	36.7	2030
Moderate	15.85	21.2	5496
High	8.50	11.3	6341
Very High	3.20	3.8	429

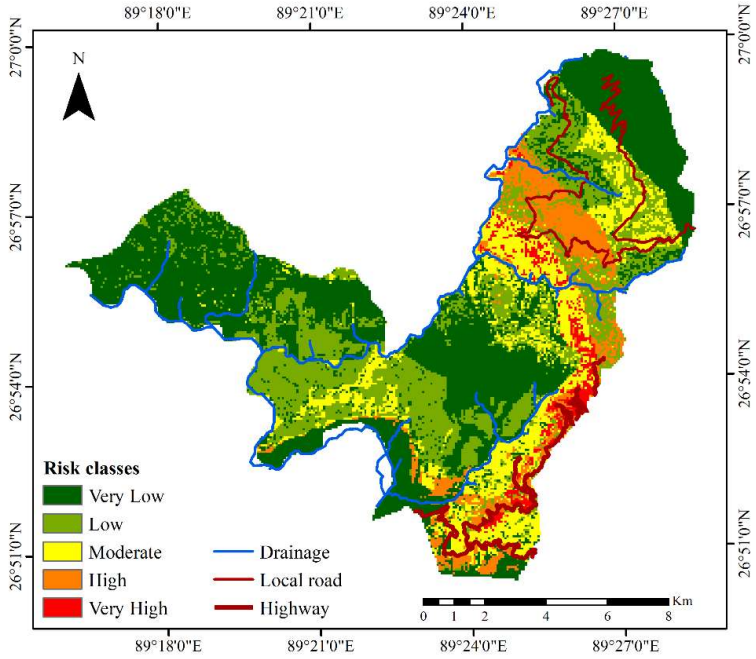


Figure 8. Landslide risk map of the study region.

Table 3. Area affected by various risk zones for all six chiwogs.

Chiwogs	Risk Zone	Area (km^2)
Lingdaen	Very Low	21.71
	Low	5.60
	Moderate	0.20
	High	0.00
	Very High	0.00
Pachhu	Very Low	1.30
	Low	11.40
	Moderate	2.80
	High	0.07
	Very High	0.00
Dzongkhag Thromde	Very Low	7.02
	Low	0.09
	Moderate	1.70
	High	2.35
	Very High	0.03
Dophugchen Wangduegatshel	Very Low	0.50
	Low	2.71
	Moderate	6.90
	High	2.82
	Very High	2.71
Deling Marpji	Very Low	14.20
	Low	6.00
	Moderate	2.60
	High	0.14
	Very High	0.01
Chong Geykha Dophulakha	Very Low	13.76
	Low	11.10
	Moderate	6.96
	High	5.95
	Very High	1.07

6. Conclusions

The occurrence of landslides and their growing incidence is a major threat to population and developments to the mountainous region, especially in developing countries. This has led to a need for comprehensive study on various aspects of landslides in such regions. However, such studies in the Bhutan Himalayas is very scarce and studies need to be conducted to understand the landslide effects in the region. Therefore, a landslide risk map has been developed for the Phuentsholing region of southwestern Bhutan. The study area has been selected since the major trade route, Phuentsholing-Thimphu Highway, which connects Bhutan with neighboring countries, passes through it and a landslide in this area can hamper people's lives and lead to economic losses. The method involved assessing the risk by multiplying the hazard and vulnerability maps. The landslide hazard map was developed by determining the spatial and temporal probabilities of landslides. The spatial probability was determined by performing a weighted sum analysis in an overlay environment for landslide causative factors (geology, slope, rainfall, vegetation, elevation), whereas temporal probability was determined using Poisson's probability model based on the already calculated regional thresholds. Similarly, landslide vulnerability was determined by providing weights based on expert knowledge gathered after carrying out extensive field study to various elements at risk involved. Finally, the landslide risk map was developed by combining the landslide hazard and vulnerability maps and was divided into five categories depending on various levels of risk associated. The analysis showed that 11.7% of the area comes under the high to very high category of risk, whereas 15.85% falls under moderate risk. One of the major problems for determining landslide risk in such a region is the unavailability of longer and precise rainfall and landslide records, with often under-reporting of the landslide damages. However, the present study can be considered as a preliminary step towards achieving a deep understanding of landslide risk at a local scale. The maps can be further modified and improved with the availability of more detailed data like the distribution of population in terms of gender, and the landslide type. The areas along the highway are expected to be at considerable risk in the near future as more hills are cut for residential and economic development purposes. Such an analysis may be used as a tool to develop a building code based on landslide risk, which Bhutan is presently lacking. This analysis would help the concerned authorities to understand the potential risk in certain sections and could be used for landslide mitigation purposes as there is a rise of infrastructure development in the region.

Author Contributions: Conceptualization, A.D. and R.S.; methodology, A.D. and S.A.; data curation, S.A.; writing—original draft preparation, A.D. and R.S.; writing—review and editing, B.P. and A.M.A.; supervision, B.P. and R.S. All authors have read and agreed to the published version of the manuscript.

Funding: The research was funded from BRACE project (NERC/GCRF NE/P016219/1) granted to Raju Sarkar. This research was supported by the Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS) in the University of Technology Sydney (UTS) under Grants 321740.2232335 and 321740.2232357; Grant 321740.2232424, and Grant 321740.2232452. This research is also supported by Researchers Supporting Project number RSP-2019/14, King Saud University, Riyadh, Saudi Arabia.

Acknowledgments: The authors are thankful to the National Center for Hydrology and Meteorology, Royal Government of Bhutan for providing rainfall data and the Border Roads Organization (Project DANTAK), Government of India for providing landslide data. The authors acknowledge the three anonymous reviewers and the editors for their useful comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Petley, D. Global patterns of loss of life from landslides. *Geology* **2012**, *40*, 927–930. [CrossRef]
2. Froude, M.J.; Petley, D.N. Global fatal landslide occurrence from 2004 to 2016. *Nat. Hazards Earth Syst. Sci.* **2018**, *18*, 2161–2181. [CrossRef]
3. Dikshit, A.; Sarkar, R.; Pradhan, B.; Jena, R.; Drukpa, D.; Alamri, A. Temporal Probability Assessment and Its Use in Landslide Susceptibility Mapping for Eastern Bhutan. *Water* **2020**, *12*, 267. [CrossRef]

4. Dikshit, A.; Satyam, N. Probabilistic rainfall thresholds in Chibo, India: Estimation and validation using monitoring system. *J. Mt. Sci.* **2019**, *16*, 870–883. [CrossRef]
5. Segoni, S.; Piciullo, L.; Gariano, S.L. A review of the recent literature on rainfall thresholds for landslide occurrence. *Landslides* **2018**, *15*, 1483–1501. [CrossRef]
6. Dikshit, A.; Sarkar, R.; Satyam, N. Probabilistic approach toward Darjeeling Himalayas landslides-A case study. *Cogent Eng.* **2018**, *5*, 1–11. [CrossRef]
7. Yin, Y.; Zheng, W.; Liu, Y.; Zhang, J.; Li, X. Integration of GPS with InSAR to monitoring of the Jiaju landslide in Sichuan, China. *Landslides* **2010**, *7*, 359–365. [CrossRef]
8. Bui, D.T.; Shahabi, H.; Omidvar, E.; Shirzadi, A.; Geertsema, M.; Clague, J.J.; Khosravi, K.; Pradhan, B.; Pham, B.T.; Chapi, K.; et al. Shallow Landslide Prediction Using a Novel Hybrid Functional Machine Learning Algorithm. *Remote. Sens.* **2019**, *11*, 931.
9. Vranken, L.; Vantilt, G.; Eeckhaut, M.V.D.; Vandekerckhove, L.; Poesen, J. Landslide risk assessment in a densely populated hilly area. *Landslides* **2014**, *12*, 787–798. [CrossRef]
10. Einstein, H.H. Landslide risk–Systematic approaches to assessment and management. *Landslide Risk Assess.* **2018**, 25–50. [CrossRef]
11. Althuwaynee, O.F.; Pradhan, B. Semi-quantitative landslide risk assessment using GIS-based exposure analysis in Kuala Lumpur City. *Geomat. Nat. Hazards Risk* **2016**, *8*, 706–732. [CrossRef]
12. Carrara, A.; Publiese-Carratelli, E.; Merenda, L. Computer-based data bank and statistical analysis of slope instability phenomena. *Z. fur Geomorphol. N.F.* **1977**, *21*, 187–222.
13. Nilsen, T.H.; Wright, R.; Vlasic, T.; Spangle, W. Relative slope stability and land-use planning in the San Francisco Bay region, California. *Prof. Pap.* **1979**, *944*. [CrossRef]
14. Lovell, D.P.; Thomas, G. Quantitative risk assessment. In Proceedings of the Food Chemical Risk Analysis; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 1997; pp. 57–86.
15. Wong, H. Landslide risk assessment for individual facilities. In *Landslide Risk Management*; CRC Press: Boca Raton, FL, USA, 2005; pp. 247–306.
16. Hadmoko, D.S.; Lavigne, F.; Sartohadi, J.; Hadi, P. Winaryo Landslide hazard and risk assessment and their application in risk management and landuse planning in eastern flank of Menoreh Mountains, Yogyakarta Province, Indonesia. *Nat. Hazards* **2010**, *54*, 623–642. [CrossRef]
17. Pradhan, B.; Abdulwahid, W.M. Landslide Risk Assessment Using Multi-hazard Scenario Produced by Logistic Regression and LiDAR-Based DEM. In *Laser Scanning Applications in Landslide Assessment*; Springer Science and Business Media LLC: Berlin/Heidelberg, German, 2017; Volime 9, pp. 253–275.
18. Liu, X.; Miao, C. Large-scale assessment of landslide hazard, vulnerability and risk in China. *Geomat. Nat. Hazards Risk* **2018**, *9*, 1037–1052. [CrossRef]
19. Goorchi, R.N.; Amini, M.; Memarian, H. A new rating system approach for risk analysis of rock slopes. *Nat. Hazards* **2017**, *91*, 75–102.
20. Saldivar-Sali, A.; Einstein, H.H. A Landslide Risk Rating System for Baguio, Philippines. *Eng. Geol.* **2007**, *91*, 85–99. [CrossRef]
21. Alejano, L.R.; Stockhausen, H.; Alonso, E.; Bastante, F.G.; Oyanguren, P.R. ROFRAQ: A statistics-based empirical method for assessing accident risk from rockfalls in quarries. *Int. J. Rock Mech. Min. Sci.* **2008**, *45*, 1252–1272. [CrossRef]
22. Lombardi, M.; Cardarilli, M.; Raspa, G. Spatial variability analysis of soil strength to slope stability assessment. *Géoméch. Eng.* **2017**, *12*, 483–503. [CrossRef]
23. Tofani, V.; Bicocchi, G.; Rossi, G.; Segoni, S.; D'Ambrosio, M.; Casagli, N.; Catani, F. Soil characterization for shallow landslides modeling: A case study in the Northern Apennines (Central Italy). *Landslides* **2017**, *14*, 755–770. [CrossRef]
24. Cardarilli, M.; Lombardi, M.; Corazza, A. Landslide risk management through spatial analysis and stochastic prediction for territorial resilience evaluation. *Int. J. Saf. Secur. Eng.* **2019**, *9*, 109–120. [CrossRef]
25. Gariano, S.L.; Sarkar, R.; Dikshit, A.; Dorji, K.; Brunetti, M.T.; Peruccacci, S.; Melillo, M. Automatic calculation of rainfall thresholds for landslide occurrence in Chukha Dzongkhag, Bhutan. *Bull. Int. Assoc. Eng. Geol.* **2018**, *78*, 4325–4332. [CrossRef]

26. Dikshit, A.; Sarkar, R.; Pradhan, B.; Acharya, S.; Dorji, K. Estimating Rainfall Thresholds for Landslide Occurrence in the Bhutan Himalayas. *Water* **2019**, *11*, 1616. [CrossRef]
27. Sarkar, R.; Dorji, K. Determination of the Probabilities of Landslide Events—A Case Study of Bhutan. *Hydrology* **2019**, *6*, 52. [CrossRef]
28. Erener, A.; Duzgun, H.S.B. A regional scale quantitative risk assessment for landslides: Case of Kumluca watershed in Bartin, Turkey. *Landslides* **2012**, *10*, 55–73. [CrossRef]
29. Luca, F.; D'Ambrosio, D.; Robustelli, G.; Rongo, R.; Spataro, W. Integrating geomorphology, statistic and numerical simulations for landslide invasion hazard scenarios mapping: An example in the Sorrento Peninsula (Italy). *Comput. Geosci.* **2014**, *67*, 163–172. [CrossRef]
30. Piciullo, L.; Calvello, M.; Cepeda, J.M. Territorial early warning systems for rainfall-induced landslides. *Earth-Sci. Rev.* **2018**, *179*, 228–247. [CrossRef]
31. Gansser, A. Geology of the Bhutan Himalaya. In *Denkschriften der Schweizerischen Naturforschenden Geselschaft*; Birkhäuser Verlag: Basel, Switzerland, 1983; p. 181.
32. Segoni, S.; Pappafico, G.; Luti, T.; Catani, F. Landslide susceptibility assessment in complex geological settings: Sensitivity to geological information and insights on its parameterization. *Landslides* **2020**, 1–11. [CrossRef]
33. Long, S.; McQuarrie, N.; Tobgay, T.; Grujic, D. Geometry and crustal shortening of the Himalayan fold-thrust belt, eastern and central Bhutan. *GSA Bull.* **2011**, *123*, 1427–1447. [CrossRef]
34. Lohmann, U.; Sausen, R.; Bengtsson, L.; Cubasch, U.; Perlwitz, J.; Roeckner, E. The Köppen climate classification as a diagnostic tool for general circulation models. *Clim. Res.* **1993**, *3*, 177–193. [CrossRef]
35. Prokop, P.; Walanus, A. Impact of the Darjeeling–Bhutan Himalayan front on rainfall hazard pattern. *Nat. Hazards* **2017**, *89*, 387–404. [CrossRef]
36. Hervás, J.; Bobrowsky, P. Mapping: Inventories, Susceptibility, Hazard and Risk. In *Landslides–Disaster Risk Reduction*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2008; pp. 321–349.
37. Reichenbach, P.; Rossi, M.; Malamud, B.; Mihir, M.; Guzzetti, F. A review of statistically-based landslide susceptibility models. *Earth-Sci. Rev.* **2018**, *180*, 60–91. [CrossRef]
38. Michael, E.A.; Samanta, S. Landslide vulnerability mapping (LVM) using weighted linear combination (WLC) model through remote sensing and GIS techniques. *Model. Earth Syst. Environ.* **2016**, *2*, 88. [CrossRef]
39. Jaiswal, P.; Van Westen, C. Estimating temporal probability for landslide initiation along transportation routes based on rainfall thresholds. *Geomorphology* **2009**, *112*, 96–105. [CrossRef]
40. Guzzetti, F.; Reichenbach, P.; Cardinali, M.; Galli, M.; Ardizzone, F. Probabilistic landslide hazard assessment at the basin scale. *Geomorphology* **2005**, *72*, 272–299. [CrossRef]
41. Lin, Q.; Wang, Y.; Liu, T.; Zhu, Y.; Sui, Q. The Vulnerability of People to Landslides: A Case Study on the Relationship between the Casualties and Volume of Landslides in China. *Int. J. Environ. Res. Public Heal.* **2017**, *14*, 212. [CrossRef]
42. Baiocchi, V.; Dominici, D.; Guarascio, M.; Lombardi, M.; Vatore, F. Mapping seismic vulnerability in buildings by means of open source tools and open data. *Geoing. Ambient. e Min.* **2017**, *151*, 27–32.
43. Glade, T.; Crozier, M.J. *The Nature of Landslide Hazard Impact*; Landslide Hazard and Risk; Wiley: Chichester, UK, 2005; pp. 43–74.
44. Van Westen, C.; Alkema, D.; Damen, M.; Kerle, N.; Kingma, N. *Multi-Hazard Risk Assessment: Distance Education Course Guide Book*; United Nations University—ITC School on Disaster Geoinformation Management: Tokyo, Japan, 2009.
45. Jena, R.; Pradhan, B.; Beydoun, G. Earthquake vulnerability assessment in Northern Sumatra province by using a multi-criteria decision-making model. *Int. J. Disaster Risk Reduct.* **2020**, *46*, 101518. [CrossRef]
46. Pourghasemi, H.R.; Mohammady, M.; Pradhan, B. Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran. *Catena* **2012**, *97*, 71–84. [CrossRef]
47. Park, S.-J.; Lee, C.-W.; Lee, S.; Lee, M.-J. Landslide Susceptibility Mapping and Comparison Using Decision Tree Models: A Case Study of Jumunjin Area, Korea. *Remote. Sens.* **2018**, *10*, 1545. [CrossRef]

48. Kayal, J.R.; Arefiev, S.S.; Baruah, S.; Tatevossian, R.; Gogoi, N.; Sanoujam, M.; Gautam, J.L.; Hazarika, D.; Borah, D. The 2009 Bhutan and Assam felt earthquakes (Mw6.3 and 5.1) at the Kopili fault in the northeast Himalaya region. *Geomat. Nat. Hazards Risk* **2010**, *1*, 273–281. [CrossRef]
49. Bhutan, N. *Population and Housing Census of Bhutan*; Royal Government of Bhutan: Thimpu, Bhutan, 2017.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

arXiv:2007.07816v1 [math.CV] 15 Jul 2020

SUBORDINATION AND RADIUS PROBLEMS FOR CERTAIN STARLIKE FUNCTIONS

S. SIVAPRASAD KUMAR AND KAMALJEET GANGANIA

Abstract. We study the following class of starlike functions

$$\mathcal{S}^*_\wp := \{f \in \mathcal{A} : zf'(z)/f(z) \prec 1 + ze^z =: \wp(z)\},$$

that are associated with the cardioid domain $\wp(\mathbb{D})$, by deriving certain convolution results, radius problems, majorization result, radius problems in terms of coefficients and differential subordination implications. Consequently, we establish some interesting generalizations of our results for the Ma-Minda class of starlike functions $\mathcal{S}^*(\psi)$. We also provide, the set of extremal functions maximizing $\Re\Phi\left(\log\left(f(z)/z\right)\right)$ or $|\Phi\left(\log\left(f(z)/z\right)\right)|$ for functions in $\mathcal{S}^*(\psi)$, where Φ is a non-constant entire function. Further T. H. MacGregor's result for the class $\mathcal{S}^*(\alpha)$ and $\mathcal{S}^*_\wp$ are obtained as special case to our result.

2010 AMS Subject Classification. Primary 30C45, Secondary 30C50, 30C80.
Keywords and Phrases. Starlike functions; Convolution; Radius problem; Extremal problem; Subordination; Majorization; Special functions.

1. Introduction

Let $\mathcal{A}_0$ be the collection of analytic functions of the form $p(z) = 1+p_1z+p_2z+\cdots$ and $\mathcal{A}$ consists of analytic functions, f normalized by the conditions $f(0) = 0$ and $f'(0) = 1$ defined in the unit disk $\mathbb{D} := \mathbb{D}_1$, where $\mathbb{D}_r := \{z : |z| < r\}$. The Carathéodory class, $\mathcal{P}$ consists of functions $p \in \mathcal{A}_0$ with $\Re p(z) > 0$. Let us denote the class of normalized univalent functions by $\mathcal{S}$. For two analytic functions f and g, we say f subordinate to g, written as $f \prec g$, if there exists a Schwarz function w such that $f(z) = g(w(z))$. If g is univalent, then $f \prec g$ if and only if $f(\mathbb{D}) \subseteq g(\mathbb{D})$ and $f(0) = g(0)$. Many subclasses of $\mathcal{S}$ were introduced and studied by several authors, the chief amongst them is $\mathcal{S}^*$ (or $\mathcal{C}$), the class of normalized starlike (or convex) functions which is characterized by the quantity $zf'(z)/f(z)$ (or $1 + zf''(z)/f'(z)$), assuming values in the right half plane. In 1992, using subordination, Ma and Minda [19] defined the classes (1.1) and (1.2), which unify many subclasses of $\mathcal{S}$ and they also studied some of it's common properties.

$$\mathcal{S}^*(\psi) := \left\{f \in \mathcal{A} : \frac{zf'(z)}{f(z)} \prec \psi(z)\right\} \tag{1.1}$$

and

$$\mathcal{C}(\psi) := \left\{f \in \mathcal{A} : 1 + \frac{zf''(z)}{f'(z)} \prec \psi(z)\right\}, \tag{1.2}$$

where $\psi \in \mathcal{P}$ is univalent and $\psi(\mathbb{D})$ starlike with respect to $\psi(0) = 1$. Note that $\mathcal{S}^*(\frac{1+z}{1-z})$ reduces to the class $\mathcal{S}^*$. Now a good amount of literature exists for different

The work of the second author is supported by University Grant Commission, New-Delhi, India under UGC-Ref. No.:1051/(CSIR-UGC NET JUNE 2017).

choices of ψ in (1.1). For example, the class $\mathcal{S}(\alpha,\beta) := \mathcal{S}^*(p_{\alpha,\beta}(z))$, where

$$p_{\alpha,\beta}(z) := 1 + \frac{\beta-\alpha}{\pi} i \log \frac{1-e^{2\pi i \frac{1-\alpha}{\beta-\alpha}} z}{1-z}, \tag{1.3}$$

$\alpha < 1$, $\beta > 1$ and $p_{\alpha,\beta}$ maps $\mathbb{D}$ onto the convex domain $\{w \in \mathbb{C} : \alpha < \Re w < \beta\}$, was introduced by Kuroki and Owa [18]. Sokół and Stankiewicz [27] introduced the class associated with lemniscate of Bernoulli $\mathcal{SL}^* := \mathcal{S}^*(\sqrt{1+z})$. The class $\mathcal{S}_e^* := \mathcal{S}^*(e^z)$ of exponential starlike functions was introduced by Mendiratta et al. [22] and the class $\mathcal{S}_b^* := \mathcal{S}^*(e^{e^z-1})$ of starlike functions related with Bell numbers was introduced by Kumar et at. [15]. For some recent classes see [24, 17, 26, 10, 12]. Evidently the functions ψ under discussion above are all convex and hence satisfy the following result:

Theorem 1.1. [19] *Let $\psi(\mathbb{D})$ be convex, $g \in \mathcal{C}$ and $f \in \mathcal{S}^*(\psi)$. Then $f * g \in \mathcal{S}^*(\psi)$.*

But the above result is not applicable when $\psi(z) = 1 + ze^z$, $z + \sqrt{1+z^2}$, e^{e^z-1} and $1 + 4z/3 + 2z^2/3$ etc. due to the fact that these are not convex. We deal with Theorem 1.1 for the case when ψ starlike and derive some radius related problems.

In Geometric function theory, we usually find the sufficient conditions in terms of it's coefficients for the normalized functions f in $\mathcal{A}$ to be in a desired class. In recent times, finding the sufficient conditions on functions to be in a desired class using the technique of differential subordination, admissibility conditions and subordination chain has attracted many researcher's attention. See [4, 12, 23].

In the present work, we consider the following first-order differential subordination implications:

$$1 + \beta \frac{zp'(z)}{p^n(z)} \prec h(z) \Rightarrow p(z) \prec q(z), \tag{1.4}$$

where $n \in \{1,2,3\}$. We find the condition on β so that (1.4) holds whenever the functions h and q are already known. Note that if p and q are two fixed analytic and univalent functions in $\mathbb{D}$ and q does not have the nice properties like, explicit inverse representation and convexity, then proving $p \prec q$ is not so easy. See that if $p \prec q$, then geometrically the boundary curve of $p(\mathbb{D})$ is contained in $q(\mathbb{D})$, which can be ensured by the technique of radial distances. Further, if the images $p(\mathbb{D})$ and $q(\mathbb{D})$ are symmetric about the real axis and both the functions p and q have the same orientation, then we can use the maximum arguments of $p(z)$ and $q(z)$ for $|z| = 1$ to prove that $p \prec q$.

Recall that for the subfamilies $\mathcal{G}_1$ and $\mathcal{G}_2$ of $\mathcal{A}$, we say that r_0 is the $\mathcal{G}_1$-radius of the class $\mathcal{G}_2$, if $r_0 \in (0,1)$ is largest number such that $r^{-1}f(rz) \in \mathcal{G}_1$, $0 < r \leq r_0$ for all $f \in \mathcal{G}_2$. Recently, the radii of starlikeness and convexity of some normalized special functions were studied as they can be represented as Hadamard factorization under certain conditions, a few such special functions are Bessel functions [1], Struve functions [5, 1], Wright functions [6], Lommel functions [5, 1] and Legendre polynomials of odd degree [8]. If f and g be analytic in $\mathbb{D}$, then $g(z)$ is said to be majorized by $f(z)$, if there exists an analytic function $\Phi(z)$ in $\mathbb{D}$ satisfying $|\Phi(z)| \leq 1$ and $g(z) = \Phi(z)f(z)$ for all $z \in \mathbb{D}$. MacGregor [20] found the largest radius r_0 so that $|g'(z)| \leq |f'(z)|$ holds in $|z| \leq r_0$ and he also proved that $r_0 = 2 - \sqrt{3}$, if f is univalent. Recently, Tang and Deng [28] obtained this majorization-radius when f belongs to some class $\mathcal{S}^*(\psi)$.

Consider the class of cardioid starlike functions:

$$\mathcal{S}^*_{\wp} := \left\{ f \in \mathcal{A} : \frac{zf'(z)}{f(z)} \prec \wp(z) := 1 + ze^z \right\}. \tag{1.5}$$

Let $\mathbb{E}_{\alpha,\beta}$ be the normalized form of the Mittag-Leffler function:

$$\mathbb{E}_{\alpha,\beta}(z) = z + \sum_{n\geq 2} \frac{\Gamma(\beta)}{\Gamma(\alpha(n-1)+\beta)} z^n, \quad (z, \alpha, \beta \in \mathbb{C}; \Re(\alpha) > 0, \beta \neq 0, -1, \cdots).$$

Then clearly $\wp(z) = 1 + \mathbb{E}_{1,1}(z)$. In [16], $\mathcal{S}^*_{\wp}$ was extensively studied with inclusion properties, radius problems, sharp coefficient's estimations including hankel determinants. Where in, it is proved that $r_c := (3-\sqrt{5})/2$ is the radius of convexity for $\wp$ however in [3], $\mathbb{E}_{\alpha,\beta}$ is shown to be convex for $\alpha \geq 1$ and $\beta \geq (3+\sqrt{17})/2$. So, $\mathbb{E}_{1,1}$ has the radius of convexity r_c.

In this paper, we continue to study further in finding the largest radius for some normalized special functions to be in $\mathcal{S}^*_{\wp}$. We also obtain largest radius r such that $f \in \mathcal{S}^*_{\wp}$ in $|z| < r$, whenever f belongs to the class $\mathcal{S}(\alpha, \beta)$ and the class $\mathcal{S}_\lambda$ defined in (1.3) and (2.5), respectively. We find that special function's radii of starlikness results obtained here naturally generalize to $\mathcal{S}^*(\psi)$-radii. Further, radius problem related to the majorization and some sufficient conditions in terms of coefficients for the normalized functions to be in the class $\mathcal{S}^*_{\wp}$ are established. We discuss the modified form of Theorem 1.1 and as an application, we find the largest radius so that Theorem 1.1 holds for $\mathcal{S}^*_{\wp}$ and the convolution of two starlike functions to be in $\mathcal{S}^*_{\wp}$. In the last section, we obtain the condition on β so that the differential subordination implication (1.4) holds, whenever $h(z) = \wp(z)$ or $q(z) = \wp(z)$.

2. Radius problems

We need the following lemma to prove the subsequent results:

Lemma 2.1. [16] *Let $\wp(z) = 1 + ze^z$. Then we have $\{w : |w-a| < R_a\} \subset \wp(\mathbb{D})$, where*

$$R_a = \begin{cases} (a-1) + 1/e, & 1 - 1/e < a \leq 1 + (e - e^{-1})/2; \\ e - (a-1), & 1 + (e-e^{-1})/2 \leq a < 1 + e. \end{cases}$$

Theorem 2.1. *Let $f \in \mathcal{S}(\alpha, \beta)$. Then $f \in \mathcal{S}^*_{\wp}$ in $\mathbb{D}_{r_0}$, where r_0 is the least positive root of the equation*

$$\frac{\beta - \alpha}{\pi} \left(\log \frac{1 + \sqrt{2(1 + \cos(2\pi \frac{1-\alpha}{\beta-\alpha}))} r + r^2}{1 - r^2} + 2 \arctan \frac{r}{1-r} \right) - \frac{1}{e} = 0. \tag{2.1}$$

Proof. Consider the analytic function

$$p_{\alpha,\beta}(z) := 1 + \frac{\beta - \alpha}{\pi} i \log q(z),$$

where

$$q(z) = \frac{1 - cz}{1 - z} \quad \text{and} \quad c = \exp\left(2\pi i \frac{1-\alpha}{\beta - \alpha} \right).$$

Note that $q(z)$ is a bilinear transformation, maps $\mathbb{D}_r$ onto the disk:

$$\left| q(z) - \frac{1 + cr^2}{1 - r^2} \right| \leq \frac{|1 + c| r}{1 - r^2},$$

which implies

$$|q(z)| \leq \frac{1+|1+c|r+r^2}{1-r^2}$$

and therefore

$$\log|q(z)| \leq \log\left(\frac{1+|1+c|r+r^2}{1-r^2}\right). \tag{2.2}$$

For any $\delta \in \mathbb{C}$ with $|\delta| = 1$, we have $1+\delta z \prec 1+z$. So to maximize $|\arg(1+\delta z)|$, it suffices to consider $|\arg(1+z)|$. Now for $|z| = r$, we have

$$|\arg(1+z)| \leq \arctan\frac{r}{1-r}. \tag{2.3}$$

Hence to apply Lemma 2.1, we need to maximize $|p_{\alpha,\beta}(z)-1|$, that is,

$$|p_{\alpha,\beta}-1| = \frac{\beta-\alpha}{\pi}\left|\log|q(z)| + i\arg\frac{1-cz}{1-z}\right|. \tag{2.4}$$

Using (2.2) and (2.3) in (2.4), we see that

$$|p_{\alpha,\beta}-1| \leq \frac{\beta-\alpha}{\pi}\left(\log\frac{1+|1+c|r+r^2}{1-r^2} + 2\arctan\frac{r}{1-r}\right) \leq \frac{1}{e}$$

holds in $|z| < r_0$ whenever r_0 is the smallest positive root of the equation (2.1). This ends the proof. □

Note that if we choose $\alpha = 1+\frac{\delta-\pi}{2\sin\delta}$ and $\beta = 1+\frac{\delta}{2\sin\delta}$, where $\pi/2 \leq \delta < \pi$, then the class $\mathcal{S}(\alpha,\beta)$ reduces to the class $\mathcal{V}(\delta)$ introduced by Kargar et al. [14].

Corollary 2.2. *Let $f \in \mathcal{V}(\delta)$. Then $f \in \mathcal{S}^*_\wp$ in $\mathbb{D}_{r_\delta}$, where r_δ is the least positive root of the equation*

$$\frac{1}{2\sin\delta}\left(\log\frac{1+\sqrt{2(1+\cos(2(\pi-\delta)))}r+r^2}{1-r^2} + 2\arctan\frac{r}{1-r}\right) - \frac{1}{e} = 0.$$

Now we consider the following class introduced in [9]:

$$\mathcal{S}_\lambda := \left\{f \in \mathcal{A} : \frac{f(z)}{z} \in P_\lambda\right\}, \tag{2.5}$$

where

$$P_\lambda := \{p \in \mathcal{A}_0 : \Re(e^{i\lambda}p(z)) > 0, \quad -\pi/2 \leq \lambda \leq \pi/2\}$$

denotes the class of tilted Carathéodory functions [29]. Note that P_0 reduces to $\mathcal{P}$, the class of Carathéodory functions. For the function $p \in P_\lambda$, upper bound on the quantity $zp'(z)/p(z)$ is given by the following lemma that will be used for our next result:

Lemma 2.2. [29] *If $p \in P_\lambda$, then $|zp'(z)/p(z)| \leq M(\lambda,r)$, where*

$$M(\lambda,r) = \begin{cases} \frac{2r\cos\lambda}{r^2-2r|\sin\lambda|+1} & \text{for} \quad r < |\tan\frac{\lambda}{2}|; \\ \frac{2r}{1-r^2} & \text{for} \quad r \geq |\tan\frac{\lambda}{2}|. \end{cases}$$

The equality holds for some point $z = re^{i\theta}$, $r \in (0,1)$ if and only if $p(z) = p_\lambda(yz)$, where $p_\lambda(z) = \frac{1+e^{-2i\lambda}z}{1-z}$ and $y = e^{i(\theta_0-\theta)}$ with

$$\theta_0 = \begin{cases} \frac{\pi}{2}+\lambda & \text{for} \quad r < -\tan\frac{\lambda}{2}; \\ -\frac{\pi}{2}+\lambda & \text{for} \quad r < \tan\frac{\lambda}{2}; \\ \arcsin\left(\frac{1+r^2}{r^2-1}\right)+\lambda & \text{for} \quad r \geq |\tan\frac{\lambda}{2}|. \end{cases}$$

We determine the largest radius r such that the function $F(z) := f(z)g(z)/z \in \mathcal{S}^*_\wp$ in $|z| < r$, whenever $f, g \in \mathcal{S}_\lambda$.

Theorem 2.3. *Let $c_\lambda = \cos\lambda$, $s_\lambda = \sin\lambda$ and $t_\lambda = |\tan(\lambda/2)|$. If $f, g \in \mathcal{S}_\lambda$, then $F \in \mathcal{S}^*_\wp$ in $\mathbb{D}_{r_0}$, where*

$$r_0 := \begin{cases} 2ec_\lambda + |s_\lambda| + \sqrt{((4e^2-1)c_\lambda + 4e|s_\lambda|)c_\lambda}, & \text{if } \ r < t_\lambda; \\ \sqrt{4e^2+1} - 2e, & \text{if } \ r \geq t_\lambda. \end{cases}$$

Proof. Since $f, g \in \mathcal{S}_\lambda$, it follows that the functions $p(z) = f(z)/z$ and $q(z) = g(z)/z$ belong to the class P_λ such that $F(z) = zp(z)q(z)$. Thus

$$\frac{zF'(z)}{F(z)} - 1 = \frac{zp'(z)}{p(z)} + \frac{zq'(z)}{q(z)}.$$

Now from Lemma 2.2, we obtain

$$\left|\frac{zF'(z)}{F(z)} - 1\right| \leq 2M(\lambda, r).$$

Therefore, using Lemma 2.1, we conclude that if $2M(\lambda, r) \leq 1/e$, then $F \in \mathcal{S}^*_\wp$. Since $2M(\lambda, r) \leq 1/e$ holds whenever

$$\frac{2rc_\lambda}{r^2 - 2|s_\lambda|r + 1} \leq \frac{1}{2e}, \quad \text{if} \quad r < t_\lambda$$

and

$$\frac{2r}{1-r^2} \leq \frac{1}{2e}, \quad \text{if} \quad r \geq t_\lambda;$$

or equivalently

$$r^2 - 2(|s_\lambda| + 2ec_\lambda)r + 1 \geq 0, \quad \text{if} \quad r < t_\lambda$$

and

$$r^2 + 4er - 1 \leq 0, \quad \text{if} \quad r \geq t_\lambda,$$

respectively. Hence the result follows with r_0 as given in the hypothesis. Further, for the functions

$$f(z) = g(z) = \frac{z(1 + e^{-2i\lambda}yz)}{1 - yz},$$

sharpness hold in view of Lemma 2.2. □

2.1. **Cardioid starlikeness radius for some special functions and generalizations.** We begin with the following: The Bessel function $\mathcal{J}_\beta$ of first kind of order $\beta \in \mathbb{C}$ is a particular solution of the homogeneous Bessel differential equation

$$z^2w''(z) + zw'(z) + (z^2 - \beta^2)w(z) = 0$$

and have the following series expansion:

$$\mathcal{J}_\beta(z) := \sum_{n\geq 1} \frac{(-1)^n}{n!\Gamma(n+\beta+1)} \left(\frac{z}{2}\right)^{2n+\beta},$$

where $z \in \mathbb{C}$ and $\beta \notin \mathbb{Z}^-$. The following three normalized functions expressed in terms of $\mathcal{J}_\beta(z)$ have been studied extensively by many authors:

$$f_\beta(z) = (2^\beta\Gamma(\beta+1)\mathcal{J}_\beta(z))^{1/\beta} = z - \frac{1}{4\beta(\beta+1)}z^3 + \cdots, \quad \beta \neq 0,$$

$$g_\beta(z) = 2^\beta\Gamma(\beta+1)z^{1-\beta}\mathcal{J}_\beta(z) = z - \frac{1}{4(\beta+1)}z^3 + \cdots,$$

$$h_\beta(z) = 2^\beta \Gamma(\beta+1) z^{1-\beta/2} \mathcal{J}_\beta(\sqrt{z}) = z - \frac{1}{4(\beta+1)} z^2 + \cdots . \tag{2.6}$$

Since the zeros of $\mathcal{J}_\beta$ are real if $\beta > 0$, therefore using Weierstrass decomposition, we have for $\beta > 0$:

$$\mathcal{J}_\beta(z) := \frac{z^\beta}{2^\beta \Gamma(\beta+1)} \prod_{n\geq 1} \left(1 - \frac{z^2}{j_{\beta,n}^2}\right),$$

where $j_{\beta,n}$ is the n-th positive zero of $\mathcal{J}_\beta$ and satisfies $j_{\beta,n} < j_{\beta,n+1}$ for $n \in \mathbb{N}$. Thus we have

$$\frac{z\mathcal{J}_\beta'(z)}{\mathcal{J}_\beta(z)} = \beta - \sum_{n\geq 1} \frac{2z^2}{j_{\beta,n}^2 - z^2}. \tag{2.7}$$

Now using the above representation, we obtain the radii of cardioid-starlikeness of the functions f_β, g_β and h_β for the case $\beta > 0$.

Theorem 2.4. *Let $\beta > 0$. Then the radii of cardioid-starlikeness $r_\wp(f_\beta)$, $r_\wp(g_\beta)$ and $r_\wp(h_\beta)$ of the functions f_β, g_β and h_β are the smallest positive roots of the following equations, respectively:*

(i) $er\mathcal{J}_\beta'(r) + \beta(e-1)\mathcal{J}_\beta(r) = 0$;
(ii) $er\mathcal{J}_\beta'(r) - (e\beta - 1)\mathcal{J}_\beta(r) = 0$;
(iii) $e\sqrt{r}\mathcal{J}_\beta'(\sqrt{r}) - (e\beta - 2)\mathcal{J}_\beta(\sqrt{r}) = 0$.

Proof. We now prove the first part. Using the representation (2.7) and equation (2.6), we get

$$\frac{zf_\beta'(z)}{f_\beta(z)} = \frac{z\mathcal{J}_\beta'(z)}{\beta\mathcal{J}_\beta(z)} = 1 - \frac{1}{\beta}\sum_{n\geq 1} \frac{2z^2}{j_{\beta,n}^2 - z^2}. \tag{2.8}$$

Further using a result [11, Lemma 3.2, p. 10] and from (2.8), $|z| = r < j_{\beta,1}$ we obtain

$$\left|\frac{zf_\beta'(z)}{f_\beta(z)} - a\right| \leq \frac{2}{\beta}\sum_{n\geq 1} \frac{j_{\beta,n}^2 r^2}{j_{\beta,n}^4 - r^4}, \tag{2.9}$$

where $a := 1 - \frac{2}{\beta}\sum_{n\geq 1} \frac{r^4}{j_{\beta,n}^4 - r^4}$ and $j_{\beta,n}$ denotes the n-th positive zero of the Bessel function $\mathcal{J}_\beta$. Also a simple calculation shows that $a \leq 1$. Thus for the disk (2.9) to lie inside $\wp(\mathbb{D})$, we need only to consider that $1 - \frac{1}{e} < a < 1 + \frac{e-e^{-1}}{2}$, and so by Lemma 2.1, we have

$$\frac{2}{\beta}\sum_{n\geq 1} \frac{j_{\beta,n}^2 r^2}{j_{\beta,n}^4 - r^4} \leq a - 1 + \frac{1}{e} = \frac{1}{e} - \frac{2}{\beta}\sum_{n\geq 1} \frac{r^4}{j_{\beta,n}^4 - r^4},$$

or equivalently,

$$\frac{2}{\beta}\sum_{n\geq 1} \frac{r^2}{j_{\beta,n}^2 - r^2} - \frac{1}{e} \leq 0. \tag{2.10}$$

Also using (2.8), (2.10) can be written as $\frac{er\mathcal{J}_\beta'(r)}{\beta\mathcal{J}_\beta(r)} + 1 - e \geq 0$. Note that in view of Lemma 2.1, we can also obtain (2.10) directly from (2.8). Now let us consider the strictly decreasing continuous function

$$\Psi(r) := \frac{1}{e} - \frac{2}{\beta}\sum_{n\geq 1} \frac{r^2}{j_{\beta,n}^2 - r^2}, \quad r \in (0, j_{\beta,1}).$$

Then $\lim_{r\to 0}\Psi(r) = 1/e > 0$ and $\lim_{r\to j_{\beta,1}}\Psi(r) = -\infty$. Also $\Psi'(r) < 0$, since $r < j_{\beta,1}$. So we may assume $r_{\wp}(f_\beta)$ be the unique root of $\Psi(r) = 0$ in $(0, j_{\beta,1})$ such that f_β in $\mathcal{S}^*_{\wp}$ in $|z| < r_{\wp}(f_\beta)$. Proof of other parts follows similarly. □

The Struve function $\mathbf{H}_\beta$ of first kind is a particular solution of the second-order inhomogeneous Bessel differential equation

$$z^2w''(z) + zw'(z) + (z^2 - \beta^2)w(z) = \frac{4(\frac{z}{2})^{\beta+1}}{\sqrt{\pi}\Gamma(\beta + \frac{1}{2})}$$

and have the following form:

$$\mathbf{H}_\beta(z) := \frac{(\frac{z}{2})^{\beta+1}}{\sqrt{\frac{\pi}{4}}\Gamma(\beta + \frac{1}{2})}{}_1F_2\left(1; \frac{3}{2}, \beta + \frac{3}{2}; -\frac{z^2}{4}\right),$$

where $-\beta - \frac{3}{2} \notin \mathbb{N}$ and ${}_1F_2$ is a hypergeometric function. Since it is not normalized, so we consider the following three normalized functions involving $\mathbf{H}_\beta$:

$$\begin{aligned} U_\beta(z) &= \left(\sqrt{\pi}2^\beta(\beta + \frac{3}{2})\mathbf{H}_\beta(z)\right)^{\frac{1}{\beta+1}}, \\ V_\beta(z) &= \sqrt{\pi}2^\beta z^{-\beta}\Gamma(\beta + \frac{3}{2})\mathbf{H}_\beta(z), \\ W_\beta(z) &= \sqrt{\pi}2^\beta z^{\frac{1-\beta}{2}}\Gamma(\beta + \frac{3}{2})\mathbf{H}_\beta(\sqrt{z}). \end{aligned} \tag{2.11}$$

Moreover, for $|\beta| \le \frac{1}{2}$, it has the Hadamard factorization given by

$$\mathbf{H}_\beta(z) = \frac{z^{\beta+1}}{\sqrt{\pi}2^\beta\Gamma(\beta + \frac{3}{2})}\prod_{n\ge 1}\left(1 - \frac{z^2}{z^2_{\beta,n}}\right), \tag{2.12}$$

where $z_{\beta,n}$ is the n-th positive root of $\mathbf{H}_\beta$ such that $z_{\beta,n+1} > z_{\beta,n}$ and $z_{\beta,1} > 1$ and also from (2.12), we obtain

$$\frac{z\mathbf{H}'_\beta(z)}{\mathbf{H}_\beta(z)} = (\beta + 1) - \sum_{n\ge 1}\frac{2z^2}{z^2_{\beta,n} - z^2}. \tag{2.13}$$

Now using the above representation, we can obtain the radii of cardioid-starlikeness of the functions U_β, V_β and W_β for the case $|\beta| \le 1/2$.

Theorem 2.5. *Let $|\beta| \le 1/2$. Then the radii of cardioid-starlikeness $r_{\wp}(U_\beta)$, $r_{\wp}(V_\beta)$ and $r_{\wp}(W_\beta)$ of the functions U_β, V_β and W_β are the smallest positive roots of the following equations, respectively:*

(i) $r\mathbf{H}'_\beta(r) - (1 - \frac{1}{e})(\beta + 1)\mathbf{H}_\beta(r) = 0$;
(ii) $r\mathbf{H}'_\beta(r) - ((1 + \beta) - \frac{1}{e})\mathbf{H}'_\beta(r) = 0$;
(iii) $\sqrt{r}\mathbf{H}'_\beta(\sqrt{r}) - (1 + \beta - \frac{2}{e})\mathbf{H}_\beta(\sqrt{r}) = 0$.

Proof. From (2.11) and (2.13), by logarithmic differentiation, we get

$$\begin{aligned} \frac{zU'_\beta(z)}{U_\beta(z)} &= \frac{1}{\beta + 1}\frac{z\mathbf{H}'_\beta(z)}{\mathbf{H}_\beta(z)} = 1 - \frac{1}{\beta + 1}\sum_{n\ge 1}\frac{2z^2}{z^2_{\beta,n} - z^2}, \\ \frac{zV'_\beta(z)}{V_\beta(z)} &= -\beta + \frac{z\mathbf{H}'_\beta(z)}{\mathbf{H}_\beta(z)} = 1 - \sum_{n\ge 1}\frac{2z^2}{z^2_{\beta,n} - z^2}, \end{aligned}$$

$$\frac{zW'_\beta(z)}{W_\beta(z)} = \frac{1-\beta}{2} + \frac{\sqrt{z}\mathbf{H}'_\beta(\sqrt{z})}{2\mathbf{H}_\beta(\sqrt{z})} = 1 - \sum_{n\geq 1} \frac{z}{z^2_{\beta,n} - z}. \tag{2.14}$$

Now applying the triangle inequality $||x| - |y|| \leq |x - y|$ and Lemma 2.1 in (2.14), we see that U_β, V_β and W_β belongs to $\mathcal{S}^*_\wp$, respectively whenever

$$\begin{aligned}
\left|\frac{zU'_\beta(z)}{U_\beta(z)} - 1\right| &\leq \frac{1}{\beta+1}\sum_{n\geq 1} \frac{2r^2}{z^2_{\beta,n} - r^2} \leq \frac{1}{e},\\
\left|\frac{zV'_\beta(z)}{V_\beta(z)} - 1\right| &\leq \sum_{n\geq 1} \frac{2r^2}{z^2_{\beta,n} - r^2} \leq \frac{1}{e},\\
\left|\frac{zW'_\beta(z)}{W_\beta(z)} - 1\right| &\leq \sum_{n\geq 1} \frac{r}{z^2_{\beta,n} - r} \leq \frac{1}{e}
\end{aligned} \tag{2.15}$$

holds, where $|z| = r < z_{\beta,1}$. Now to find the largest positive radius for which (2.15) holds. Let us consider the strictly increasing continuous functions

$$\begin{aligned}
\Psi_1(r) &:= \frac{1}{\beta+1}\sum_{n\geq 1} \frac{2r^2}{z^2_{\beta,n} - r^2} - \frac{1}{e},\\
\Psi_2(r) &:= \sum_{n\geq 1} \frac{2r^2}{z^2_{\beta,n} - r^2} - \frac{1}{e},\\
\Psi_3(r) &:= \sum_{n\geq 1} \frac{r}{z^2_{\beta,n} - r} - \frac{1}{e}.
\end{aligned}$$

Since $\lim_{r\to 0}\Psi_i(r) < 0$, $\Psi'_i(r) > 0$ for $i = 1$ to 3, $\lim_{r\to z_{\beta,1}}\Psi_i(r) > 0$ for $i = 1, 2$ and $\lim_{r\to z^2_{\beta,1}}\Psi_3(r) > 0$, there exist the unique positive roots, $r_\wp(U_\beta), r_\wp(V_\beta) \in (0, z_{\beta,1})$ and $r_\wp(W_\beta) \in (0, z^2_{\beta,1})$ for Ψ_i, respectively so that the inequalities in (2.15) holds in $|z| < r_\wp(U_\beta)$, $|z| < r_\wp(V_\beta)$ and $|z| < r_\wp(W_\beta)$, respectively. Further using (2.14) in $\Psi_i(r) = 0$, respectively, we obtain the desired equations. This completes the proof. □

The Lommel function $\mathcal{L}_{u,v}$ of first kind is a particular solution of the second-order inhomogeneous Bessel differential equation

$$z^2w''(z) + zw'(z) + (z^2 - v^2)w(z) = z^{u+1},$$

where $u \pm v \notin \mathbb{Z}^-$ and is given by

$$\mathcal{L}_{u,v} = \frac{z^{u+1}}{(u-v+1)(u+v+1)}{}_1F_2\left(1; \frac{u-v+3}{2}, \frac{u+v+3}{2}; -\frac{z^2}{4}\right),$$

where $\frac{1}{2}(-u \pm v - 3) \notin \mathbb{N}$ and ${}_1F_2$ is a hypergeometric function. Since it is not normalized, so we consider the following three normalized functions involving $\mathcal{L}_{u,v}$:

$$\begin{aligned}
f_{u,v}(z) &= ((u-v+1)(u+v+1)\mathcal{L}_{u,v}(z))^{\frac{1}{u+1}},\\
g_{u,v}(z) &= (u-v+1)(u+v+1)z^{-u}\mathcal{L}_{u,v}(z),\\
h_{u,v}(z) &= (u-v+1)(u+v+1)z^{\frac{1-u}{2}}\mathcal{L}_{u,v}(\sqrt{z}).
\end{aligned} \tag{2.16}$$

Authors in [5, 1] proved the radius of starlikeness for the normalized functions expressed in terms of $\mathcal{L}_{u,v}$:

$$f_{u-\frac{1}{2},\frac{1}{2}}(z), \quad g_{u-\frac{1}{2},\frac{1}{2}}(z) \quad \text{and} \quad h_{u-\frac{1}{2},\frac{1}{2}}(z), \tag{2.17}$$

where $0 \neq u \in (-1,1)$. Now we find the radii of cardioid-starlikeness of the functions defined in (2.17). For simplicity, we write these as f_u, g_u and h_u, respectively.

Theorem 2.6. *Let $0 \neq u \in (-1,1)$ and write $\mathcal{L}_{u-\frac{1}{2},\frac{1}{2}}(z) =: \mathcal{L}_u(z)$. Then the radii of cardioid-starlikeness $r_\wp(f_u)$, $r_\wp(g_u)$ and $r_\wp(h_u)$ of the functions f_u, g_u and h_u are the smallest positive roots of the following equations, respectively:*

(i) $\begin{cases} 2er\mathcal{L}'_u(r) - (2u+1)(e-1)\mathcal{L}_u(r) = 0, & for \quad u \in (-\frac{1}{2}, 1) \\ 2er\mathcal{L}'_u(r) - (2u+1)(e+1)\mathcal{L}_u(r) = 0, & for \quad u \in (-1, -\frac{1}{2}); \end{cases}$

(ii) $2er\mathcal{L}'_u(r) - (2eu+e-2)\mathcal{L}_u(r) = 0;$

(iii) $2e\sqrt{r}\mathcal{L}'_u(\sqrt{r}) - (2eu+e-4)\mathcal{L}_u(\sqrt{r}) = 0.$

Proof. We prove the first part. Let $0 \neq u \in (0,1)$. Then using a result from [7] (also see [5, Lemma 1, p. 3358]), we can write the Lommel function $\mathcal{L}_{u-\frac{1}{2},\frac{1}{2}}$ as follows:

$$\mathcal{L}_{u-\frac{1}{2},\frac{1}{2}}(z) = \frac{z^{u+\frac{1}{2}}}{u(u+1)}{}_1F_2\left(1; \frac{u+2}{2}, \frac{u+3}{2}; -\frac{z^2}{4}\right) = \frac{z^{u+\frac{1}{2}}}{u(u+1)}\phi_0(z), \tag{2.18}$$

where

$$\phi_0(z) = \prod_{n\geq 1}\left(1 - \frac{z^2}{z^2_{u,0,n}}\right),$$

and $z_{u,0,n}$ is the simple and real n-th positive root of ϕ_0. Also $z_{u,0,n} \in (n\pi, (n+1)\pi)$ which ensures $z_{u,0,n} > z_{u,0,1} > \pi > 1$. Now with this representation, after logarithmic differentiation, from (2.16) we get

$$\frac{zf'_u(z)}{f_u(z)} = \frac{z\mathcal{L}'_{u-\frac{1}{2},\frac{1}{2}}(z)}{(u+\frac{1}{2})\mathcal{L}_{u-\frac{1}{2},\frac{1}{2}}(z)} = 1 - \frac{1}{u+\frac{1}{2}}\sum_{n\geq 1}\frac{2z^2}{z^2_{u,0,n} - z^2}.$$

Using the triangle inequality and Lemma 2.1, we have $f_u \in \mathcal{S}^*_\wp$ provided

$$T(r) := \frac{1}{u+\frac{1}{2}}\sum_{n\geq 1}\frac{2r^2}{z^2_{u,0,n} - r^2} - \frac{1}{e} \leq 0$$

holds for $|z| = r < z_{u,0,1}$, where $T(r)$ is a strictly increasing continuous function in $(0, z_{u,0,1})$. Since $\lim_{r\to 0} T(r) < 0$, $\lim_{r\to z_{u,0,1}} T(r) > 0$ and $T'(r) > 0$, there exists a root $r_\wp(f_u) \in (0, z_{u,0,1})$ so that $f_u \in \mathcal{S}^*_\wp$ in $|z| < r_\wp(f_u)$. Now for the case $u \in (-1,0)$, we proceed as in the case when $u \in (0,1)$, just replacing u by $u+1$ and ϕ_0 by ϕ_1, where

$$\phi_1(z) = {}_1F_2\left(1; \frac{u+1}{2}, \frac{u+2}{2}; -\frac{z^2}{4}\right) = \prod_{n\geq 1}\left(1 - \frac{z^2}{z^2_{u,1,n}}\right)$$

and $z_{u,1,n}$ be the n-th positive root of ϕ_1.

Proof for the part (ii) and (iii) follows in a similar fashion as in the Theorem 2.5 by applying Lemma 2.1 on the following two equations, respectively together with the triangle inequality $||x| - |y|| \leq |x - y|$:

$$\frac{zg'_u(z)}{g_u(z)} = -u + \frac{1}{2} + \frac{z\mathcal{L}'_{u-\frac{1}{2},\frac{1}{2}}(z)}{(u+\frac{1}{2})\mathcal{L}_{u-\frac{1}{2},\frac{1}{2}}(z)} = 1 - \sum_{n\geq 1}\frac{2z^2}{z^2_{u,0,n} - z^2}$$

and

$$\frac{zh_u'(z)}{h_u(z)} = \frac{3-2u}{4} + \frac{\sqrt{z}\mathcal{L}'_{u-\frac{1}{2},\frac{1}{2}}(\sqrt{z})}{2\mathcal{L}_{u-\frac{1}{2},\frac{1}{2}}(\sqrt{z})} = 1 - \sum_{n\geq 1} \frac{z}{z_{u,0,n}^2 - z},$$

where $z_{u,0,n}$ is the n-th positive root of the function ϕ_0. □

The Legendre polynomials P_n are the solutions of the Legendre differential equation:

$$((1-z^2)P_n'(z))' + n(n+1)P_n(z) = 0,$$

where $n \in \mathbb{Z}^+$ and using Rodrigues′ formula, P_n can be represented in the form:

$$P_n(z) = \frac{1}{2^n n!}\frac{d^n(z^2-1)^n}{dz^n}$$

and it also satisfies the geometric condition $P_n(-z) = (-1)^n P_n(z)$. Moreover, the odd degree Legendre polynomials $P_{2n-1}(z)$ have only real roots which satisfy

$$0 = z_0 < z_1 < \cdots < z_{n-1} \quad \text{and} \quad -z_1 < \cdots < -z_{n-1}. \tag{2.19}$$

Thus the normalized form is as follows:

$$\mathcal{P}_{2n-1}(z) := \frac{P_{2n-1}(z)}{P_{2n-1}'(0)} = z + \sum_{k=2}^{2n-1} a_k z^k = a_{2n-1} z \prod_{k=1}^{n-1}(z^2 - z_k^2). \tag{2.20}$$

Theorem 2.7. *The radii of cardioid-starlikeness $r_\wp(\mathcal{P}_{2n-1}) \in (0, z_1)$ of the normalized odd degree Legendre polynomial is the smallest positive root of the following equation:*

$$er\mathcal{P}_{2n-1}'(r) - (e-1)\mathcal{P}_{2n-1}(r) = 0.$$

Proof. From (2.20), after logarithmic differentiation, we obtain

$$\frac{z\mathcal{P}_{2n-1}'(z)}{\mathcal{P}_{2n-1}(z)} = 1 - \sum_{k=1}^{n-1} \frac{2z^2}{z_k^2 - z^2}. \tag{2.21}$$

Now applying Lemma 2.1 on (2.21), we have $\mathcal{P}_{2n-1} \in \mathcal{S}_\wp^*$ whenever

$$\left|\frac{z\mathcal{P}_{2n-1}'(z)}{\mathcal{P}_{2n-1}(z)} - 1\right| \leq \sum_{k=1}^{n-1} \frac{2r^2}{z_k^2 - r^2} \leq \frac{1}{e}, \tag{2.22}$$

where $|z| = r < z_1$ and z_k satisfies the condition given in (2.19). Now let us consider the strictly increasing continuous function

$$T(r) := \sum_{k=1}^{n-1} \frac{2r^2}{z_k^2 - r^2} - \frac{1}{e}, \quad r \in (0, z_1).$$

We have to show that $T(r) \leq 0$ in $|z| \leq r < z_1$ so that (2.22) holds. Since $\lim_{r\to 0} T(r) < 0$, $\lim_{r\to z_1} T(r) > 0$ and $T'(r) > 0$, there exists a unique positive root $r_\wp(\mathcal{P}_{2n-1}) \in (0, z_1)$ of $T(r)$ such that $\mathcal{P}_{2n-1} \in \mathcal{S}_\wp^*$ in $|z| < r_\wp(\mathcal{P}_{2n-1})$. □

Remark 2.1 (**Generalization for $\mathcal{S}^*(\psi)$**). Consider the Ma-Minda function ψ as defined in (1.1). Let $a \in \psi(\mathbb{D}) \cap \mathbb{R}$, r_a is the radius depending on a and assume the maximal disk $|w-a| < r_a$ such that

$$\{w : |w-a| < r_a\} \subset \psi(\mathbb{D}).$$

Then we can easily extend Theorem 2.1 ($\psi(z) \neq (1+z)/(1-z)$), Theorem 2.3, Theorem 2.4, Theorem 2.5, Theorem 2.6 and Theorem 2.7 for $\mathcal{S}^*(\psi)$ using the corresponding radius r_1.

Here we mention the $\mathcal{S}^*(\psi)$-radius for the special functions without proof using Remark (2.1). In particular, we show that there exists an $\alpha \in (0,1]$ such that $R[\mathcal{S}^*(\psi)] \geq R[\mathcal{S}^*(1+\alpha z)]$, where we denote $\mathcal{S}^*(\psi)$-radius by $R[\mathcal{S}^*(\psi)]$.

In the following theorems we mean $R[\mathcal{S}^*(\psi)] = R[\mathcal{S}^*(1+r_1 z)]$.

Theorem 2.8 (Bessel function $\mathcal{J}_\beta$)**.** *Let $\beta > 0$. Then the $\mathcal{S}^*(\psi)$-radii $r_\wp(f_\beta)$, $r_\wp(g_\beta)$ and $r_\wp(h_\beta)$ of the functions f_β, g_β and h_β are the smallest positive root of the following equations, respectively:*

(i) $r\mathcal{J}'_\beta(r) + \beta(1-r_1)\mathcal{J}_\beta(r) = 0$;
(ii) $r\mathcal{J}'_\beta(r) - (\beta - r_1)\mathcal{J}_\beta(r) = 0$;
(iii) $\sqrt{r}\mathcal{J}'_\beta(\sqrt{r}) - (\beta - 2r_1)\mathcal{J}_\beta(\sqrt{r}) = 0$.

Theorem 2.9 (Struve function $\mathbf{H}_\beta$)**.** *Let $|\beta| \leq 1/2$. Then the $\mathcal{S}^*(\psi)$-radii $r_\wp(U_\beta)$, $r_\wp(V_\beta)$ and $r_\wp(W_\beta)$ of the functions U_β, V_β and W_β are the smallest positive root of the following equations, respectively:*

(i) $r\mathbf{H}'_\beta(r) - (1-r_1)(\beta+1)\mathbf{H}_\beta(r) = 0$;
(ii) $r\mathbf{H}'_\beta(r) - ((1+\beta) - r_1)\mathbf{H}'_\beta(r) = 0$;
(iii) $\sqrt{r}\mathbf{H}'_\beta(\sqrt{r}) - (1+\beta-2r_1)\mathbf{H}_\beta(\sqrt{r}) = 0$.

Theorem 2.10 (Lommel function $\mathcal{L}_{u,v}$)**.** *Let $0 \neq u \in (-1,1)$ and write $\mathcal{L}_{u-\frac{1}{2},\frac{1}{2}}(z) =: \mathcal{L}_u(z)$. Then the $\mathcal{S}^*(\psi)$-radii $r_\wp(f_u)$, $r_\wp(g_u)$ and $r_\wp(h_u)$ of the functions f_u, g_u and h_u are the smallest positive root of the following equations, respectively:*

(i) $\begin{cases} 2r\mathcal{L}'_u(r) - (2u+1)(1-r_1)\mathcal{L}_u(r) = 0, & for \quad u \in (-\frac{1}{2}, 1) \\ 2r\mathcal{L}'_u(r) - (2u+1)(1+r_1)\mathcal{L}_u(r) = 0, & for \quad u \in (-1, -\frac{1}{2}); \end{cases}$
(ii) $2r\mathcal{L}'_u(r) - (2u+1-2r_1)\mathcal{L}_u(r) = 0$;
(iii) $2\sqrt{r}\mathcal{L}'_u(\sqrt{r}) - (2u+1-4r_1)\mathcal{L}_u(\sqrt{r}) = 0$.

Theorem 2.11 (Legendre polynomials P_n)**.** *The $\mathcal{S}^*(\psi)$-radius $r_\wp(\mathcal{P}_{2n-1}) \in (0, z_1)$ of the normalized odd degree Legendre polynomial is the smallest positive root of the following equation:*

$$r\mathcal{P}'_{2n-1}(r) - (1-r_1)\mathcal{P}(r) = 0.$$

Corollary 2.12. *If $\alpha = r_1$ be the radius of the lagest disk $\{w : |w-1| < \alpha\}$ inside $\psi(\mathbb{D})$, where*

(i) $\alpha = \min\left\{\left|1 - \frac{1+A}{1+B}\right|, \left|1 - \frac{1-A}{1-B}\right|\right\} = \frac{A-B}{1+|B|}$ *when* $\psi(z) = \frac{1+Az}{1+Bz}$, *where* $-1 \leq B < A \leq 1$;
(ii) $\alpha = \sqrt{2 - 2\sqrt{2} + \sqrt{-2+2\sqrt{2}}}$ *when* $\psi(z) = \sqrt{2} - (\sqrt{2}-1)\sqrt{\frac{1-z}{1+2(\sqrt{2}-1)z}}$;
(iii) $\alpha = \sqrt{2} - 1$ *when* $\psi(z) = \sqrt{1+z}$;
(iv) $\alpha = e - 1$ *when* $\psi(z) = e^z$;
(v) $\alpha = 2 - \sqrt{2}$ *when* $\psi(z) = z + \sqrt{1+z^2}$;
(vi) $\alpha = \frac{e-1}{e+1}$ *when* $\psi(z) = \frac{2}{1+e^{-z}}$;
(vii) $\alpha = \sin 1$ *when* $\psi(z) = 1 + \sin z$.

Then Theorem 2.8, Theorem 2.9, Theorem 2.10 and Theorem 2.11 hold true for the class $\mathcal{S}^*(\psi)$.

2.2. **Majorization result for $\mathcal{S}^*_\wp$.**

Theorem 2.13. *Let $f \in \mathcal{A}$ and $g \in \mathcal{S}^*_\wp$. If f is majorized by g in $\mathbb{D}$. Then*

$$|f'(z)| \leq |g'(z)| \quad in \quad |z| \leq r_0, \tag{2.23}$$

where $r_0 \approx 0.380056$ is the smallest positive root of the equation

$$(1-r^2)(1-re^r) - r = 0.$$

Proof. Let $f \in \mathcal{A}$ and $g \in \mathcal{S}^*_\wp$, then we have

$$\frac{zg'(z)}{g(z)} = 1 + \omega(z)e^{\omega(z)}$$

or

$$\frac{g(z)}{g'(z)} = \frac{z}{1+\omega(z)e^{\omega(z)}}, \tag{2.24}$$

where ω is a Schwarz function. Let $\omega(z) = Re^{i\theta}$, where $R \leq r = |z| < 1$, then we have

$$\left|\frac{g(z)}{g'(z)}\right| = \left|\frac{z}{1+\omega(z)e^{\omega(z)}}\right| \leq \frac{r}{1-re^r}. \tag{2.25}$$

Now by definition of majorization, we have $f(z) = \Phi(z)g(z)$, where $\Phi(z)$ is analytic function satisfying $|\Phi(z)| \leq 1$ in $\mathbb{D}$ such that

$$f'(z) = g'(z)\left(\Phi'(z)\frac{g(z)}{g'(z)} + \Phi(z)\right).$$

Thus we have

$$|f'(z)| \leq |g'(z)|\left(\frac{1-|\Phi(z)|^2}{1-r^2}\frac{r}{1-re^r} + |\Phi(z)|\right) = |g'(z)|h(r,\beta), \tag{2.26}$$

where

$$h(r,\beta) = \frac{(1-\beta^2)r}{(1-r^2)(1-re^r)} + \beta.$$

Now to arrive at equation (2.23), it suffices to find r_0 such that $h(r,\beta) \leq 1$ holds. Or equivalently,

$$0 \leq (1-r^2)(1-re^r) - (1+\beta)r =: k(r,\beta).$$

Since $\frac{\partial}{\partial\beta}k(r,\beta) = -r$, therefore $\min_\beta k(r,\beta) = k(r,1)$. Write $k(r,1)$ as $k(r)$ and note that $k(0) = 1$ and $k(1) = -1$. So by continuity of $k(r)$, there exists a point r_0 such that $k(r) \geq 0$ for $0 \leq r \leq r_0$, where r_0 is the smallest positive root of the equation

$$(1-r^2)(1-re^r) - r = 0.$$

This ends the proof. □

2.3. **Convolution radius.** First, we need to recall the following result due to Ruscheweyh and Sheil-Small:

Lemma 2.3 ([25], p. 126)**.** *Suppose that either $g \in \mathcal{C}$, $h \in \mathcal{S}^*$ or else $g, h \in \mathcal{S}^*_{1/2}$. Then for any analytic function G in $\mathbb{D}$, we have*

$$\frac{g * hG(z)}{g * h(z)} \in \overline{co}G(\mathbb{D}),$$

where $\overline{co}G(\mathbb{D})$ is the closed convex hull of $G(\mathbb{D})$.

Observe that if we consider the function $\phi \in \mathcal{P}$, which is starlike but not convex then we can not directly apply Theorem 1.1 as such function do exist, for instance $\phi(z) = 1 + ze^z$. But keenly observing the proof of Lemma 2.3, we see that the unit disk $\mathbb{D}$ can be replaced by the sub-disk $\mathbb{D}_r := \{z : |z| < r\}$, where $0 < r \leq 1$ and consequently, we obtain the following modified result. Since the proof is similar, so it is omitted here.

Lemma 2.4. *Suppose either $g \in \mathcal{C}$, $h \in \mathcal{S}^*$ or else $g, h \in \mathcal{S}^*_{1/2}$. Then for any analytic function G in $\mathbb{D}_r$, we have $(g * hG(z))/(g * h(z)) \in \overline{co}G(\mathbb{D}_r)$, where $r \in [0, 1]$.*

This immediately gives

Theorem 2.14. *Let r_0 be the radius of convexity of ψ. If $g \in \mathcal{C}$ and $f \in \mathcal{S}^*(\psi)$. Then $f * g \in \mathcal{S}^*(\psi)$ in $|z| < r$, where $r = \min\{r_0, 1\}$.*

Corollary 2.15. *Let $f \in \mathcal{S}^*_\wp$ and $g \in \mathcal{C}$. Then $f * g \in \mathcal{S}^*_\wp$ in $\mathbb{D}_{r_0}$, where $r_0 = (3 - \sqrt{5})/2$ is the radius of convexity of $\wp$.*

Now consider the operators $\mathcal{F}_i : \mathcal{A} \to \mathcal{A}$ defined by

$$\mathcal{F}_1(f)(z) = f * g_1(z) = zf'(z)$$

$$\mathcal{F}_2(f)(z) = f * g_2(z) = \frac{1}{2}(f(z) + zf'(z))$$

$$\mathcal{F}_3(f)(z) = f * g_3(z) = \frac{k+1}{z^k}\int_0^z t^{k-1}f(t)dt, \quad \Re k > 0,$$

where $g_3(z) = \sum_{n=1}^{\infty}(k+1)/(k+n)z^n$, $g_2(z) = (z - z^2/2)/(1-z^2)^2$ and $g_1(z) = z/(1-z)^2$. Note that the function g_1 is convex in $|z| < 2 - \sqrt{3}$, g_2 is convex in $|z| < 1/2$ while $g_3 \in \mathcal{C}$. The above defined operators were introduced by Alexander, Livingston and Bernardi, respectively. Now we obtain the following result, where $\mathcal{S}^*_{SG} := \mathcal{S}^*(\frac{2}{e^{-z}+1})$, $\mathcal{S}^*_C := \mathcal{S}^*(1 + 4z/3 + 2z^2/3)$ and $\mathcal{S}^*_s := \mathcal{S}^*(1 + \sin z)$:

Corollary 2.16. *Let $\mathcal{F}_i$, $i = 1$ to 3 be the operators as defined above.*

(*i*) *Let $f \in \mathcal{S}^*_\wp$. Then $\mathcal{F}_i(f) \in \mathcal{S}^*_\wp$ in $\mathbb{D}_{r_i}$, where $r_1 = 2 - \sqrt{3}$, $r_2 = (3 - \sqrt{5})/2$ and $r_3 = (3 - \sqrt{5})/2$.*

(*ii*) *Let $f \in \mathcal{S}^*_C$. Then $\mathcal{F}_i(f) \in \mathcal{S}^*_C$ in $\mathbb{D}_{r_i}$, where $r_1 = 2 - \sqrt{3}$, $r_2 = 1/2$ and $r_3 = 1/2$.*

(*iii*) *Let $f \in \mathcal{S}^*_s$. Then $\mathcal{F}_i(f) \in \mathcal{S}^*_s$ in $\mathbb{D}_{r_i}$, where $r_1 = 2 - \sqrt{3}$, $r_2 = 0.345$ and $r_3 = 0.345$.*

(*iv*) *Let $f \in \mathcal{S}^*_{SG}$. Then $\mathcal{F}_i(f) \in \mathcal{S}^*_{SG}$ in $\mathbb{D}_{r_i}$, where $r_1 = 2 - \sqrt{3}$, $r_2 = 1/2$ and $r_3 = 1$.*

In 2010, Ali et al. [2] dealt with the problem of finding $\mathcal{S}^*(\psi)$-radii of the convolution $f * g$, between two starlike functions. In particular, they showed that if $f, g \in \mathcal{S}^*$ and $h_\rho(z) = f * g(\rho z)/\rho$, then $h_\rho \in \mathcal{SL}^*$ for $0 \leq \rho \leq (\sqrt{5} - 2)/(\sqrt{2} - 1) \approx 0.09778$. They used the property of the function ψ being convex. Now using Theorem 2.14, we can obtain the result even for the case when $\psi(\mathbb{D})$ is starlike. Here we have shown the usability of radius of convexity of ψ.

Theorem 2.17. *Let $f, g \in \mathcal{S}^*$ and $h_\rho(z) := f * g(\rho z)/\rho$. Then*

(*i*) *$h_\rho \in \mathcal{S}^*_\wp$ for $0 \leq \rho \leq (2e - \sqrt{4e^2 - 2e + 1})/(2e - 1) \approx 0.0957$,*

(*ii*) *$h_\rho \in \mathcal{S}^*_C$ for $0 \leq \rho \leq (3 - \sqrt{7})/2 \approx 0.177124$,*

(*iii*) $h_\rho \in \mathcal{S}_s^*$ *for* $0 \le \rho \le (\sqrt{\sin 1^2 + 2\sin 1 + 4} - 2)/(2 + \sin 1) \approx 0.185835$,
(*iv*) $h_\rho \in \mathcal{S}_b^*$ *for* $0 \le \rho \le (2e - \sqrt{3e^2 + e^{2/e}})/(e + e^{1/e}) \approx 0.122919$,
(*v*) $h_\rho \in \mathcal{S}_{SG}^*$ *for* $0 \le \rho \le (\sqrt{7e^2 + 6e + 3} - 2(1 + e))/(3e + 1) \approx 0.108309$.

The constants are best possible.

Proof.

(*i*) Let $H(z) = z + \sum_{n=2}^{\infty} n^2 z^n = (z(1+z))/(1-z)^3$. It is easy to see that

$$\left| \frac{zH'(z)}{H(z)} - \frac{1+r^2}{1-r^2} \right| \le \frac{4r}{1-r^2}, \quad |z| = r < 1. \tag{2.27}$$

Now by Lemma 2.1, the disk (2.27) lies inside the cardioid $\wp(\mathbb{D})$, provided

$$\frac{4r}{1-r^2} \le \frac{1+r^2}{1-r^2} - 1 + \frac{1}{e}$$

which in turn gives $r \le r_0 := (2e - \sqrt{4e^2 - 2e + 1})/(2e - 1)$. Define the function $h : \mathbb{D} \to \mathbb{C}$ by $h(z) := f(z) * g(z)$. Then $h(z) = F(z) * G(z) * H(z)$, where F and G are, respectively defined as $zF'(z) = f(z)$ and $zG'(z) = g(z)$. Since $f, g \in \mathcal{S}^*$, it follows that $F * G \in \mathcal{C}$. Also, $H(r_0 z)/r_0 \in \mathcal{S}_\wp^*$. Hence, using Theorem 2.14, we have

$$F(z) * G(z) * H(\rho_0)/\rho_0 \in \mathcal{S}_\wp^*,$$

where $\rho_0 = \min\{r_0, r_c\} = r_0$ and $r_c = (3 - \sqrt{5})/2$ is the radius of convexity of $\wp$. For $z = -\rho_0$, $zH'(z)/H(z) = (1 + 4z + z^2)/(1 - z^2) = 1 - 1/e$, which implies that ρ_0 is sharp.

Rest part's proof also follow in a similar fashion. □

Theorem 2.18. *Let* $f, g \in \mathcal{S}^*$ *and* $h_\rho(z) := f * g(\rho z)/\rho$. *Then* $h_\rho(z) \in \mathcal{S}^*\left(\frac{1+Az}{1+Bz}\right)$ *for*

$$0 \le \rho \le \frac{2(B^2 - 1) + \sqrt{4(1 - B^2)^2 + (A - B)^2}}{A - B} =: \rho_0,$$

where $-1 < B < A \le 1$.

Proof. Since for the function $p(z) \prec (1 + Az)/(1 + Bz)$, we have

$$\left| p(z) - \frac{1 - AB}{1 - B} \right| \le \frac{A - B}{1 - B^2}. \tag{2.28}$$

Therefore, for the disk (2.27) to lie inside the disk (2.28), we must have

$$\frac{1 - AB}{1 - B^2} - \frac{A - B}{1 - B^2} \le \frac{1 + r^2}{1 - r^2} \le \frac{1 - AB}{1 - B^2} + \frac{A - B}{1 - B^2}$$

and

$$\frac{4r}{1 - r^2} \le \frac{A - B}{1 - B^2}$$

which upon simplification hold for $r \le r_0 = \sqrt{(A - B)/(2 + A + B)}$ and $r \le \rho_0$ respectively, where ρ_0 is the smallest positive root of the following equation

$$(A - B)r^2 + 4(1 - B^2)r - (A - B) = 0.$$

Since $\min\{r_0, \rho_0\} = \rho_0$ and the class $\mathcal{S}^*\left(\frac{1+Az}{1+Bz}\right)$ is closed under convolution with convex functions, now the result follows in a similar way as in the part (*i*) of Theorem 2.17. □

3. Extremal problem for the class $\mathcal{S}^*(\psi)$

In 1961, Goluzin [13] obtained the set of extremal functions $f(z) = z/(1-xz)^2$, $|x| = 1$ for the problem of maximization of the quantity $\Re\Phi\left(\log(f(z)/z)\right)$ or $|\Phi\left(\log(f(z)/z)\right)|$ over the class $\mathcal{S}^*$, where ϕ is a non-constant entire function. In 1973, MacGregor [21] proved the result for the class $\mathcal{S}^*(\alpha) := \{f \in \mathcal{A} : \Re(zf'(z)/f(z)) > \alpha, \alpha \in [0,1)\}$. We observed that the proof given by MacGregor can be generalized to the Ma-Minda class $\mathcal{S}^*(\psi)$. Thus we have the following result:

Theorem 3.1. *Suppose Φ is a non-constant entire function and $0 < |z_0| < 1$ and assume that the class $\mathcal{S}^*(\psi)$ is closed. Then maximum of either*

$$\Re\Phi\left(\log\frac{f(z_0)}{z_0}\right) \quad or \quad \left|\Phi\left(\log\frac{f(z_0)}{z_0}\right)\right| \tag{3.1}$$

for functions in the class $\mathcal{S}^*(\psi)$ is attained only when the function is of the form

$$f(z) = z\exp\int_0^{\zeta z}\frac{\psi(t)-1}{t}dt, \tag{3.2}$$

where $|\zeta| = 1$.

Proof. Since the class $\mathcal{S}^*(\psi)$ is compact, therefore the problem under consideration has a solution. Moreover, in view of a result of Goluzin [13], in (3.1) it suffices to consider the continuous functional

$$\Re\Phi\left(\log\frac{f(z_0)}{z_0}\right).$$

Let $f \in \mathcal{S}^*(\psi)$. Then using a result from [19], $f(z)/z \prec f_0(z)/z =: F(z)$, where $f_0(z) = z\exp\int_0^z \frac{\psi(t)-1}{t}dt$ or equivalenlty $\log(f(z)/z) \prec \log F(z)$. Thus,

$$g(z) = \Phi\left(\log\frac{f(z)}{z}\right) \prec \Phi(\log F(z)) = G(z).$$

Note that G is also non-constant as is Φ. Thus for each $r \in (0,1)$ by subordination principle, we obtain $g(\overline{\mathbb{D}}_r) \subset G(\overline{\mathbb{D}}_r) = \Omega$. Since $G(xz) \prec G(z)$ for $|x| \leq 1$ is obvious, therefore for $|z_0| = r$, we have $\{g(z_0) : g \prec G \text{ in } \mathbb{D}\} = \Omega$. Now by considering a support line to the compact set Ω, we conclude that

$$\max_{f\in\mathcal{S}^*(\psi)} \Re\Phi\left(\log\frac{f(z_0)}{z_0}\right) = \Re w_1, \quad w_1 \in \partial\Omega.$$

Since G is also an open map, therefore there exists a point z_1 where $|z_1| = r$ and $G(z_1) = w_1$ such that among finitely many w_1, for one suitable w_1, we have

$$\Phi\left(\log\frac{f(z_0)}{z_0}\right) = w_1,$$

where f is the solution for the extremal problem. Now by the well known Lindelöf Principle, we have

$$\Phi\left(\log\frac{f(z)}{z}\right) = \Phi(\log F(xz)), \tag{3.3}$$

that is, if f is the desired solution, then (3.3) holds for some x, $|x| = 1$. Since Φ is non-constant analytic function, so we may write

$$\Phi(w) = c_0 + c_n w^n + c_{n+1}w^{n+1} + \cdots;\ c_n \neq 0.$$

If we set $\log(f(z)/z) = \alpha_1 z + \alpha_2 z^2 + \cdots$ and $\log(F(z)) = \beta_1 z + \beta_2 z^2 + \cdots$, then from (3.3), comparing the coefficients, we get $c_n \alpha_1^n = c_n \beta_1^n$. Or equivalently, $\alpha_1^n = \beta_1^n$, which in particular implies that $|\alpha_1| = |\beta_1|$. Since $\log(f(z)/z) \prec \log F(xz)$, $|\alpha_1| = |\beta_1|$ is possible only if $\log(f(z)/z) = \log F(xyz)$ for some $|y| = 1$. Therefore, we conclude that

$$f(z) = z \exp \int_0^{uz} \frac{\psi(t) - 1}{t} dt,$$

where $|u| = 1$ if f is a solution to the extremal problem. □

Now as an application of the Theorem 3.1, we obtain a result due to MacGregor [21]:

Corollary 3.2. [21] *Suppose Φ is a non-constant entire function and $0 < |z_0| < 1$. Then the maximum of the expression* (3.1) *for functions in the class $\mathcal{S}^*(\alpha)$ is attained only when the function is of the form*

$$f(z) = \frac{z}{(1 - \zeta z)^{2-2\alpha}}, \ |\zeta| = 1.$$

Proof. If $f \in \mathcal{S}^*(\alpha)$, then $f(z)/z \prec 1/(1-z)^{2-2\alpha}$ and hence the result.

Corollary 3.3. *Suppose Φ is a non-constant entire function and $0 < |z_0| < 1$. Then the maximum of the expression* (3.1) *for functions in the class $\mathcal{S}^*_\wp$ is attained only when the function is of the form*

$$f(z) = z \exp(e^{\zeta z} - 1), \ |\zeta| = 1.$$

Proof. If $f \in \mathcal{S}^*_\wp$, then $f(z)/z \prec \exp(e^z - 1)$ and hence the result. □

4. Some sufficient conditions for $\mathcal{S}^*_\wp$

In this section, we determine the sufficient conditions for the functions $z/(1 + \sum_{k=1}^{\infty} a_k z^k)$, $z/(1+z^n)^k$ and certain other types of functions to be in $\mathcal{S}^*_\wp$.

Theorem 4.1. *Let $f(z) = z/(1 + \sum_{k=1}^{\infty} a_k z^k)$. If the coefficients of f satisfy*

$$|1 - a| + \sum_{k=1}^{\infty} (R_a + |1 - a - k|)|a_k| \leq R_a,$$

where a and R_a is as defined in Lemma 2.1. Then $f \in \mathcal{S}^*_\wp$.

Proof. For $f(z) = z/(1 + \sum_{k=1}^{\infty} a_k z^k)$, we have

$$\left| \frac{zf'(z)}{f(z)} - a \right| = \left| 1 - a - \frac{\sum_{k=1}^{\infty} k a_k z^k}{1 + \sum_{k=1}^{\infty} a_k z^k} \right|.$$

Thus by Lemma 2.1, $f \in \mathcal{S}^*_\wp$, if

$$\left| 1 - a - \frac{\sum_{k=1}^{\infty} k a_k z^k}{1 + \sum_{k=1}^{\infty} a_k z^k} \right| \leq R_a.$$

The above inequality holds whenever

$$|1 - a| + \sum_{k=1}^{\infty} |1 - a - k||a_k| r^k \leq R_a (1 - \sum_{k=1}^{\infty} |a_k| r^k)$$

or equivalently,

$$|1-a|+\sum_{k=1}^{\infty}(|1-a-k|+R_a)|a_k|r^k \leq R_a.$$

Letting r tends to 1^-, completes the proof. □

Theorem 4.2. *Let $f(z)=z/(1+z^k)^n$, where $n,k\in\mathbb{Z}^+$ are fixed. Then $f\in\mathcal{S}^*_\wp$ for*

$$|z|<\left(\frac{R_a-|1-a|}{R_a+|1-a-kn|}\right)^{1/k},$$

where a and R_a is as defined in Lemma 2.1.

Proof. For $f(z)=z/(1+z^k)^n$, we have

$$\left|\frac{zf'(z)}{f(z)}-a\right|=\left|1-a\frac{knz^k}{1+z^k}\right|.$$

Thus by Lemma 2.1, $f\in\mathcal{S}^*_\wp$, if

$$\left|1-a-\frac{knz^k}{1+z^k}\right|<R_a.$$

The above inequality holds whenever

$$|1-a|+|1-a-kn||z|^k<R_a(1-|z|^k)$$

which simplifies to

$$|z|^k<\frac{R_a-|1-a|}{R_a+|1-a-kn|}.$$

Hence the result. □

Theorem 4.3. *Let $p(z)$ be a polynomial such that $p(0)=1$ and $\deg p(z)=m$. Let $R=\min\{|z|:p(z)=0,z\neq 0\}$. Then the function $f(z)=z(p(z))^{\beta/m}\in\mathcal{S}^*_\wp$ for*

$$|z|<\frac{R(R_a-|1-a|)}{|\beta|+R_a-|1-a|},$$

where a and R_a is as defined in Lemma 2.1.

Proof. Assume that z_k, $(k=1,2,...,m)$ are zeros of the polynomial $p(z)$. For the function $f(z)=z(p(z))^{\beta/m}$, we have

$$\frac{zf'(z)}{f(z)}=1+\frac{\beta}{m}\sum_{k=1}^{\infty}\frac{z}{z-z_k}$$

or equivalently,

$$\frac{zf'(z)}{f(z)}-a=1-a+\frac{\beta}{m}\sum_{k=1}^{\infty}\left(\frac{z}{z-z_k}+\frac{r^2}{R^2-r^2}-\frac{r^2}{R^2-r^2}\right).$$

Thus by Lemma 2.1, $f\in\mathcal{S}^*_\wp$, if

$$(|1-a|-R_a)(R^2-r^2)+|\beta|(Rr+r^2)<0.$$

The above inequality is satisfied if $|z|=r<R(R_a-|1-a|)/(|\beta|+R_a-|1-a|)$. This completes the proof. □

5. Subordination results for $\mathcal{S}^*_\wp$

To prove our differential subordination results, we will need the following lemma due to Miller and Mocanu:

Lemma 5.1. [23] *Let q be the univalent in $\mathbb{D}$. Let θ and ϕ be analytic in a domain D containing $q(\mathbb{D})$, with $\phi(w) \neq 0$, when $w \in q(\mathbb{D})$. Set $Q(z) = zq'(z)\phi(q(z))$, $h(z) = \theta(q(z)) + Q(z)$ and suppose that Q is starlike. In addition, assume that*

$$\Re\frac{zh'(z)}{Q(z)} = \Re\left(\frac{\theta'(q(z))}{\phi(q(z))} + \frac{zQ'(z)}{Q(z)}\right) > 0.$$

If p is analytic in $\mathbb{D}$, with $p(0) = q(0)$, $p(\mathbb{D}) \subset D$ and

$$\theta(p(z)) + zp'(z)\phi(p(z)) \prec h(z), \tag{5.1}$$

then $p \prec q$ and q is the best dominant.

Theorem 5.1. *Let $0 < \alpha < 1$, $0 < B < A < 1$ and $k = 1 + \sqrt{2}$. If $p \in \mathcal{A}_0$ and satisfies the differential subordination*

$$1 + \beta zp'(z) \prec \wp(z),$$

then

(*i*) $p(z) \prec \phi_0(z) := 1 + \frac{z}{k}\left(\frac{k+z}{k-z}\right)$ *for* $\beta \geq \frac{k(e-1)(k+1)}{e(k-1)} \approx 3.68427$.
(*ii*) $p(z) \prec \sqrt{1+z}$ *for* $\beta \geq \frac{e-1}{\sqrt{2}-1}$.
(*iii*) $p(z) \prec \frac{1+Az}{1+Bz}$ *for* $\beta \geq \max\left\{\frac{(1-1/e)(1-B)}{A-B}, \frac{(e-1)(1+B)}{A-B}\right\}$.
(*iv*) $p(z) \prec 1 + \sin z$ *for* $\beta \geq \frac{e-1}{\sin 1}$.
(*v*) $p(z) \prec z + \sqrt{1+z^2}$ *for* $\beta \geq \frac{e-1}{\sqrt{2}}$.
(*vi*) $p(z) \prec e^z$ *for* $\beta \geq 1$.

The bounds are sharp.

Proof. We set $\theta(w) = 1$ and $\phi(w) = \beta \neq 0$. Let $\Psi_\beta(z, p(z)) := 1 + \beta zp'(z)$.

(*i*) The differential equation $\Psi_\beta(z, p(z)) = \wp(z)$ has an analytic solution given by

$$q_\beta(z) = 1 + \frac{1}{\beta}(e^z - 1).$$

Since the function $Q(z) = zq'_\beta(z)\phi(q_\beta(z)) = \wp(z) - 1 = ze^z \in \mathcal{S}^*$ and the function $h(z) = \theta(q_\beta(z)) + Q(z)$ satisfies

$$\Re\left(\frac{zh'(z)}{Q(z)}\right) = \Re\left(\frac{zQ'(z)}{Q(z)}\right) > 0.$$

Therefore, by Lemma 5.1, the following differential subordination implication holds:

$$\Psi_\beta(z, p(z)) \prec 1 + \beta zq'_\beta(z) \Rightarrow p \prec q_\beta.$$

Now to show that $p \prec \phi_0$, we need to show that $q_\beta \prec \phi_0$ and for $q_\beta \prec \phi_0$ it is necessary that $\phi_0(-1) \leq q_\beta(-1)$ and $q_\beta(1) \leq \phi_0(1)$ which gives $\beta \geq \beta_1$ and $\beta \geq \beta_2$ respectively, where

$$\beta_1 = \frac{k(e-1)(k+1)}{e(k-1)} \quad \text{and} \quad \beta_2 = \frac{k(e-1)(k-1)}{k+1}.$$

Let $\beta \geq \max\{\beta_1, \beta_2\} = \beta_1$. Since the difference of the square of the distances from the point $(1,0)$ to the points on the boundary curves $\phi_0(e^{i\theta})$ and $q_\beta(e^{i\theta})$, respectively is given by

$$T(\theta) := \frac{4(1+\sin^2\theta)}{(k^2-2k\cos\theta+1)^2} - \frac{e^{2\cos\theta}+1}{\beta^2} + \frac{2e^{\cos\theta}\cos(\sin\theta)}{\beta}, \quad 0 \leq \theta \leq \pi.$$

A calculation shows that $T'(\theta) \leq 0$. Therefore, $T(\theta)$ decreases on $[0,\pi]$, which further implies that the condition $\phi_0(-1) \leq q_\beta(-1) \leq q_\beta(1) \leq \phi_0(1)$, that is $\beta \geq \beta_1$ is also sufficient for $q_\beta \prec \phi_0$. Hence the result. Rest of the results follow in a similar fashion. □

Theorem 5.2. *Let $0 < \alpha < 1$, $0 < B < A < 1$ and $k = 1+\sqrt{2}$. If $p \in \mathcal{A}_0$ and satisfies the differential subordination*

$$1 + \beta\frac{zp'(z)}{p(z)} \prec \wp(z),$$

then

(*i*) $p(z) \prec e^z$ *for* $\beta \geq e-1$.
(*ii*) $p(z) \prec \phi_0(z)$ *for* $\beta \geq (1/e-1)/\log\left(1-\frac{k-1}{k(k+1)}\right) \approx 3.3583$.
(*iii*) $p(z) \prec \sqrt{1+z}$ *for* $\beta \geq \frac{e-1}{\log\sqrt{2}}$.
(*iv*) $p(z) \prec \frac{1+Az}{1+Bz}$ *for* $\beta \geq \max\left\{\frac{1-1/e}{\log((1-B)/(1-A))}, \frac{e-1}{\log((1+A)/(1+B))}\right\}$.
(*v*) $p(z) \prec 1+\sin z$ *for* $\beta \geq \frac{e-1}{\log(1+\sin 1)}$.
(*vi*) $p(z) \prec z+\sqrt{1+z^2}$ *for* $\beta \geq \frac{e-1}{\log(k)}$.

The bounds are sharp.

Proof. Let $\theta(w) = 1$ and $\phi(w) = \beta/\omega \neq 0$. Let $\Phi_\beta(z,p(z)) := 1+\beta\frac{zp'(z)}{p(z)}$.

(*i*) The differential equation $\Phi_\beta(z,p(z)) = \wp(z)$ has an analytic solution given by

$$q_\beta(z) = \exp\left(\frac{e^z-1}{\beta}\right).$$

Since the function $Q(z) = zq'_\beta(z)\phi(q_\beta(z)) = \wp(z) - 1 = ze^z \in \mathcal{S}^*$ and $h(z) = \theta(q_\beta(z)) + Q(z)$ satisfies $\Re(zh'(z)/Q(z)) = \Re(zQ'(z)/Q(z)) > 0$. Therefore, by Lemma 5.1, the following differential subordination implication holds:

$$\Phi_\beta(z,p(z)) \prec 1+\beta\frac{zq'_\beta(z)}{q_\beta(z)} \Rightarrow p \prec q_\beta.$$

Now to show that $p \prec e^z$, we need to show that $q_\beta \prec e^z$ and for $q_\beta \prec e^z$, it is necessary that $e^{-1} \leq q_\beta(-1)$ and $q_\beta(1) \leq e$ which gives $\beta \geq \beta_1$ and $\beta \geq \beta_2$ respectively, where

$$\beta_1 = (1-1/e) \quad \text{and} \quad \beta_2 = e-1.$$

Let $\beta \geq \beta_2$. To prove that $q_\beta(z) \prec e^z$, now it only suffices to show that $(e^z-1)/\beta \prec z$, which is equivalent to prove that for $z \in \overline{\mathbb{D}}$

$$|(e^z-1)/\beta| \leq 1, \tag{5.2}$$

and thus, for $z = e^{i\theta}$, we have

$$|(e^{e^{i\theta}}-1)/\beta| = \sqrt{(e^{2\cos\theta}+1-2e^{\cos\theta}\cos(\sin\theta))/\beta^2} =: \sqrt{T(\theta)}.$$

Note that $T(\theta) = T(-\theta)$. So we may consider the interval $0 \leq \theta \leq \pi$. Now since

$$T'(\theta) = 2e^{\cos\theta}(\sin(\theta - \sin\theta) - \sin\theta)/\beta^2 = 0$$

if and only if

$$\cos\left(\theta - \frac{\sin\theta}{2}\right)\sin\left(\frac{\sin\theta}{2}\right) = 0$$

if and only if $\theta \in \{0, \theta_0, \pi\}$, where $\theta_0 \approx 2.02098$ is the only root of the equation

$$2\theta - \sin\theta - \pi = 0.$$

Thus, we see that $T'(\theta) < 0$ over $(0, \theta_0)$ and $T'(\theta) > 0$ over (θ_0, π), which implies that $T(\theta)$ decreases over $[0, \theta_0]$ and increases over $[\theta_0, \pi]$ and therefore,

$$\max_{0\leq\theta\leq\pi} T(\theta) = \max\{T(0), T(\pi)\} = T(0) \leq 1$$

and (5.2) holds. Hence, $\beta \geq \beta_2$ is also sufficient for $q_\beta(z) \prec e^z$. Rest of the results follow in a similar fashion. □

Theorem 5.3. *Let $0 < \alpha < 1$, $0 < B < A < 1$ and $k = 1 + \sqrt{2}$. If $p \in \mathcal{A}_0$ and satisfies the differential subordination*

$$1 + \beta\frac{zp'(z)}{p^2(z)} \prec \wp(z),$$

then

(i) $p(z) \prec e^z$ *for* $\beta \geq \frac{e-1}{1-1/e}$.

(ii) $p(z) \prec \phi_0(z)$ *for* $\beta \geq (e-1)/\left(1 - \frac{1}{1+\frac{1}{k}\left(\frac{k+1}{k-1}\right)}\right)$.

(iii) $p(z) \prec \sqrt{1+z}$ *for* $\beta \geq \frac{(e-1)(k-1)}{k-2}$.

(iv) $p(z) \prec \frac{1+Az}{1+Bz}$ *for* $\beta \geq \max\left\{\frac{(1-1/e)(1-A)}{A-B}, \frac{(e-1)(1+A)}{A-B}\right\}$.

(v) $p(z) \prec 1 + \sin z$ *for* $\beta \geq \frac{(e-1)(1+\sin 1)}{\sin 1}$.

(vi) $p(z) \prec z + \sqrt{1+z^2}$ *for* $\beta \geq \frac{(e-1)k}{k-1}$.

The bounds are sharp.

Proof. We set $\theta(w) = 1$ and $\phi(w) = \beta/w^2 \neq 0$. Let $\mho_\beta(z, p(z)) := 1 + \beta\frac{zp'(z)}{p^2(z)}$.

(i) The differential equation $\mho_\beta(z, p(z)) = \wp(z)$ has an analytic solution given by

$$q_\beta(z) = \frac{1}{1 - \frac{1}{\beta}(e^z - 1)}.$$

Since the function $Q(z) = zq'_\beta(z)\phi(q_\beta(z)) = \wp(z) - 1 = ze^z \in \mathcal{S}^*$ and the function $h(z) = \theta(q_\beta(z)) + Q(z)$ satisfies

$$\Re\left(\frac{zh'(z)}{Q(z)}\right) = \Re\left(\frac{zQ'(z)}{Q(z)}\right) > 0.$$

Therefore, by Lemma 5.1, the following differential subordination implication holds:

$$\Psi_\beta(z, p(z)) \prec 1 + \beta z q'_\beta(z) \Rightarrow p \prec q_\beta.$$

Now to show that $p \prec e^z$, we need to show that $q_\beta \prec e^z$ and for $q_\beta \prec e^z$, it is necessary that $e^{-1} \leq q_\beta(-1)$ and $q_\beta(1) \leq e$ which gives $\beta \geq \beta_1$ and $\beta \geq \beta_2$ respectively, where

$$\beta_1 = \frac{1-1/e}{e-1} \quad \text{and} \quad \beta_2 = \frac{e-1}{1-1/e}.$$

Let $\beta \geq \max\{\beta_1, \beta_2\} = \beta_2$. Note that e^z and e^{-z} both maps $\mathbb{D}$ onto $\Omega := \{w \in \mathbb{C} : w = e^z \text{ and } z \in \mathbb{D}\}$. Therefore, to prove that $q_\beta \prec e^z$, or equivalently $q_\beta(z) \in \Omega$, it suffices to show that $q_\beta \prec e^{-z}$, which is equivalent to show that

$$1 - \frac{1}{\beta}(e^z - 1) \prec e^z. \tag{5.3}$$

Now since the difference of the square of the distances from the point $(1, 0)$ to the points on the boundary curves $e^{e^{i\theta}}$ and $1/q_\beta(e^{i\theta})$, respectively is given by

$$T(\theta) := \left(1 - \frac{1}{\beta^2}\right) S(\theta),$$

where $S(\theta) = 1 + e^{2\cos\theta} - 2e^{\cos\theta}\cos(\sin\theta)$. Here, we may consider the interval $0 \leq \theta \leq \pi$, since $T(\theta) = T(-\theta)$. A calculation shows that $S'(\theta) \leq 0$ for $\theta \in [0, \pi]$. Therefore, $T(\theta)$ decreases on $[0, \pi]$ and also

$$\min_{0\leq\theta\leq\pi} T(\theta) = \left(1 - \frac{1}{\beta^2}\right) S(\pi) = \left(1 - \frac{1}{\beta^2}\right)(1 + e^{-2} - 2e^{-1}) > 0,$$

which means (5.3) is true. Hence $\beta \geq \beta_2$ is also sufficient for $q_\beta \prec e^z$. Proof of the rest of the results are much akin and so are omitted here. □

Remark 5.1. If $p \in \mathcal{A}_0$, $\beta > 0$ and satisfies the differential subordination

$$1 + \beta\frac{zp'(z)}{p^n(z)} \prec \wp(z), \quad n \geq 2$$

then

$$p(z) \prec q_\beta(z) := \left(\frac{1}{1 - \frac{n-1}{\beta}(e^z - 1)}\right)^{\frac{1}{n-1}},$$

where q_β is the best dominant.

Theorem 5.4. *Let p be an analytic function in $\mathbb{D}$ with $p(0) = 1$. Let $0 < \alpha < 1$, $0 < B < A < 1$, $k = 1 + \sqrt{2}$ and $\wp(z) = 1 + ze^z$. Set*

$$\Psi_\beta(z, p(z)) = 1 + \beta z p'(z).$$

If $p \in \mathcal{A}_0$ and satisfies any of the following differential subordinations:

- *(i)* $\Psi_\beta(z, p(z)) \prec \phi_0(z)$ *for* $\beta \geq e(-1 + 2k\log(1 + 1/k))/k \approx 0.75822$.
- *(ii)* $\Psi_\beta(z, p(z)) \prec \sqrt{1+z}$ *for* $\beta \geq 2e(1 - \log 2) \approx 1.66822$.
- *(iii)* $\Psi_\beta(z, p(z)) \prec \varphi_\alpha(z) := 1 + \frac{z}{1-\alpha z^2}$ *for* $\beta \geq e\log(\frac{1+\sqrt{\alpha}}{1-\sqrt{\alpha}})/(2\sqrt{\alpha})$.
- *(iv)* $\Psi_\beta(z, p(z)) \prec \frac{1+Az}{1+Bz}$ *for* $\beta \geq \max\{\frac{e(A-B)}{B}\log(1 - B), \frac{A-B}{eB}\log(1 + B)\}$.
- *(v)* $\Psi_\beta(z, p(z)) \prec 1 + \sin z$ *for* $\beta \geq e\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!(2n+1)} \approx 2.57172$.
- *(vi)* $\Psi_\beta(z, p(z)) \prec z + \sqrt{1+z^2}$ *for* $\beta \geq e(3 - k + \log(k/2)) \approx 2.10399$.
- *(vii)* $\Psi_\beta(z, p(z)) \prec e^z$ *for* $\beta \geq e\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{(n!)(n)} \approx 2.16538$.

Then $p \prec \wp$ and the bounds are sharp.

Proof. We set $\theta(w) = 1$ and $\phi(w) = \beta \neq 0$. To prove our results, we make use of the inequality:

$$\wp(-1) \leq q_\beta(-1) \leq q_\beta(1) \leq \wp(1), \tag{5.4}$$

if $q_\beta \prec \wp$ which gives necessary as well as sufficient condition for the required subordination.

(i) The differential equation $\Psi_\beta(z, p(z)) = \phi_0(z)$ has an analytic solution given by

$$q_\beta(z) = 1 - \frac{1}{\beta k}\left(z + 2k \log\left(1 - \frac{z}{k}\right)\right).$$

Since the function

$$Q(z) = zq'_\beta(z)\phi(q_\beta(z)) = \phi_0(z) - 1 = \frac{z}{k}\left(\frac{k+z}{k-z}\right) \in \mathcal{S}^*$$

and the function $h(z) = \theta(q_\beta(z)) + Q(z)$ satisfies

$$\Re\left(\frac{zh'(z)}{Q(z)}\right) = \Re\left(\frac{zQ'(z)}{Q(z)}\right) > 0.$$

Therefore, by Lemma 5.1, the following differential subordination implication holds:

$$\Psi_\beta(z, p(z)) \prec 1 + \beta z q'_\beta(z) \Rightarrow p \prec q_\beta.$$

Now to show that $p \prec \wp$, we need to show that $q_\beta \prec \wp$. Now if $q_\beta \prec \wp$, then (5.4) must hold, that is, the inequalities $\wp(-1) \leq q_\beta(-1)$ and $q_\beta(1) \leq \wp(1)$ respectively gives $\beta \geq \beta_1$ and $\beta \geq \beta_2$, where

$$\beta_1 = \frac{e}{k}\left(2k \log\left(\frac{k+1}{k}\right) - 1\right) \quad \text{and} \quad \beta_2 = \frac{1}{ek}\left(2k \log\left(\frac{k}{k-1}\right) + 1\right).$$

Let $\beta \geq \max\{\beta_1, \beta_2\} = \beta_1$. Note that $q_\beta(\mathbb{D})$ is convex, symmetric about real axis and $q_\beta(\mathbb{D}) \subset q_{\beta_1}(\mathbb{D})$ for each $\beta \geq \beta_1$. Also, $|\arg(1+z)| \leq \arctan(r/(1-r))$ for $|z| = r < 1$. Now to prove $q_\beta \prec \wp$, in view of Lemma 2.1, it only suffices to show that $\max_{0\leq\theta\leq\pi} |\arg q_{\beta_1}(e^{i\theta})| = \max_{0\leq\theta\leq\pi} \left|\arg \wp(e^{i\theta})\right|$. Since

$$\begin{aligned}
&\max_{0\leq\theta\leq\pi} |\arg q_{\beta_1}(e^{i\theta})| \\
&= \max_{0\leq\theta\leq\pi} \left|\arctan(\Im q_{\beta_1}(e^{i\theta})/\Re q_{\beta_1}(e^{i\theta}))\right| \\
&= \max_{0\leq\theta\leq\pi} \left|\arctan\left(\frac{(-2k \arg\left(1 - \frac{e^{i\theta}}{k}\right) - \sin\theta)}{M - \cos\theta - k\log((1 + \frac{1}{k^2} - \frac{2\cos\theta}{k})}\right)\right| \\
&\leq \arctan\left(\frac{2k \arctan(\frac{1}{k-1}) + 1}{M(1 - 1/e)}\right) \approx 0.7719 \\
&< \max_{0\leq\theta\leq\pi} |\arg(1 + ze^z)| \approx 1.41022,
\end{aligned}$$

where $M = e(-1 + 2k\log(1 + 1/k))$. Therefore, $\beta \geq \beta_1$ is also sufficient for $q_\beta \prec \wp$ to hold. Hence the result. Other results follow similarly. □

Let $f \in \mathcal{A}$ and setting $p(z) = zf'(z)/f(z)$ and $p(z) = f(z)/z$. Then using Theorem 5.4, we obtain sufficient conditions for f to be in $\mathcal{S}^*_\wp$.

Corollary 5.5. *Let $f \in \mathcal{A}$ and set*

$$\Psi_\beta\left(z, \frac{zf'(z)}{f(z)}\right) = 1 + \beta\frac{zf'(z)}{f(z)}\left(1 - \frac{zf'(z)}{f(z)} + \frac{zf''(z)}{f'(z)}\right).$$

If $\Psi_\beta\left(z, \frac{zf'(z)}{f(z)}\right)$ satisfies any of the conditions (i)-(vii) of the Theorem 5.4, then $f \in \mathcal{S}^*_\wp$.

Corollary 5.6. *Let $f \in \mathcal{A}$ and set*

$$\Psi_\beta\left(z, \frac{f(z)}{z}\right) = 1 + \beta\left(f'(z) - \frac{f(z)}{z}\right).$$

If $\Psi_\beta\left(z, \frac{f(z)}{z}\right)$ satisfies any of the conditions (i)-(vii) of the Theorem 5.4, then $f \in \mathcal{S}^*_\wp$.

Theorem 5.7. *Let $0 < \alpha < 1$, $0 < B < A < 1$, $\gamma = \log(e/(e-1))$, $k = 1 + \sqrt{2}$ and $\wp(z) = 1 + ze^z$ and $p \in \mathcal{A}_0$. Set*

$$\Phi_\beta(z, p(z)) = 1 + \beta\frac{zp'(z)}{p(z)}.$$

If p satisfies any of the following differential subordinations:

- *(i)* $\Phi_\beta(z, p(z)) \prec \phi_0(z)$ *for* $\beta \geq (-1 + 2k\log(1 + 1/k))/\gamma k \approx 0.60812$.
- *(ii)* $\Phi_\beta(z, p(z)) \prec \sqrt{1+z}$ *for* $\beta \geq 2(1 - \log 2)/\gamma \approx 1.33799$.
- *(iii)* $\Phi_\beta(z, p(z)) \prec \varphi_\alpha(z)$ *for* $\beta \geq \log(\frac{1+\sqrt{\alpha}}{1-\sqrt{\alpha}})/(2\gamma\sqrt{\alpha})$.
- *(iv)* $\Phi_\beta(z, p(z)) \prec (1+Az)/(1+Bz)$ *for* $\beta \geq \max\left\{\frac{(A-B)\log(1/(1-B))}{\gamma B}, \frac{(A-B)\log(1+B)}{B\log(1+e)}\right\}$.
- *(v)* $\Phi_\beta(z, p(z)) \prec 1 + \sin z$ *for* $\beta \geq (1/\gamma)\sum_{n=0}^{\infty} \frac{(-1)^(3n+2)}{(2n+1)!(2n+1)} \approx 2.06264$.
- *(vi)* $\Phi_\beta(z, p(z)) \prec z + \sqrt{1+z^2}$ *for* $\beta \geq (3 - k + \log(k/2))/\gamma \approx 1.6875$.
- *(vii)* $\Phi_\beta(z, p(z)) \prec e^z$ *for* $\beta \geq (1/\gamma)\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{(n!)(n)} \approx 1.736740$.

Then $p \prec \wp$ and all the above bounds are sharp.

Proof. Let $\theta(w) = 1$ and $\phi(w) = \beta/\omega \neq 0$. To prove our results, we make use of the following inequality:

$$\wp(-1) \leq q_\beta(-1) \leq q_\beta(1) \leq \wp(1) \tag{5.5}$$

so that $q_\beta \prec \wp$. Observe that (5.5) gives necessary as well as sufficient condition for the required subordination.

(*i*) The differential equation $\Phi_\beta(z, p(z)) = \phi_0(z)$ has an analytic solution given by

$$q_\beta(z) = \exp\left(-(z + 2k\log(1 - z/k))/\beta k\right). \tag{5.6}$$

Since the function

$$Q(z) = zq'_\beta(z)\phi(q_\beta(z)) = \phi_0(z) - 1 = \frac{z}{k}\left(\frac{k+z}{k-z}\right) \in \mathcal{S}^*$$

and $h(z) = \theta(q_\beta(z)) + Q(z)$ satisfies $\Re(zh'(z)/Q(z)) = \Re(zQ'(z)/Q(z)) > 0$. Therefore, by Lemma 5.1, the following differential subordination implication holds:

$$\Phi_\beta(z, p(z)) \prec 1 + \beta\frac{zq'_\beta(z)}{q_\beta(z)} \Rightarrow p \prec q_\beta.$$

Now to show that $p \prec \wp$, we need to show that $q_\beta \prec \wp$. Thus, (5.5) gives necessary condtion for the subordination $p \prec \wp$ to hold. That is, $\beta \geq \beta_1$ and $\beta \geq \beta_2$, respectively, where

$$\beta_1 = \frac{-1 + 2k\log(1 + 1/k)}{k\log(e/(e-1))} \quad \text{and} \quad \beta_2 = \frac{-1 + 2k\log(k/(k-1))}{k\log(1+e)}.$$

Let $\beta \geq \beta_1$. Also for simplicity, we write the function given in (5.6) as follows for $z = e^{i\theta}$:

$$q_\beta(z) := \exp(u + iv),$$

where u and v are function of $\theta \in [-\pi, \pi]$ given as

$$u(\theta) := \left(\frac{-1}{\beta k}\right)\left(\cos\theta + 2k\log\sqrt{1 + \frac{1}{k^2} - \frac{2\cos\theta}{k}}\right)$$

and

$$v(\theta) := \left(\frac{-1}{\beta k}\right)\left(\sin\theta + 2k\arctan\left(\frac{\sin\theta}{\cos\theta - k}\right)\right).$$

Further, without loss of generality, we may assume that $\beta = \beta_1$. Now the difference of the square of the distances from the point $(1, 0)$ to the points on the boundary curves $\wp(e^{i\theta})$ and $q_\beta(e^{i\theta})$, respectively is given by

$$d(\theta) := e^{2\cos\theta} - 1 - e^{2u(\theta)} + 2e^{u(\theta)}\cos(v(\theta)).$$

Since $d(\theta) = d(-\theta)$. So we may consider the interval $0 \leq \theta \leq \pi$. Now we need to prove that $d(\theta) \geq 0$ and so we consider

$$d'(\theta) = -e^{2\cos\theta}\sin\theta - 2e^{u(\theta)}(e^{u(\theta)}u'(\theta) + u'(\theta)\cos(v(\theta)) - \sin(v(\theta))v'(\theta)).$$

It is easy to see that $d'(\theta) = 0$ for $\theta = 0$ and π. Further a numerical computation also shows that $d'(\theta) = 0$ for a unique $\theta_0 \in (0, \pi)$ such that $d(\theta)$ increases on $[0, \theta_0]$ and decreases on $[\theta_0, \pi]$, where $\theta_0 \approx 0.351807$. Note that $d(0) > 0$ and $d(\pi) = 0$. Therefore,

$$\min_{0 \leq \theta \leq \pi} d(\theta) = 0.$$

Hence $\beta \geq \beta_1$ is also sufficient for $q_\beta \prec \wp$ and this completes the proof. Similarly we can prove the other results. □

Let $f \in \mathcal{A}$ and set $p(z) = zf'(z)/f(z)$ and $p(z) = f(z)/z$. Then using Theorem 5.7, we obtain sufficient conditions for f to be in $\mathcal{S}^*_\wp$.

Corollary 5.8. *Let $f \in \mathcal{A}$ and set*

$$\Phi_\beta\left(z, \frac{zf'(z)}{f(z)}\right) = 1 + \beta\left(1 - \frac{zf'(z)}{f(z)} + \frac{zf''(z)}{f(z)}\right).$$

If $\Phi_\beta\left(z, \frac{zf'(z)}{f(z)}\right)$ satisfies any of the conditions (i)-(vii) of Theorem 5.7, then $f \in \mathcal{S}^*_\wp$.

Corollary 5.9. *Let $f \in \mathcal{A}$ and set*

$$\Phi_\beta\left(z, \frac{f(z)}{z}\right) = \beta\frac{zf'(z)}{f(z)} + (1 - \beta).$$

If $\Phi_\beta\left(z, \frac{f(z)}{z}\right)$ satisfies any of the conditions (i)-(vii) of Theorem 5.7, then $f \in \mathcal{S}^*_\wp$.

The proof of the following result is omitted here as it is much akin to the proof of Theorem 5.4 and 5.7.

Theorem 5.10. *Assume that* $0 < \alpha < 1$, $0 < B < A < 1$, $k = 1 + \sqrt{2}$ *and* $\wp(z) = 1 + ze^z$. *Let* $p \in \mathcal{A}_0$. *Set*

$$\mho_\beta(z, p(z)) = 1 + \beta \frac{zp'(z)}{p^2(z)}.$$

If p *satisfies any of the following differential subordinations:*

(*i*) $\mho_\beta(z, p(z)) \prec \phi_0(z)$ *for* $\beta \geq -((e+1)/ke)(1 + 2k\log(1 - 1/k)) \approx 0.896489$.
(*ii*) $\mho_\beta(z, p(z)) \prec \sqrt{1+z}$ *for* $\beta \geq 2(e-1)(1 - \log 2) \approx 1.05451$.
(*iii*) $\mho_\beta(z, p(z)) \prec \varphi_\alpha(z)$ *for* $\beta \geq ((e-1)/2\sqrt{\alpha}) \log(\frac{1+\sqrt{\alpha}}{1-\sqrt{\alpha}})$.
(*iv*) $\mho_\beta(z, p(z)) \prec 1 + \sin z$ *for* $\beta \geq (e-1)\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!(2n+1)} \approx 1.62563$.
(*v*) $\mho_\beta(z, p(z)) \prec e^z$ *for* $\beta \geq ((e+1)/e)\sum_{n=1}^{\infty} \frac{1}{n!n} \approx 1.80273$.

Then $p \prec \wp$ *and the bounds are sharp.*

Let $f \in \mathcal{A}$ and set $p(z) = zf'(z)/f(z)$ and $p(z) = f(z)/z$. Then using Theorem 5.10, we obtain the following results:

Corollary 5.11. *Let* $f \in \mathcal{A}$ *and set*

$$\mho_\beta\left(z, \frac{zf'(z)}{f(z)}\right) = 1 + \beta\left(\frac{zf'(z)}{f(z)}\right)^{-1}\left(1 - \frac{zf'(z)}{f(z)} + \frac{zf''(z)}{f(z)}\right).$$

If $\mho_\beta\left(z, \frac{zf'(z)}{f(z)}\right)$ *satisfies any of the conditions (i)-(v) of the Theorem 5.10, then* $f \in \mathcal{S}^*_\wp$.

Corollary 5.12. *Let* $f \in \mathcal{A}$ *and set*

$$\mho_\beta\left(z, \frac{f(z)}{z}\right) = 1 + \beta\left(\frac{z}{f(z)}\left(\frac{zf'(z)}{f(z)} - 1\right)\right).$$

If $\mho_\beta\left(z, \frac{f(z)}{z}\right)$ *satisfies any of the conditions (i)-(v) of the Theorem 5.10, then* $f \in \mathcal{S}^*_\wp$.

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] Aktaş İ, Baricz Á and Orhan H. Bounds for radii of starlikeness and convexity of some special functions. Turkish J Math, 2018, **42**(1): 211–226

[2] Ali R M, Ravichandran V and Jain N K. Convolution of certain analytic functions. J Anal, 2010, **18**: 1–8

[3] Bansal D and Prajapat J K. Certain geometric properties of the Mittag-Leffler functions. Complex Variables and Elliptic Equations, 2016, **61**(3): 338–350

[4] Banga S and Kumar S S. Applications of differential subordinations to certain classes of starlike functions. J Korean Math Soc, 2020, **57**(2): 331–357

[5] Baricz Á, Dimitrov D K, Orhan H, and Yağmur N. Radii of starlikeness of some special functions. Proc Amer Math Soc, 2016, **144**(8): 3355–3367

[6] Baricz Á, Toklu E and Kadioğlu E. Radii of starlikeness and convexity of Wright functions. Math Commun, 2018, **23**(1): 97–117

[7] Biernacki M and Krzyż J. On the monotonity of certain functionals in the theory of analytic functions. Ann Univ Mariae Curie-Skłodowska Sect A,1957, **9**(1955): 135–147

[8] Bulut S and Engel O. The radius of starlikeness, convexity and uniform convexity of the Legendre polynomials of odd degree. Results Math, 2019, **74**(1), Art 48, 9 pp.

[9] Cho N E, Kumar S, Kumar v and Ravichandran V. Convolution and radius problems of analytic functions associated with the tilted Carathéodory functions. Math Commun, 2019, **24**(2): 165–179

[10] Cho N E, Kumar V, Kumar S S and Ravichandran V. Radius problems for starlike functions associated with the sine function. Bull Iranian Math Soc, 2019, **45**(1): 213–232

[11] Gangadharan A, Ravichandran V and Shanmugam T N. Radii of convexity and strong starlikeness for some classes of analytic functions. J Math Anal Appl, 1997, **211**(1): 301–313

[12] Goel P and Kumar S S. Certain Class of Starlike Functions Associated with Modified Sigmoid Function. Bull Malays Math Sci Soc, 2020, **43**(1): 957–991

[13] Goluzin G M. On a variational method in the theory of analytic functions. Amer Math Soc Transl, 1961, **18**(2): 1–14

[14] Kargar R, Ebadian A and Sokół J. Radius problems for some subclasses of analytic functions. Complex Anal Oper Theory, 2017, **11**(7): 1639–1649

[15] Kumar V, Cho N E, Ravichandran V and Srivastava H M. Sharp coefficient bounds for starlike functions associated with the Bell numbers. Math Slovaca, 2019, **69**(5): 1053–1064

[16] Kumar S S and Kamaljeet G. A Cardioid Domain and Starlike Functions. Analysis and Mathematical Physics, 2020 (Communicated)

[17] Kumar S and Ravichandran V. A subclass of starlike functions associated with a rational function. Southeast Asian Bull Math, 2016, **40**(2): 199–212

[18] Kuroki K, Owa S. Notes on new class for certain analytic functions. RIMS Kokyuroku, 2011, **1772**: 21–25

[19] Ma W C and Minda D. A unified treatment of some special classes of univalent functions. *Proceedings of the Conference on Complex Analysis, Tianjin*, Conf Proc Lecture Notes Anal, I Int Press, Cambridge, MA, 1992: 157–169

[20] MacGregor T H. Majorization by univalent functions. Duke Math J, 1967, **34**: 95–102

[21] MacGregor T H. Hull subordination and extremal problems for starlike and spirallike mappings. Trans Amer Math Soc, 1973, **183**: 499–510

[22] Mendiratta R, Nagpal S and Ravichandran V. On a subclass of strongly starlike functions associated with exponential function. Bull Malays Math Sci Soc, 2015, **38**(1): 365–386

[23] Miller S S and Mocanu P T. *Differential subordinations*, Monographs and Textbooks in Pure and Applied Mathematics. New York: Marcel Dekker, Inc, 2000

[24] Raina R K and Sokół J. Some properties related to a certain class of starlike functions. C R Math Acad Sci Paris, 2015, **353**(11): 973–978

[25] Ruscheweyh S and Sheil-Small T. Hadamard products of Schlicht functions and the Pólya-Schoenberg conjecture. Comment Math Helv, 1973, **48**: 119–135

[26] Sharma K, Jain N K and Ravichandran V. Starlike functions associated with a cardioid. Afr Mat, 2016, **27**(5-6): 923–939

[27] Sokół J and Stankiewicz J. Radius of convexity of some subclasses of strongly starlike functions. Zeszyty Nauk Politech Rzeszowskiej Mat, 1996, (19): 101–105

[28] Tang H and Deng G. Majorization problems for some subclasses of starlike functions. J Math Res Appl, 2019, **39**(2): 153–159.

[29] Wang L M. The tilted Carathéodory class and its applications. J Korean Math Soc, 2012, **49**: 671–686

Department of Applied Mathematics, Delhi Technological University, Delhi–110042, India

E-mail address: spkumar@dce.ac.in

Department of Applied Mathematics, Delhi Technological University, Delhi–110042, India

E-mail address: gangania.m1991@gmail.com

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3007079, IEEE Transactions on Industrial Electronics

Suppression of Lower Order Harmonics for the full modulation range for a 2-level inverter fed IM drive with a switched capacitive filter technique forming a 42-sided Voltage Space Vector Structure

Rahul Dewani, *Student Member, IEEE*, Rakesh R, *Student Member, IEEE*,
K. Gopakumar, *Fellow, IEEE*, Umanand Loganathan
Dariusz Zielinski, *Member, IEEE*, Leopoldo G. Franquelo, *Fellow, IEEE*

***Abstract*—The work proposes an inverter topology for the first time, for suppression of all lower order harmonics (5^{th}, 7^{th}, 11^{th}, 13^{th}, ...) up to 39^{th} harmonic. The lower order harmonic suppression is achieved by switching the 42-sided polygonal vectors, which eliminates harmonics in phase voltage without using passive LC filtering technique. The power circuit topology uses a single DC link supply for primary inverter. The secondary inverter is fed with a capacitive supply. A novel topology is proposed for the dense 6-level secondary inverter. The lower order harmonics in phase voltage are suppressed using secondary inverter, which acts like a switched-capacitive filter. The work also proposes a novel capacitor voltage balancing scheme for the first time, which achieves voltage balancing for secondary inverter capacitors. The voltages of secondary inverter capacitors are balanced using the power balance relationship between primary inverter and motor load. An experimental study of the topology and capacitor balancing scheme is presented for the validation of proposed scheme.**

***Index Terms*—Harmonic suppression, Polygonal space vector structure, Induction Motor drives, Transformer-less Filtering, Lower Order Harmonic Elimination, Pulse width modulation (PWM).**

I. Introduction

INDUCTION MOTORS have emerged as the preferred solution for medium to high power industrial applications. The induction motor can be driven with a 3-phase, 2-level inverter, fed with active DC link supply for variable speed applications. The 2-level inverter generates a hexagonal Space Vector (SV) structure which has lower linear modulation range of operation. When the inverter is driven in over-modulation range, lower order harmonics in phase voltage, cause undesirable lower order harmonic torque pulsations. The lower order torque ripple is not suitable for industrial applications, and can even lead to mechanical breakdown of motor shaft [1], [2].

The lower order torque harmonics can be eliminated by suppressing lower order harmonics in phase voltage. The lower order harmonics can be attenuated by using a passive LC filter [3], [4]. Since the phase voltage has 5^{th} and 7^{th} order harmonics in over-modulation, the cut-off frequency of LC filter, must be lower than $5.\omega_o$ (ω_o is the fundamental frequency). An LC filter of such a low cut-off frequency will be very bulky and expensive. The LC filter also limits the bandwidth of the speed control loop, degrading transient performance. Also, in linear modulation range the 2-level inverter needs to be switched at high frequencies to suppress lower order harmonics resulting in higher switching losses. Hence, this solution for harmonic suppression has many drawbacks in implementation.

Another technique used for suppression of lower order harmonics in phase voltage is Selective Harmonic Elimination(SHE) [5]–[8]. The technique is based on mathematical computations to find the optimum PWM switching sequence for the inverter operation. The resulting phase voltage will be devoid of the harmonics for which the PWM scheme has been designed. The technique lacks intuitive insight, as it involves solving an additional non-linear equation for each harmonic being eliminated. Hence, the non-linear equations are precomputed and optimal PWM sequence is stored in look-up tables. The method is limited in the number of lower order harmonics that can be suppressed, as the number of non-linear equations to be solved keeps increasing with the number of harmonics that need to be suppressed.

The multi-level topologies proposed in literature [9]–[12], achieve harmonic suppression in linear range of operation. Since the closest levels are time-averaged to generate the reference pole voltage, the switching ripple is reduced in the phase voltage. While this method is useful for harmonic suppression in linear range of operation, in general, multi-level topologies generate lower order harmonics in over-modulation range of operation. This imposes a limitation on the performance that can be achieved with multi-level inverters.

The topology proposed in this work overcomes the draw-

Manuscript received March 14, 2020; revised May 24, 2020; accepted June 21, 2020.

Rahul Dewani, Rakesh R, K. Gopakumar and Umanand Loganathan are with the Department of Electronic Systems Engineering,Indian Institute of Science,Bangalore-560012,India.(email: rahuldewani@iisc.ac.in; rrakesh225@gmail.com; kgopa@iisc.ac.in; lums@iisc.ac.in).

Dariusz Zielinski is with the Department of Electrical Drives and Machines, Lublin University of Technology, Lublin, Poland. (email: d.zielinski@pollub.pl).

Leopoldo G. Franquelo is with the Electronic Engineering Department, University of Seville, 41092, Spain. (email: lgfranquelo@ieee.org).

0278-0046 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 06:36:29 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3007079, IEEE Transactions on Industrial Electronics

backs of other works in literature, by rejecting lower order harmonics in phase voltage without compromising on the bandwidth of speed control loop. The harmonic rejection scheme is extended over the entire range of linear modulation up to the base speed, without the need of complex computations. The power circuit topology proposed in literature [13], for polygonal structures have reported the necessity of multiple DC link sources. In the work presented in this paper, a harmonic elimination scheme with a single DC link supply (feeding primary inverter) is proposed. A secondary inverter fed with capacitive supply is used to suppress lower order harmonics in phase voltage.

The primary and secondary inverters are feeding an Open End IM (OEIM) from opposite ends in this scheme. The primary inverter delivers the active power for motor operation. The secondary inverter delivers only reactive power for harmonic elimination, hence a capacitive supply can be used for secondary inverter [14], [15]. A new power circuit topology for secondary inverter is proposed in this work for the first time, which generates a 6-level Space Vector Structure (SVS). The primary and secondary space vectors are superimposed to generate a 42-sided polygonal SVS.

The secondary inverter capacitor voltages are balanced using a novel capacitor voltage balancing scheme for the first time, which does not compromise on suppression of lower order harmonics. The works proposed in literature for balancing of secondary inverter [15], [16], achieve capacitor balancing by modifying the timings of the vectors being switched, by a factor referred as K_{ratio} correction factor in literature. This brings about an error in the reference vector being generated for motor operation, which can cause lower order harmonics in phase voltage. This work overcomes the drawbacks for the first time, by introducing a novel capacitor voltage balancing scheme. The capacitor balancing is achieved by controlling the power balance of primary inverter and motor load. The capacitor voltages are controlled by modifying the radius of the 42-sided polygon, by small percentage, so as to achieve power balance.

II. Inverter Topology

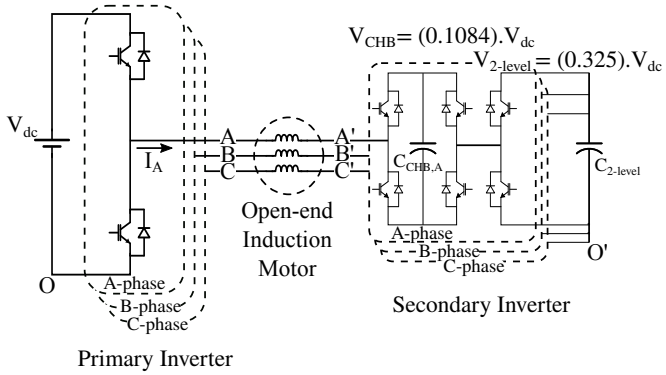


Fig. 1. Proposed Power Circuit Topology

The power circuit topology proposed in this work, is shown in Fig.1. The primary 2-level inverter fed by active DC link supply, provides the active power for motor operation. The secondary inverter, feeding the other end of OEIM is used for switched-capacitive filtering of lower order harmonics in motor phase voltage. The secondary inverter consists of a 2-level inverter fed with capacitive supply, followed by Cascaded H-Bridge (CHB) stages in the three phases.

A dense 6-level SV structure with reduced component count is generated by using asymmetric voltages of $3xV_{dc}$ and xV_{dc} for 2-level and CHB stages, where x is a constant ratio. The secondary 2-level stage generates a 2-level SV structure of radius $3xV_{dc}$, and the CHB stage generates a 3-level SV structure of radius $2xV_{dc}$. The combination of the 2-level and 3-level SV structures results in a dense 6-level SV structure of radius $5xV_{dc}$ for secondary inverter. The secondary structure is superimposed on the primary SV structure to generate the resultant 42-sided vectors, as shown in Fig.2. The figure also shows the 42-sided polygonal vectors (OP_1, OP_2, ...). The 42-sided vectors are generated by superposition of primary and secondary vectors. For example, the 42-sided vector OP_4 is generated by switching $V_{pri,1}$ vector from primary inverter and $O'P_4$ from secondary inverter. The vector $O'P_4(= 0.4374V_{dc}\angle 108.5^o)$ can be generated from the secondary SV structure, if the vector $O'P_4$ is inside the secondary 6-level SV structure of radius $5xV_{dc}$. The ratio x is designed to be 0.1084 so that $O'P_4$, can be generated by the secondary SV structure. The 2-level and CHB capacitors are maintained at voltages of $0.325V_{dc}$ and $0.1084V_{dc}$ for 42-sided polygonal switching.

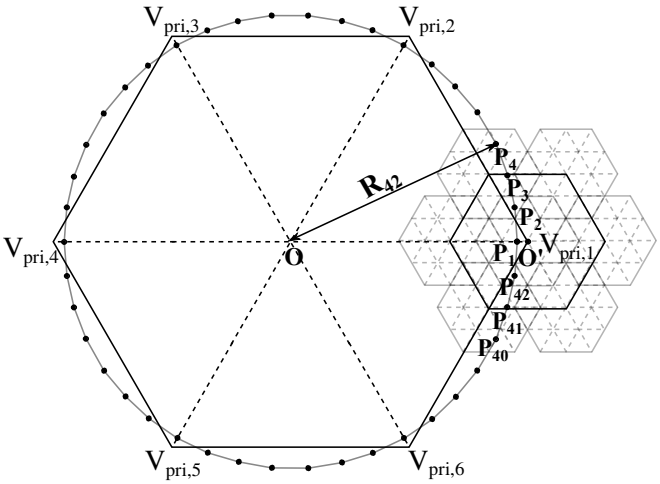


Fig. 2. SV structure of secondary inverter, superimposed on primary vector $\mathrm{V_{pri,1}}$, along with 42-sided vectors.

For power balance of secondary inverter capacitors, the power drawn form primary inverter should be same as the power delivered to the motor load. The power balance is maintained, if the fundamental phase voltage generated by the primary inverter, is equal to the fundamental phase voltage generated by switching the vectors of the 42-sided polygon. Using this condition, at the extreme case of over-modulation (where primary 2-level inverter is in six step mode of operation), the radius of 42-sided polygon can be determined. The fundamental component of phase voltage in six step mode of operation, $V_{pri,fund}$ for a hexagonal SV structure of radius V_{dc} is given by,

$$V_{pri,fund} = 0.6366V_{dc} \tag{1}$$

0278-0046 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 06:36:29 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3007079, IEEE Transactions on Industrial Electronics

Similarly, for a 42-sided polygon of radius R_{42} operating in 42-stepped mode of operation, $V^*_{pri,fund}$, the fundamental component of phase voltage, can be determined to be,

$$V^*_{pri,fund} = 0.6660R_{42} \tag{2}$$

From the power balance condition, the fundamental phase voltage should be same in the two cases,

$$\begin{aligned} &V^*_{pri,fund} = V_{pri,fund} \\ \implies &R_{42} = 0.956V_{dc} \end{aligned} \tag{3}$$

Where, R_{42} is the nominal radius of the 42-sided polygonal vectors that is necessary to maintain the active power balance.

III. Capacitor Balancing strategy

Based on the explanation in the previous section, the radius of the 42-sided polygonal reference vectors is maintained at $0.956V_{dc}$ (Equation-(3)). The PWM switching pattern for primary and secondary inverters can be obtained by assuming the secondary inverter capacitors are balanced to their nominal voltages, shown in Fig.1. For the 42-sided PWM scheme, the real power will flow from primary inverter to motor load. Secondary inverter only provides reactive power, therefore, it can be fed with capacitive supply maintained at a nominal voltage using the capacitor balancing scheme proposed next.

Assume that the secondary inverter CHB capacitors are maintained at the nominal voltage of $0.1084V_{dc}$, as shown in Fig.1. The secondary inverter 2-level capacitor can get discharged due to circuit non-idealities (or overcharged, overcharging is discussed later). To balance the 2-level capacitor voltage, it needs to be charged back to it's nominal voltage by sourcing active power to secondary inverter. To source power to secondary inverter, a fraction of active power drawn from primary inverter is diverted to secondary inverter as explained next. To source active power to secondary inverter, the radius of 42-sided polygonal vectors (OP_1, OP_2, OP_3,.., OP_{42}) is reduced slightly from the nominal value of $0.956V_{dc}$, by changing the PWM of the secondary inverter alone. By reducing the radius of 42-sided polygonal reference vectors, the phase voltage of motor load is reduced momentarily. As a result, the power going to motor load is reduced momentarily. During this process, the PWM pattern of primary inverter is not disturbed. Since the primary inverter PWM is not disturbed, the power drawn from primary inverter does not change. Thus, the power drawn from primary inverter is maintained constant, while the power delivered to motor load is reduced. Since, the power drawn from primary inverter is now greater than the power drawn by motor load, the excess power is diverted to secondary inverter 2-level capacitor momentarily. The power diverted to secondary inverter increases the energy stored in 2-level capacitor, by charging the 2-level capacitor voltage back to it's nominal value.

The secondary 2-level capacitor may get overcharged due to circuit non-idealities. To balance the 2-level capacitor voltage, it needs to be discharged back to it's nominal voltage by drawing active power from secondary inverter. To draw power from secondary inverter, the active power delivered to motor load is increased momentarily by increasing motor phase voltage as explained next. To draw active power from secondary inverter, the radius of 42-sided polygonal vectors is increased from the nominal value of $0.956V_{dc}$, by changing the PWM of the secondary inverter alone. By increasing the radius of 42-sided polygonal reference vectors, the phase voltage of motor load is increased momentarily. As a result, the power going to motor load is increased momentarily. During this process, the PWM pattern of primary inverter is not disturbed. Since the primary inverter PWM is not disturbed, the power drawn from primary inverter does not change. Thus, the power drawn from primary inverter is maintained constant, while the power delivered to motor load is increased. Since, the power drawn from primary inverter is now lesser than the power drawn by motor load, power is drawn from the secondary inverter 2-level capacitor momentarily. The power drawn from secondary inverter decreases the energy stored in 2-level capacitor, by discharging the 2-level capacitor voltage back to it's nominal value. Thus, the above mentioned control scheme is used to maintain the 2-level capacitor voltage within $\pm 5\%$ of $0.325V_{dc}$.

In the balancing scheme for 2-level secondary capacitor discussed so far, the CHB capacitors of secondary inverter are assumed to be balanced at their nominal voltages. In practice, they may deviate from their nominal voltages. A fast switching cycle capacitor voltage balancing scheme is proposed for CHB stage capacitors alone in sub-section.III-B.

A. Voltage Balancing of 2-level stage capacitor

Assuming that the CHB stage capacitor voltages are balanced at their nominal value, there may be a momentary disturbance in the secondary 2-level capacitor voltage due to circuit non-idealities. The 2-level capacitor needs to be balanced by diverting power appropriately to secondary inverter. The PWM of secondary inverter has to be modified in such a way that, either a fraction of real power from primary inverter is given to secondary inverter, or real power is drawn from secondary inverter and given to motor load. During this process the PWM pattern of primary inverter is not disturbed.

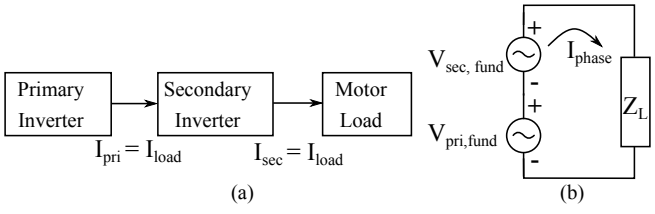


Fig. 3. **(a)** Block diagram of power circuit. **(b)** Single phase equivalent of power circuit.

The primary and secondary phase voltages are generated by PWM switching pattern of the inverters. The currents for primary and secondary inverters is same as motor phase current (Fig.3(a)). Thus, a single phase equivalent model can be visualized as shown in Fig.3(b). The primary and secondary fundamental phase voltages can be visualized as voltage sources in series. If the phase voltage of load impedance (motor load) is maintained to be at phase voltage of primary inverter, then secondary inverter does not contribute any fundamental phase voltage. If the fundamental phase voltage of secondary inverter is maintained to be zero, it does not contribute any real power.

0278-0046 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 06:36:29 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3007079, IEEE Transactions on Industrial Electronics

To balance the secondary 2-level capacitor voltage, the radius of 42-sided vectors is changed, by changing the switching pattern of secondary inverter alone. By changing the radius of 42-sided vectors from the nominal value, fundamental phase voltage of motor load is changed. Since, the PWM of primary inverter is not disturbed, the phase voltage of primary inverter remains same. If the motor phase voltage is increased, the secondary inverter generates positive phase voltage and delivers real power to motor load. Whereas, if the motor phase voltage is reduced, the secondary inverter generates negative phase voltage and absorbs a fraction of the real power given by primary inverter (refer Fig.3).

The real power drawn from primary inverter is given by,

$$P_{pri} = 3/2 * V_{pri,fund} * I_{phase} * cos(\theta) \tag{4}$$

Where, $V_{pri,fund}$ and I_{phase}, represent the amplitudes of fundamental components of phase voltage and current of the primary inverter, respectively, and $cos(\theta)$ represents the power factor. Similarly, the real power delivered to the motor can be expressed as,

$$P_{motor} = 3/2 * V_{motor,fund} * I_{phase} * cos(\theta) \tag{5}$$

Where, $V_{motor,fund}$ represents the amplitude of fundamental component of motor phase voltage. In steady state, the power drawn from the primary inverter, equals the power delivered to the motor load. During transients, if the power drawn from the primary inverter is more than the power delivered to the motor load, the excess power will be stored in the secondary inverter capacitors. This relation can be expressed as,

$$\begin{aligned} P_{pri} &= P_{motor} + P_{sec} \\ \implies P_{sec} &= P_{pri} - P_{motor} \end{aligned} \tag{6}$$

Where, P_{sec} is the real power going into secondary inverter. Substituting P_{pri} and P_{motor} from (4) and (5), in (6),

$$P_{sec} = 3/2 * I_{phase} * cos(\theta) * [V_{pri,fund} - V_{motor,fund}]$$

Note that, $cos(\theta)$ and I_{phase} are the Power Factor and amplitude of phase current, hence the two are always positive. Thus, the power realtion can be expressed as,

$$P_{sec} = K * [V_{pri,fund} - V_{motor,fund}]$$

Where, $K = 3/2 * I_{phase} * cos(\theta)$, is always positive for motoring operation. For $K > 0$, the above relation can be expressed as,

$$\begin{aligned} &if, \quad V_{pri,fund} > V_{motor,fund} \\ &then, V_{pri,fund} - V_{motor,fund} > 0 \\ &\implies P_{sec} > 0 \end{aligned} \tag{7}$$

$$\begin{aligned} &elseif, V_{pri,fund} < V_{motor,fund} \\ &then, \ V_{pri,fund} - V_{motor,fund} < 0 \\ &\implies \ P_{sec} < 0 \end{aligned} \tag{8}$$

Using the above relations, the real power of secondary inverter, P_{sec}, can be controlled by changing the fundamental component of motor phase voltage alone, without disturbing $V_{pri,fund}$. The fundamental motor phase voltage is controlled by changing the radius of 42-sided vectors.

Since the secondary inverter has a capacitive supply, when real power is supplied to secondary inverter, the capacitor voltage changes to store the real power in the capacitor. So, if the capacitor voltage deviates from it's nominal value, the fundamental motor phase voltage is momentarily changed to bring the capacitor voltage back to it's nominal value. The control scheme for voltage balancing of 2-level capacitor is summarized as follows:

1) Sense the voltage of 2-level capacitor of secondary inverter and compare the voltage with the desired nominal value. The sensed voltage may be greater (or lesser) than the nominal voltage due to momentary disturbances caused by circuit non-idealities. Initially, the primary inverter phase voltage and motor phase voltage are equal.
$$\begin{aligned} V_{motor,fund} &= V_{pri,fund} \\ \implies P_{sec} &= 0 \end{aligned}$$
2) If the sensed voltage is greater (or lesser), then active power should be drawn (or supplied) from the secondary inverter. The motor phase voltage is increased (or decreased) to $V^*_{motor,fund}$ by changing the PWM of secondary inverter alone, so that the power drawn by motor load exceeds the power supplied by primary inverter. As a result, the secondary inverter delivers active power to motor load.
$$\begin{aligned} V^*_{motor,fund} &= V_{pri,fund}.(1 + \alpha\%) \\ \implies V_{pri,fund} &< V^*_{motor,fund} \\ \implies P^*_{sec} &< 0 \end{aligned}$$
Where, α is less than 4%. Real power, P^*_{sec}, is drawn from the secondary inverter due to control action. As a result, the overcharged 2-level capacitor starts discharging.
3) When the secondary capacitor voltage is close to its nominal value, the motor phase voltage is reduced back to $V_{pri,fund}$, such that,
$$\begin{aligned} V_{motor,fund} &= V_{pri,fund} \\ \implies P_{sec} &= 0 \end{aligned}$$

B. Voltage Balancing of CHB stage capacitors

The CHB stage capacitor voltages are balanced within every switching cycle. The fast charge balancing for CHB stage capacitor reduces the capacitor size and voltage ripple. To achieve fast capacitor voltage balancing, SV redundancies for every vector in secondary inverter SV structure are identified.

TABLE I
SWITCHING STATE CODES CORRESPONDING TO SECONDARY INVERTER VOLTAGES.

Voltage	Switching State code
$+0.325 * V_{dc}$	+3
$+0.1084 * V_{dc}$	+1
0	0
$-0.1084 * V_{dc}$	-1

The CHB stages can be switched to generate voltages of $+0.1084V_{dc}$, 0 or $-0.1084V_{dc}$. The capacitor fed 2-level stage of secondary inverter can be switched to generate voltages of 0 or $+0.325V_{dc}$. The CHB stage voltages and 2-level voltages

0278-0046 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 06:36:29 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3007079, IEEE Transactions on Industrial Electronics

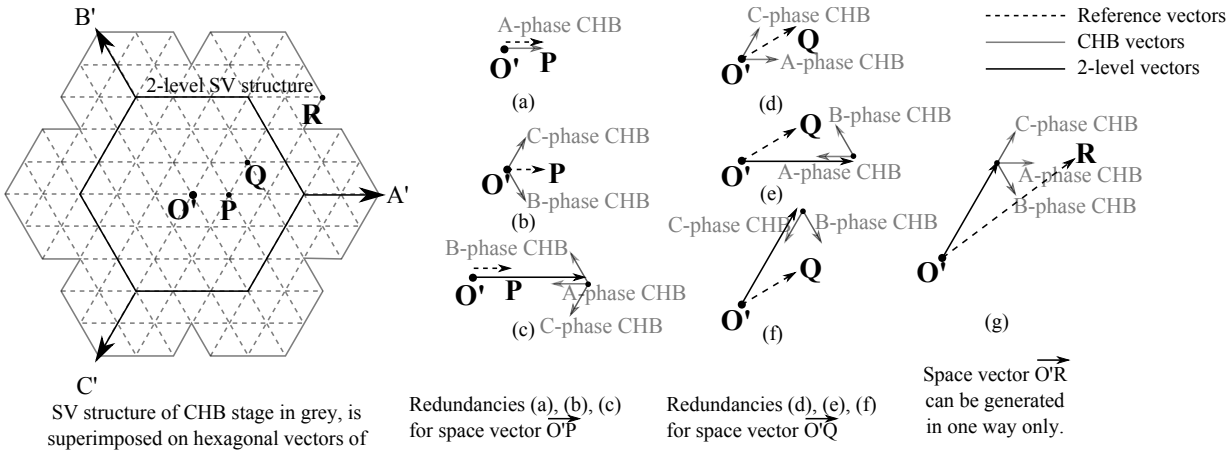


Fig. 4. Space Vector redundancies for reference vectors OP, OQ and OR.

in the three phases can be represented by the switching state codes represented in Table.I.

Consider the vector $O'P$ (Fig.4), which is to be applied for t_p time, to generate the 42-sided vectors, in a switched averaged sense. $O'P$ has three SV redundancies (a), (b), (c), shown in Fig.4. The dashed lines show the reference vector $O'P$ that should be generated by superposition of CHB stage vectors and 2-level stage vectors. The CHB stage vectors are marked by gray lines, and the 2-level stage vectors are marked by bold lines. To generate redundancy (a), state $+1$ (voltage of $0.1084V_{dc}$) is applied in A-phase of CHB stage alone, and state 0, in B and C phases. The 2-level stage applies state 0 in all 3 phases. Similarly, to generate redundancy (c), states [-1,+1,+1] are applied in the three phases CHB stage and states [+3,0,0] are applied in the three phases of 2-level stage. Since, the polarity of A-phase CHB stage is opposite in the two redundancies (+1 for redundancy (a) and -1 for redundancy (c)), one of the redundancies charges the A-phase capacitor, whereas the other redundancy discharges the capacitor. The charge balance can be maintained on A-phase capacitor, if the two redundancies (a) and (c), are applied for equal time. Similarly, for redundancy (b), the states for CHB stage and 2-level stage are [0,-1,-1] and [0,0,0]. The CHB states for redundancies (b) and (c), are opposite in B and C phases, if the two redundancies are applied for equal times, then charge balance is maintained in B and C phase capacitor voltages.

The secondary vector $O'P$ can be generated by applying any of the three redundancies. Thus, the three redundancies are applied for times t_{p1}, t_{p2} and t_{p3}, such that,

$$t_p = t_{p1} + t_{p2} + t_{p3}$$

Applying the the redundancies for times t_{p1}, t_{p2}, t_{p3}, is equivalent to applying the vector $O'P$ for time t_p. The three redundancy times are selected to be $t_p/3$, for voltage balancing of capacitors in the three phases, as explained earlier. The switching state codes in the three phases for CHB stage and 2-level stage for the three redundancies are shown in Table.II.

If the capacitor voltages are disturbed due to circuit non-idealities, the timings for each of the redundancies is slightly adjusted, so as to reduce the error in capacitor voltages.

The 2-level capacitive supply is assumed to be stiff voltage, in the discussion for CHB stage capacitor balancing, as it is maintained at the nominal voltage by power balance relations discussed earlier.

TABLE II
SPACE VECTOR REDUNDANCIES FOR VECTOR O'P

Reference Vector($O'P$) redundancy	CHB state	2-level state
(a)	[+1,0,0]	[0,0,0]
(b)	[0,-1,-1]	[0,0,0]
(c)	[-1,+1,+1]	[+3,0,0]

TABLE III
SPACE VECTOR REDUNDANCIES FOR VECTOR O'Q

Reference Vector($O'Q$) redundancy	CHB state	2-level state
(d)	[+1,0,-1]	[0,0,0]
(e)	[-1,+1,0]	[+3,0,0]
(f)	[0,-1,+1]	[+3,+3,0]

The vector $O'Q$ has three SV redundancies (d), (e), (f), shown in Fig.4. The states for CHB stage and 2-level stage are shown in Table.III. A-phase CHB capacitor is balanced by applying redundancies (d) and (e) for equal time, since they have opposite polarity for A-phase CHB state. Similarly, B-phase CHB capacitor is balanced by applying redundancies-(e) and (f) for equal time, since they have opposite polarity for B-phase CHB state. If the 3 redundancies are applied for equal times, then charge balance is maintained on all three CHB capacitors. Thus, if the vector $O'Q$ needs to be applied for t_q time, then each of the redundancies are applied for $t_q/3$ duration to balance the CHB capacitors in the 3 phases. If the capacitor voltages are disturbed due to circuit non-idealities, the timings for each of the redundancies is slightly adjusted, so as to reduce the error in capacitor voltages.

Vector $O'R$ does not have sufficient redundancies to achieve switching cycle balancing. If the vector $O'R$ is to be applied for t_r time, the capacitors will get charged or discharged for t_r time. The capacitors are balanced back to their nominal voltage during the application of next vector. It can be noted that the outer vectors like $O'R$ only need to be switched while generation of 42-sided vectors OP_{40} and OP_4, shown in Fig.2. The vectors that need to be applied to generate, other 42-sided vectors (OP_{41}, OP_{42}, OP_1, OP_2 and OP_3), shown in Fig.2, have three redundancies, so that CHB stage capacitor voltages can be balanced during there vector periods. Thus, the CHB stage capacitors can be balanced in switching cycle periods.

0278-0046 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 06:36:29 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3007079, IEEE Transactions on Industrial Electronics

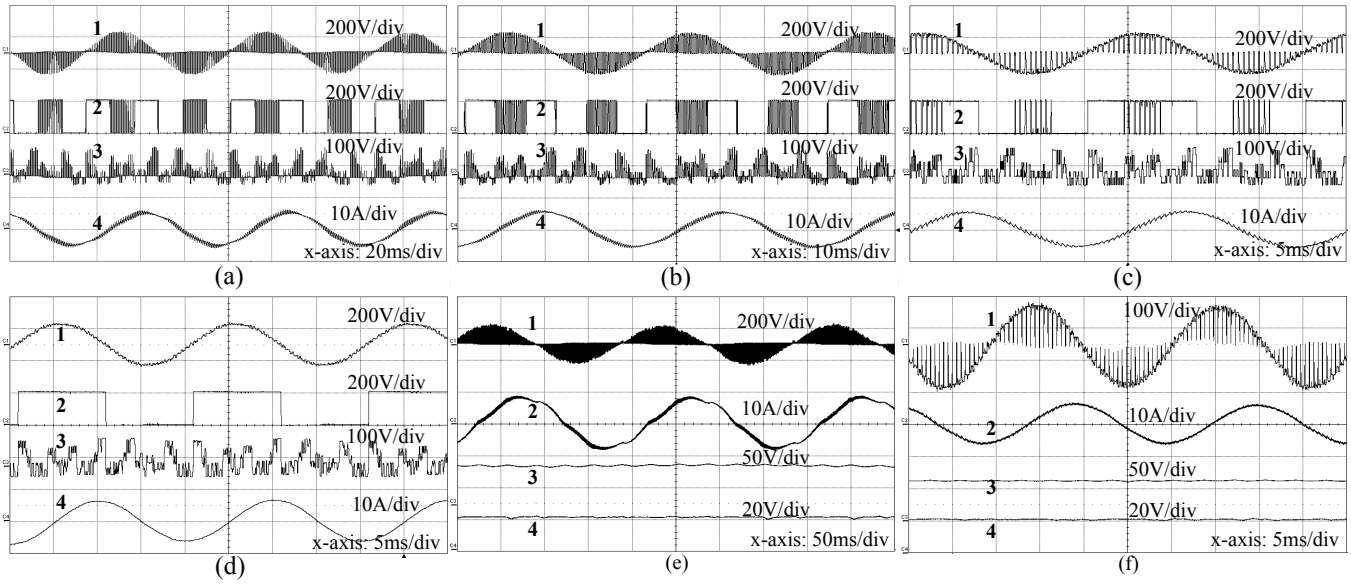


Fig. 5. Steady state Hardware results at- **(a)** 15Hz **(b)** 24Hz **(c)** 40Hz **(d)** 50Hz
Traces: **(1)** Motor phase voltage ($V_{AA'}$), **(2)** Primary pole voltage (V_{AO}), **(3)** Secondary pole voltage ($V_{A'O'}$), **(4)** Phase current (I_A).
Hardware results: Capacitor balancing results at- **(e)** 5Hz **(f)** 48Hz
Traces: **(1)** Motor phase voltage ($V_{AA'}$), **(2)** Phase current (I_A), **(3)** Secondary 2-level capacitor voltage ($V_{2-level}$), **(4)** A-phase CHB capacitor voltage ($V_{CHB,A}$).

IV. Experimental Results

TI-DSP TMS320F28335 and FPGA Spartan-XC3S200 platform are used for the hardware implementation of the proposed scheme. SKM75GB12T4 half-bridge modules are used to realize inverter switches. The 3-phase induction motor used for experiments is a 4-pole, 5kW, 200V, 50Hz Induction Motor. The block diagram of experimental setup is shown in Fig.6.

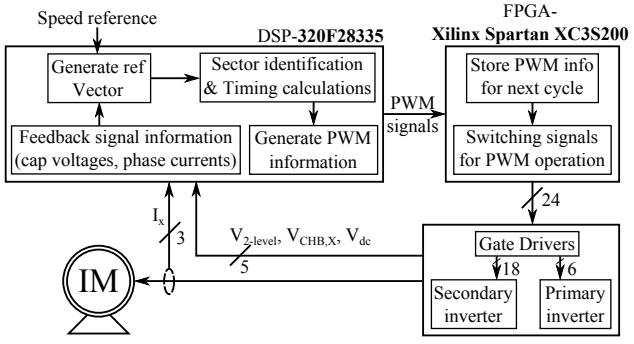


Fig. 6. Block Diagram of Experimental Setup.

The experimental results for steady state inverter operation are shown in Fig.5. The figure shows motor phase voltage, primary pole voltage, secondary pole voltage and phase current for 15Hz, 24Hz and 40Hz operation in steady state. The secondary inverter CHB capacitors and 2-level capacitor are balanced at 21.7V ($0.1084V_{dc}$) and 65V ($0.325V_{dc}$) for 200V DC link supply voltage. The secondary 6-level inverter pole voltage can vary from -21.7V to 86.7V. The envelope of phase voltage is identical for different speeds of operation. For lower speeds of operation the fundamental phase voltage is reduced, by switching the zero vector to reduce the average voltage being applied. The motor inductance further attenuates the harmonics in phase currents. Hence, the phase current is a sinusoidal signal. Fig.5(d) shows the experimental results for 50Hz operation, the extreme case of over modulation. The motor phase voltage and current are nearly sinusoidal signals. The primary inverter pole voltage is a square wave at the extreme case of over modulation. The low frequency switching on primary inverter results in lower switching losses. The secondary inverter fed with capacitive supply, is switched at higher frequencies for suppression of lower order harmonics. Since, the secondary inverter operates on voltages much lower than DC link supply, the switching losses are reduced.

Fig.5(e)-(f) show motor phase voltage, phase current, secondary 2-level capacitor voltage and A-phase CHB capacitor voltage for 5Hz and 48Hz operation in steady state. It can be observed that the capacitor voltages are balanced at their nominal voltages of 65V and 21.7V for 2-level and CHB capacitors of secondary inverter respectively.

The capacitor voltages are balanced in one switching cycle period. The capacitors are designed to have within $\pm 5\%$ deviation from nominal voltages, in the worst case. The capacitance value is determined using the relation,

$$C = \frac{I_{pk}.T_{sw}}{\Delta V_c} \tag{9}$$

Where, I_{pk}, T_{sw} and ΔV_c, are the peak of phase current, switching cycle period and deviation in capacitor voltage. For the base speed of 50Hz, the switching cycle period for 42-sided vectors is $476\mu sec$ $(= 1/(50 * 42))$. For peak current of 10A and $\pm 5\%$ voltage deviation in the worst case, the capacitance should be greater than $4.38mF$. Capacitance values of $4.4mF$ are used for experimental set up.

Fig.7 shows the FFT results for 50Hz operation, the extreme case of over modulation. The primary pole voltage generates lower order harmonics that are canceled by the secondary inverter pole voltage. The lower order harmonics are suppressed

0278-0046 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 06:36:29 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3007079, IEEE Transactions on Industrial Electronics

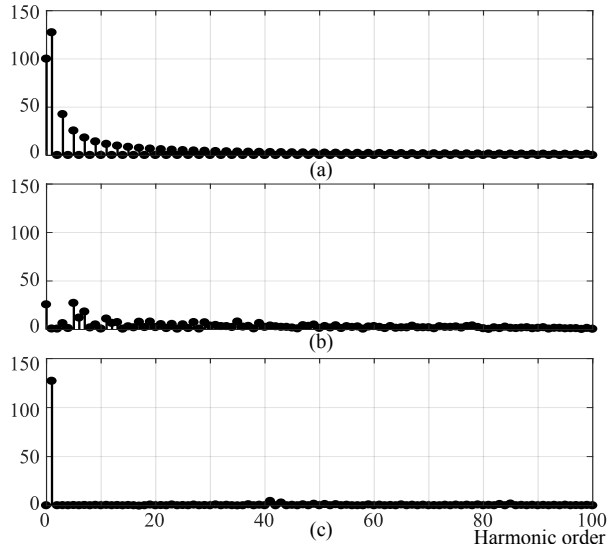


Fig. 7. FFT results at 50Hz operation. **(a)** Primary pole voltage (V_{AO}) **(b)** Secondary pole voltage ($V_{A'O'}$) **(c)** Motor phase voltage ($V_{AA'}$)

in the motor phase voltage, resulting in smooth torque profile for motor operation.

The experimental results for motor acceleration from 5Hz to 50Hz acceleration are shown in Fig.8. The waveforms for motor phase voltage, phase current, 2-level capacitor voltage and CHB A-phase capacitor voltage are shown for acceleration results. The motor phase current increases during acceleration. The capacitor voltages are balanced close to their nominal value throughout transients. Thus, the inverter topology is experimentally tested for steady state and transient operation.

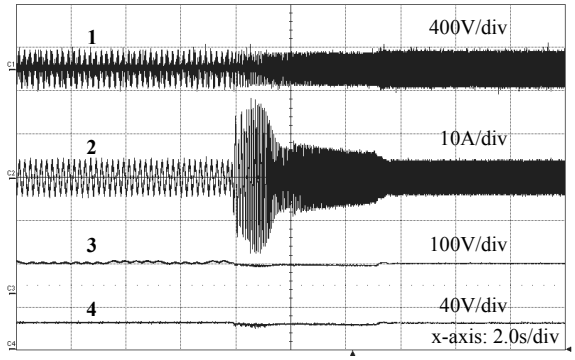


Fig. 8. Transient results: Accelaration from 5Hz to 50Hz.
Traces: **(1)** Phase voltage ($V_{AA'}$), **(2)** Phase current (I_A), **(3)** Secondary 2-level capacitor voltage ($V_{2-level}$), **(4)** A-phase CHB capacitor voltage ($V_{CHB,A}$)

Fig.9 shows the comparison of WTHD performance of a 2-level inverter and the proposed scheme. The proposed scheme achieves better harmonic performance due to suppression of lower order harmonics by secondary inverter.

V. Conclusion

A scheme for suppression of all lower order harmonics up to 39^{th} harmonic in motor phase voltage is proposed. The harmonic suppression is achieved by generating a 42-sided voltage space vector structure. The harmonics generated by the primary inverter are suppressed by using a secondary inverter, which acts like a switched-capacitive filter. The new power circuit topology proposed for secondary inverter

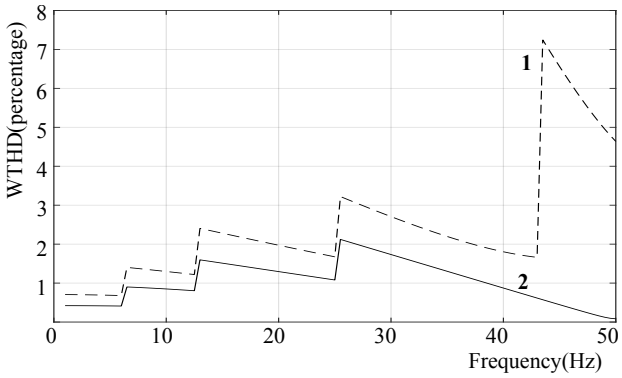


Fig. 9. Weighted Total Harmonic Distortion (WTHD) plots.
Traces: **(1)** WTHD of phase voltage for 2-level inverter **(2)** WTHD of phase voltage for proposed 42-sided polygonal SVS.

generates a dense 6-level SV structure. A switching cycle balancing scheme for CHB stage capacitors using space vector redundancies is proposed. A new capacitor balancing scheme based on power balance is proposed in this work for the first time, which balances secondary inverter 2-level capacitor voltage while suppressing all lower order harmonics up to 39^{th} harmonic. The harmonic elimination can be observed in the FFT waveforms of phase voltage. The power circuit topology proposed in the work has been studied and verified extensively in hardware. The work offers a new solution for elimination of lower order harmonics in phase voltage using minimum component count, without using conventional bulky LC filters. Here, high frequency switching is on low voltage side alone, the component size and system cost are reduced, and with the use of switched capacitive filtering, the lower order harmonics are eliminated for the full modulation range up to the base speed.

References

[1] A. Tripathi and G. Narayanan, "Evaluation and minimization of low-order harmonic torque in low-switching-frequency inverter-fed induction motor drives," *IEEE Transactions on Industry Applications*, vol. 52, no. 2, pp. 1477–1488, March 2016.

[2] J. Song-Manguelle, J. Nyobe-Yome, and G. Ekemb, "Pulsating torques in pwm multi-megawatt drives for torsional analysis of large shafts," *IEEE Transactions on Industry Applications*, vol. 46, no. 1, pp. 130–138, Jan 2010.

[3] V. Dzhankhotov and J. Pyrhönen, "Passive lc filter design considerations for motor applications," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 10, pp. 4253–4259, Oct 2013.

[4] V. Dzhankhotov, J. Pyrhonen, P. Silventoinen, M. Kuisma, and T. A. Minav, "A new passive hybrid air-core foil filter for modern power drives," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 5, pp. 1757–1762, May 2011.

[5] D. Ahmadi, K. Zou, C. Li, Y. Huang, and J. Wang, "A universal selective harmonic elimination method for high-power inverters," *IEEE Transactions on Power Electronics*, vol. 26, no. 10, pp. 2743–2752, Oct 2011.

[6] J. R. Wells, B. M. Nee, P. L. Chapman, and P. T. Krein, "Selective harmonic control: a general problem formulation and selected solutions," *IEEE Transactions on Power Electronics*, vol. 20, no. 6, pp. 1337–1345, Nov 2005.

[7] Li Li, D. Czarkowski, Yaguang Liu, and P. Pillay, "Multilevel selective harmonic elimination pwm technique in series-connected voltage inverters," *IEEE Transactions on Industry Applications*, vol. 36, no. 1, pp. 160–170, Jan 2000.

[8] H. Zhao, T. Jin, S. Wang, and L. Sun, "A real-time selective harmonic elimination based on a transient-free inner closed-loop control for cascaded multilevel inverters," *IEEE Transactions on Power Electronics*, vol. 31, no. 2, pp. 1000–1014, Feb 2016.

0278-0046 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 06:36:29 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3007079, IEEE Transactions on Industrial Electronics

[9] N. Flourentzou, V. G. Agelidis, and G. D. Demetriades, "Vsc-based hvdc power transmission systems: An overview," *IEEE Transactions on Power Electronics*, vol. 24, no. 3, pp. 592–602, March 2009.
[10] H. Abu-Rub, J. Holtz, J. Rodriguez, and G. Baoming, "Medium-voltage multilevel converters—state of the art, challenges, and requirements in industrial applications," *IEEE Transactions on Industrial Electronics*, vol. 57, no. 8, pp. 2581–2596, Aug 2010.
[11] L. G. Franquelo, J. I. Leon, and E. Dominguez, "Recent advances in high-power industrial applications," in *2010 IEEE International Symposium on Industrial Electronics*, July 2010, pp. 5–10.
[12] S. Rohner, S. Bernet, M. Hiller, and R. Sommer, "Modelling, simulation and analysis of a modular multilevel converter for medium voltage applications," in *2010 IEEE International Conference on Industrial Technology*, March 2010, pp. 775–782.
[13] A. Das, K. Sivakumar, R. Ramchand, C. Patel, and K. Gopakumar, "A combination of hexagonal and 12-sided polygonal voltage space vector pwm control for im drives using cascaded two-level inverters," *IEEE Transactions on Industrial Electronics*, vol. 56, no. 5, pp. 1657–1664, May 2009.
[14] N. A. Azeez, S. Pramanick, K. Gopakumar, and C. Cecati, "A 5th and 7th order harmonic suppression scheme using capacitive filtering for 2-level vsi fed im drive," in *2014 International Symposium on Power Electronics, Electrical Drives, Automation and Motion*, June 2014, pp. 780–783.
[15] S. Pramanick, N. A. Azeez, R. Sudharshan Kaarthik, K. Gopakumar, and C. Cecati, "Low-order harmonic suppression for open-end winding im with dodecagonal space vector using a single dc-link supply," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 9, pp. 5340–5347, Sep. 2015.
[16] K. Raj R, K. Gopakumar, M. Boby, A. K. Yadav, L. G. Franquelo, and S. S. Williamson, "Multilevel 24-sided polygonal voltage-space-vector structure generation for an im drive using a single dc source," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 2, pp. 1023–1031, Feb 2019.



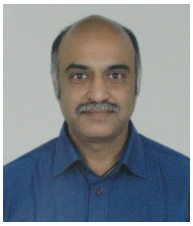
Rahul Dewani (S'19) received the B.Tech. and M.Tech. degrees from the Delhi Technological University, Delhi, India and Indian Institute of Technology, Delhi, India, in 2014 and 2016, respectively. He was with Texas Instruments, Bangalore, India in 2016, and Aura Semiconductor, Bangalore, India from 2016 to 2018. He is currently working towards the Ph.D. degree at the Department of Electronic Systems Engineering, Indian Institute of Science. His research interests include renewable energy and IM drive applications.



Rakesh R (S'18) received the B.Tech. degree in electrical and electronics engineering from the College of Engineering, Trivandrum, India, in 2014 and the M.Tech. degree in electronic systems engineering from the Indian Institute of Science, Bangalore, India, in 2017. He is currently working towards the Ph.D. degree at the Department of Electronic Systems Engineering, Indian Institute of Science. His research interests include areas of power converters and drives.



K. Gopakumar (M'94-SM'96-F'11) received the B.E., M.Sc. (Engg.), and Ph.D. degrees from the Indian Institute of Science, Bangalore, India, in 1980, 1984, and 1994, respectively. He was with the Indian Space Research Organization, Bangalore, India from 1984 to 1987. He currently holds the position of professor at the Department of Electronic Systems Engineering, Indian Institute of Science. Dr. Gopakumar is a Fellow of Institution of Electrical and Telecommunication Engineers, India and Indian National Academy of Engineers. He is currently a Co-Editor in Chief of IEEE Transaction on Industrial Electronics and also a Distinguished Lecturer of IEEE Industrial Electronics Society(IES). His research interests include PWM converters and high power drives.



Umanand Loganathan is an Associate Professor at Department of Electronics Systems Engineering (DESE) in Indian Institute of Science (IISc). He received the Bachelor's degree in electronics and communication from Bangalore University, Bangalore, India, in 1987, the M.Tech. degree in electronics design, and the Doctoral degree in the area of control of high-performance induction motor drives from the Indian Institute of Science, Bangalore, India, in 1989 and 1996 respectively. His major research interests are in photovoltaic system design, bond graph modeling of power electronic systems, high performance control of induction motor, designing for reliability and hybrid electric vehicles.



Dariusz Zielinski received his Ph.D. degree in Electrical Engineering from the Lublin University of Technology in 2017. Currently he supervises the research Laboratory for the conversion and storage of energy from renewable sources. He plays a leading role in the projects PLUGinEV (electric vehicle fast-charging system) and Active Prosumer, financed by the National Centre for Research and Development. His research interests cover power electronics, grid-tied converters, microgrids and energy storage systems. His most significant works focus on control algorithms for multi-channel power converters, bidirectional power converters, hybrid energy storage systems, grid synchronisation, island mode detection methods and DC-DC converters. He is a member of IEEE.



Leopoldo G. Franquelo (M'84–SM'96–F'05) was born in Malaga, Spain. He received the M.Sc. and Ph.D. degrees in electrical engineering from the University of Sevilla, Seville, Spain, in 1977 and 1980, respectively. His research interests include modulation techniques for multilevel inverters and its application to power electronic systems for renewable energy systems. Dr. Franquelo was a Member-at-Large of IES AdCom from 2002 to 2003, the Vice President for Conferences from 2004 to 2007, and the President Elect of the IES from 2008 to 2009. He was the recipient of a number of best paper awards from journals of the IEEE, and the Eugene Mittelmann Award and the Anthony J. Hornfeck Service Award from Industrial Electronics Society (IES) in 2012 and 2015, respectively. He has been an IES Distinguished Lecturer since 2006, an Associate Editor for the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS since 2007, the Co-Editor in Chief since 2014, and the Editor in Chief from 2016 to 2018. He was the President of the IEEE Industrial Electronics Society.

0278-0046 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 06:36:29 UTC from IEEE Xplore. Restrictions apply.

Nanoscale Advances

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Check for updates

Cite this: DOI: 10.1039/d0na00309c

The prospective application of a graphene/MoS_2 heterostructure in Si-HIT solar cells for higher efficiency†

Chandra Kamal Borah,[a] Pawan K. Tyagi[bc] and Sanjeev Kumar *[a]

The efficiency of a Si-HIT (heterojunction with intrinsic thin layer) solar cell based on a graphene/MoS_2 heterostructure has been optimized by varying the various parameters of graphene (Gr) as a transparent conducting electrode (TCE) and n-type molybdenum disulfide (n-MoS_2) as an emitter layer. The photovoltaic performance of a graphene/n-MoS_2/a-Si:H/p-cSi/Au single facial HIT solar cell has been studied using AFORS-HET v2.5 simulation software. A maximum output efficiency of 25.61% has been achieved. The obtained results were compared with the results from a commercially available a-Si:H layer and p-cSi wafer after simulation. Moreover, the dependence of the cell performance on changes in the TCE and the back contact materials has also been studied. Finally, it has been demonstrated that the graphene layer and n-MoS_2 layer could act as a TCE and an efficient emitter layer, respectively, in a n-MoS_2/p-cSi based HIT solar cell.

Received 18th April 2020
Accepted 8th June 2020

DOI: 10.1039/d0na00309c

rsc.li/nanoscale-advances

Open Access Article. Published on 23 June 2020. Downloaded on 6/24/2020 4:25:19 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

1. Introduction

The output efficiency (η) of a simple silicon (Si) heterojunction solar cell degrades owing to the presence of a large number of defects at the surface of the Si wafer. This deficiency of heterojunction solar cells can be overcome by the introduction of a thin intrinsic layer between the two doped layers of the heterojunction. Such a structure is known as a heterojunction with intrinsic thin layer (HIT) structure.[1] In a HIT structured solar cell, the intrinsic layer-like a-Si:H reduces the interfacial recombination state density to the minimum level between two doped layers[2] and reduces the leakage current.[3] Recently, SANYO Ltd reported attaining a 20% output efficiency in HIT solar cells based on n-type Si.[4] Furthermore, Panasonic has achieved the world's highest efficiency of 25.6% in a HIT cell based on an n-type c-Si.[5]

In a solar cell, a transparent conductive electrode (TCE) is required to ensure the easy transportation of carriers from the emitter layer to the external circuit. The low sheet resistance (10 Ω sq^{-1}) and high optical transparency (85%) mean that indium tin oxide (ITO) is a commonly used TCE in photovoltaic (PV) applications. However, the high cost, rarity and mechanical rigidity of ITO have limited its use for future flexible devices and these drawbacks initiated the search for new TCE materials.[6] Recently, pristine graphene has been reported to have an optical transparency as high as 97.7% (single layer graphene) with a work function of 4.31–4.5 eV (ref. 7 and 8) and a sheet resistance of 120 Ω sq^{-1} in few-layer graphene.[6] In addition, the intercalated graphene with ferric chloride is reported to have a high optical transparency of approximately 96% with a low sheet resistance of 8.8 Ω sq^{-1} and a high carrier density of 8.9×10^{14} cm^{-2} at room temperature and a high stability in air.[6] At room temperature, the carrier mobility in the graphene layer exceeds 10^5 cm^2 V^{-1} s^{-1}.[9] These extraordinary properties have led to the application of multilayer graphene as a TCE. Recently, Lazar *et al.*[7] reported that doping with boron (B) and $-NH_2$ can increase the work function of graphene (Φ_{Gr}) to 1.20 eV and 0.46 eV from 4.31 eV (pristine graphene). Kim *et al.*[8] demonstrated that, after being irradiated by an α beam the work function of graphene increases to 4.9 eV. Conversely, the low photoresponsivity, poor external quantum efficiency, and low absorption of light reduces the application of graphene as an emitter layer instead of the TCE layer in HIT solar cell.[10] Unlike graphene, another 2D material, molybdenum disulfide (MoS_2) has emerged as a complement 2D material for graphene as an emitter layer in HIT solar cells. MoS_2 has a mobility of 200 cm^2 V^{-1} s^{-1} in the monolayer and approximately 517 cm^2 V^{-1} s^{-1} in few layers, as well as a high on/off ratio in the order of 10^8 (ref. 11 and 12) and a fast photo-response time of 40 μs.[13] MoS_2 is reported to have an indirect bandgap of 1.2 eV in the bulk form and a direct bandgap of 1.9 eV in the single layer form.[14] Moreover, double layer MoS_2 has more than 85% optical transparency in the visible region.[15] Recently, it has been reported that the work function of MoS_2 changes with a change in the number of layers.[16] It has also been demonstrated that MoS_2

[a]Centre of Advanced Research, Department of Physics, Rajiv Gandhi University, Arunachal Pradesh-791112, India. E-mail: sanjeev.kumar@rgu.ac.in
[b]Department of Physics, Central University of Haryana, Haryana-123029, India
[c]Department of Applied Physics, Delhi Technological University, Delhi-110042, India

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0na00309c

Open Access Article. Published on 23 June 2020. Downloaded on 6/24/2020 4:25:19 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

can be used to form significant n-type doping with a Benzyl Viologen (BV) dopant[17] and polyethyleneimine (PEI)[18]. PEI reduces the sheet resistance of MoS_2 2.6 times.[18] These properties mean that MoS_2 can be utilized as an emitter layer in MoS_2/silicon HIT solar cells.

Experimentally and theoretically, it has been demonstrated that MoS_2 can act as an emitter layer in heterojunction silicon solar cells. Tsai *et al.*[19] fabricated a heterojunction solar cell of monolayer MoS_2 and p-cSi and achieved a power conversion efficiency of 5.23%. Hao *et al.*[20] also fabricated a bulk p–n junction after depositing a 40 nm thick layer of n-MoS_2 onto p-cSi substrates and a 1.3% photoconversion efficiency was achieved.

The high transparency, as well as the high mobility of the graphene layer with a tunable bandgap and work function, along with practical and theoretical demonstrations of graphene/MoS_2 and MoS_2/Si heterojunctions, motivated us to propose a single facial HIT solar cell structure of graphene/n-MoS_2/a-Si:H/p-cSi/Au. Here, we have used graphene as a TCE material. No reports have been published on the use of graphene as a TCE in HIT solar cells so far. In addition to graphene and MoS_2, there are a few interesting candidates such as carbon nanotubes, black phosphorous and MXene, that show potential for use in silicon solar cells. So far, a power conversion as high as 17.2% has been observed in carbon nanotube based silicon solar cells.[21–23] After proposing the structure of the cell, the influence of various parameters of the layers used on the cell performance were studied using AFORS-HET v2.5 simulation software. The main objective of the present simulation is to study the independent effects owing to the variation of the parameters of the layers on the cell performance and to demonstrate the prospective application of graphene as an effective TCE material in a n-MoS_2/p-cSi HIT solar cell.

2. Physical model, structural model and formation of the junctions

2.1 Physical model

AFORS-HET is solar cell simulation software, used to study and analyze the various properties of a heterojunction solar cell. For the proposed structure, this software solves Poisson's equation and the transport equation for electrons and holes in one dimension. These equations are as follows:

$$\frac{\varepsilon_0\varepsilon_r}{q}\frac{\partial^2\varphi(x,t)}{\partial x^2} = p(x,t) - n(x,t) + N_D - N_A + \sum_{\text{defects}}\rho_t(x,t)$$

$$-\frac{1}{q}\frac{\partial j_n(x,t)}{\partial x} = G_n(x,t) - R_n(x,t) - \frac{\partial}{\partial t}n(x,t)$$

$$\frac{1}{q}\frac{\partial j_p(x,t)}{\partial x} = G_p(x,t) - R_p(x,t) - \frac{\partial}{\partial t}p(x,t)$$

In which, n represents the electron density, p is the hole density, φ is the electric potential, q is the electron charge, $\varepsilon_0\varepsilon_r$ is the absolute/relative dielectric constant, and $N_{D/A}$ represents the donors/acceptors concentrations, which are assumed to be completely ionized. Here, n, p and φ are the independent variables for which the system of differential equations is solved.

The charge stored in the defects is described by a distribution function, f_t, specifying the probability that defects with a defect density N_t are at position E if the bandgap is occupied with electrons:

$$\text{Acceptor-type defect}: \quad \rho_t(x,t) = -\int dE f_t(E,x,t)N_t(E)$$

$$\text{Donor-type defect}: \quad \rho_t(x,t) = -\int dE[1 - f_t(E,x,t)]N_t(E)$$

The electron/hole currents ($j_{n/p}$) flow owing to the gradient of the corresponding quasi-Fermi energy ($E_{F_{n/p}}$). Within the semiconducting layer, this is equivalent to the sum of the diffusion and drift current with the corresponding mobility $\mu_{n/p}$.

$$j_n(x,t) = q\mu_n n(x,t)\frac{\partial E_{F_n}(x,t)}{\partial x} = -\frac{\mu_n kT}{q}\frac{\partial n(x,t)}{\partial x} + \mu_n n(x,t)\frac{\partial\varphi(x,t)}{\partial x}$$

$$j_p(x,t) = q\mu_p p(x,t)\frac{\partial E_{F_p}(x,t)}{\partial x} = -\frac{\mu_p kT}{q}\frac{\partial p(x,t)}{\partial x} - \mu_p p(x,t)\frac{\partial\varphi(x,t)}{\partial x}$$

If the heterostructure is illuminated (specifying the spectral distribution $\phi(\lambda)$ of the incoming photon flux), the super-bandgap optical generation rate $G_n = G_p\left(\text{for } \frac{hc}{\lambda} \geq E_g\right)$ from the valence band into the conduction band of the semiconductor layers can be evaluated by considering the Lambert–Beer absorption model by specifying the spectral absorption coefficient α, as well as specifying the dielectric properties (n,k) of each layer. An optical sub-bandgap generation $\left(\text{for } \frac{hc}{\lambda} < E_g\right)$ from a defect to the conduction/valence band can be defined by specifying the optical emission coefficients $e_n^o(E)$,$e_p^o(E) \neq 0$ for the defect state.

Recombination from the conduction band into the valence band may occur directly (band to band recombination, Auger recombination) and *via* trap states (Shockley–Read–Hall recombination, (SHR)).

$$R_{n,p}(x,t) = R_{n,p}^{\text{Band–Band}}(x,t) + R_{n,p}^{\text{Auger}}(x,t) + R_{n,p}^{\text{SRH}}(x,t)$$

The transportation across the heterojunction interface is modeled by drift-diffusion. In order to do so, an interface layer is considered. The specified interface defects are distributed homogeneously within this layer. The cell currents have to be

 This journal is © The Royal Society of Chemistry 2020

Open Access Article. Published on 23 June 2020. Downloaded on 6/24/2020 4:25:19 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

evaluated directly from the gradient of the corresponding quasi-Fermi energies. The electron affinity (χ), the bandgap (E_g) and the effective conduction/valence band density of states (N_C, N_V) will depend on the position x within the interface layer.

$$E_{F_n}(x) = -q\chi(x) + q\varphi(x) - kT \ln \frac{n(x)}{N_C(x)}$$

$$E_{F_p}(x) = -q\chi(x) + q\varphi(x) + kT \ln \frac{n(x)}{N_V(x)}$$

The electric potential is fixed to zero at one contact. At the second contact, a boundary condition has to be specified, which relates the external cell voltage/current to the internal quantities. Furthermore, the electron and hole currents moving into the metal contacts need to be modeled. The Schottky contact can either be voltage controlled, or current controlled.

2.2. Structural model

Layered structure MoS_2 can be directly deposited on Si substrates. However, owing to their structural mismatch (*i.e.* layered structure of MoS_2 and diamond-like structure of Si), large quantities of lattice defects are produced and this increases the recombination at the interface. Thus, the introduction of an intrinsic (i) buffer layer a-Si:H could be beneficial to balance the carrier injection and reduce the leakage current

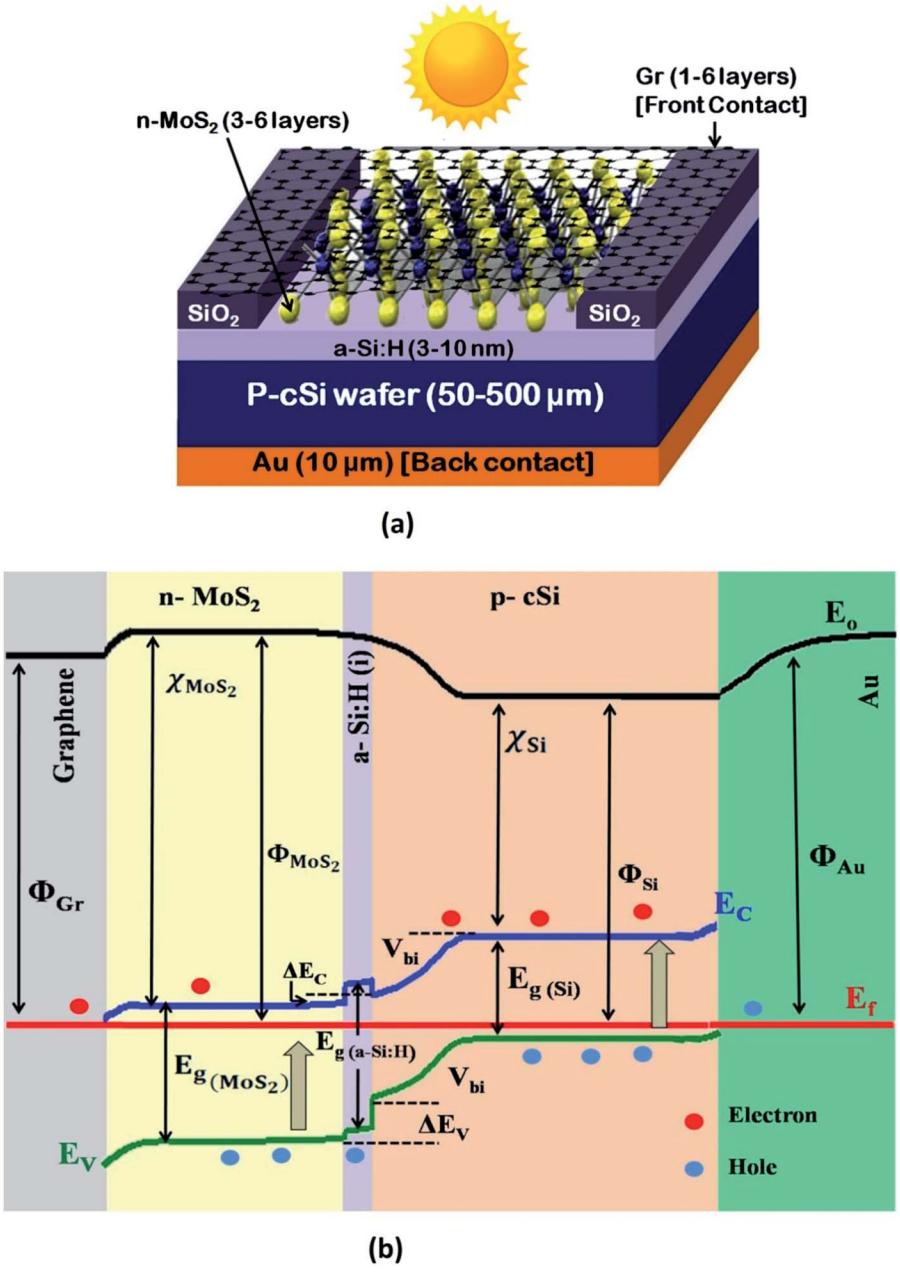


Fig. 1 (a) The proposed HIT solar cell structure, and (b) band diagram as generated using AFORS-HET software for the proposed structure.

Open Access Article. Published on 23 June 2020. Downloaded on 6/24/2020 4:25:19 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

in the cell.[3] Therefore, the as-synthesized MoS_2 is an n-type semiconductor[24] and p-type silicon is more popular in the PV industries.[4] Thus, in this simulation, we have used p-cSi as a base material. Moreover, the minority carrier diffusion length in p-cSi is also reported to be higher than n-cSi.[25] Furthermore, Fan *et al.*[26] have reported that ultrathin gold (Au) contacts provide a better electrical conductivity and optical reflectance-scattering to maintain the performance of the cell compared with ITO. Also, it has been demonstrated that an Au/Si interface formed by using novel transfer printing technology makes excellent electrical contact and exhibits a low contact resistance in comparison to that formed by vacuum deposition of Au on Si.[27] Therefore, the proposed structure of the single facial HIT solar cell is configured as graphene/n-MoS_2/a-Si:H(i)/p-cSi/Au, as shown in Fig. 1. No lattice mismatch between the layers has been considered for this simulation. Ideally, a semiconductor is called 2D if it satisfies the condition $\langle t \rangle \leq \lambda_F$, in which $\langle t \rangle$ is the average system of the 2D electron system, λ_F is the Fermi wavelength which is equal to $\sqrt{\frac{2\pi}{n_{2D}}}$ nm, and n_{2D} is the carrier density in cm^{-2}.[28] Following this condition as reported by Sarma *et al.*[29] and Suh *et al.*[30] even $n \sim 10^5$ cm^{-3} in graphene and $n \sim 1.8 \times 10^{14}$ cm^{-3} in MoS_2 cannot satisfy the above described conditions respectively. It has also been reported that doping in 2D materials increases the interlayer spacing.[31] Thus, we consider graphene and MoS_2 quasi 3D in nature instead of 2D. The quantum confinement in the structure was achieved by controlling the layer number of n-MoS_2. In the proposed HIT structure, as shown in Fig. 1a the carrier transport occurred along the *c*-axis. Details of the front contact and the back contact parameters of the proposed structure are given in Table 1[6,7,26,32–37] and the parameters, as well as the ranges of the n-MoS_2, a-Si:H and p-cSi are given in Table 2. The thickness of the single graphene was taken as 0.334 nm and the monolayer MoS_2 as 0.65 nm.[19,38,39] However, we started our simulation from five-layer (5L) graphene (1.67 nm) and six-layer (6L) MoS_2 (3.9 nm) by considering the high overlapping probability of the graphene and p-cSi. It was also reported that the Raman spectrum of 6L MoS_2 is very close to the bulk[37] and the optical transparency of 5L graphene (88.5%) is more than that of ITO (85%).[6] The parameters considered for optimization of the MoS_2 layer were the donor concentration (N_D), N_C/N_V, E_g, layer thickness, χ, dielectric constant (ε_r) and electron mobility (μ_n). The values of these parameters were kept within range and these and the results of the above mentioned previous literature are tabulated in Table 2.[3,40–47] To simulate and optimize the other parameters the bulk and interfacial defect distribution and recombination rates were kept constant in the software. Furthermore, the MS-Schottky contact model was selected for the graphene/n-MoS_2 and p-cSi/Au interfaces and the drift-diffusion model was selected for the n-MoS_2/a-Si:H and a-Si:H/p-cSi interfaces. The values and parameters of these interfaces[48] are listed in Table 3. The values of these parameters were varied accordingly under air mass 1.5 (AM1.5) illuminations under a power density of 100 mW cm^{-2} and a temperature of 300 K. In order to study the effect of the layer parameters on the cell performance, the open-circuit voltage (V_{OC}), short circuit current density (J_{SC}), fill factor (FF) and efficiency (η) were analyzed. After optimizing the parameters of n-MoS_2 further simulation was extended to the a-Si:H layer and the thickness and bandgap were optimized. The values and ranges of these parameters were obtained from previously published studies.[4,49–51] Further simulations were performed to optimize the parameters, such as the N_A, N_C/N_V, χ, and thickness for p-cSi. The values and the ranges of the parameters for the p-cSi layer were taken from the previously published research.[4,47*and*48] The defect properties were kept as the default. Finally, we optimized the parameters of the contact materials, that is graphene, Au, ITO, Ag, and so forth. The work function (Φ) and the number of layers were optimized and other parameters were fixed, as found in the literature.[7,8,34–37,52]

Table 1 Front and back contact parameters

	Front contact			Back contact	
Parameter	Initially	After optimization		Initially	After optimization
Material	Graphene	ITO	Al	Au	Ag
Thickness	1L–5L (1L = 0.334 nm)	80 nm	80 nm	10 μm	10 μm
Optical properties	$n = 2.7$ $k = 1.45$	ITO *nk* (default)	$n = 1.19$ $k = 7.05$	Au *nk* (default)	Ag *nk* (default)
Work function (eV)	4.31 4.50 4.54 4.77 4.90	4.4 4.45 4.5	4.06 4.20 4.26	5.4	4.64 4.74
Absorption loss	0.023 (constant)	ITO abs (default)	0.056	0	0
External reflection	1L = 0 2L = 0.023 3L = 0.046 4L = 0.070 5L = 0.092	ITO Ref (default)	0.868	0	0
Surface condition	Plane	Plane	Plane	Plane	Plane
Internal reflection	0	0	0	0	0

 This journal is © The Royal Society of Chemistry 2020

Open Access Article. Published on 23 June 2020. Downloaded on 6/24/2020 4:25:19 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Table 2 Parameters of the layers

Parameter	n-MoS_2	a-Si:H(i)	p-cSi
Thickness	3L–6L (1L = 0.65 nm)	3–10 nm	50–500 μm
Dielectric constant (ε_r)	4–14	11.9	11.9
Electron affinity χ (eV)	3.74–4.45	3.9	4.05
Band gap, E_g (eV)	1.41–1.48	1.6–2.0	1.12
Effective conduction band density, N_C (cm^{-3})	3×10^{18} to 9×10^{20}	2.5×10^{20}	3×10^{19} to 1×10^{21}
Effective valence band density, N_V (cm^{-3})	3×10^{18} to 9×10^{20}	2.5×10^{20}	3×10^{19} to 1×10^{21}
Effective electron mobility, μ_n ($cm^2\ V^{-1}\ s^{-1}$)	517	20	1104
Effective hole mobility, μ_p ($cm^2\ V^{-1}\ s^{-1}$)	8.5	5	420
Acceptor concentration N_A (cm^{-3})	0	0	10^{15} to 10^{17}
Donor concentration N_D (cm^{-3})	10^{12} to 10^{18}	0	0
Electron thermal velocity, V_e (cm s^{-1})	10^7	10^7	10^7
Hole thermal velocity, V_h (cm s^{-1})	10^7	10^7	10^7
Layer density, (g cm^{-1})	5.06	2.328	2.328
Auger electron recombination coefficient, ($cm^6\ s^{-1}$)	$\sim 10^{-24}$	0	2.2×10^{-31}
Auger hole recombination coefficient, ($cm^6\ s^{-1}$)	$\sim 10^{-24}$	0	9.9×10^{-32}
Band-to-band recombination coefficient, ($cm^3\ s^{-1}$)	$\sim 10^{-7}$	0	9.5×10^{-15}
Defect type	Single	Conduction tail	Single
Defect charge	Acceptor	Acceptor	Acceptor
Total defect density (cm^{-3})	10^{18}	6.4×10^{19}	10^{10}
Specific defect density ($cm^{-3}\ eV^{-1}$)	10^{18}	1.83×10^{21}	10^{10}
Defect level energy (eV)	0.6	0.035 (Urbach)	0.56
Electron and hole capture cross section (cm^{-2})	10^{-14} (default)	7×10^{-16}	10^{-14}
Optical properties	$n = 4.47$; $k = 1.01$	a-Si *nk* (default)	c-Si *nk* (default)

Table 3 Interfacial Parameters

Interface	Numerical model	Remarks
Graphene/n-MoS_2	MS-Schottky contact	Surface recombination velocity for electrons and holes = 10^7 cm s^{-1} (default)
p-cSi/Au		
n-MoS_2/a-Si:H	Drift diffusion model	Interface charge = 10^{11} cm^{-2}
a-Si:H/p-cSi		

2.3 Formation of the junctions

If n-MoS_2 and p-cSi are brought into contact, a heterojunction forms and bending of the energy bands occur close to the Fermi level. This built-in potential (V_{bi}) is developed at the MoS_2/p-cSi interface. This voltage facilitates the separation of photo-generated carriers in the solar cell. Generally, in the p–n junction solar cell, the interfacial defect density reduces the overall efficiency of the cell. Therefore, better surface passivation between the two layers is required to reduce the defect density at the interface. Here, an intrinsic layer a-Si:H is inserted between n-MoS_2 and p-cSi. This layer can be deposited on Si using the plasma-enhanced chemical vapor deposition method at a low temperature (<200 °C) without degrading the bulk properties of the Si substrate.[4] Also, it introduces band offsets between the active and emitter band which are also beneficial to obstructing the reverse flow of the charge carriers. The schematic energy band diagram of the proposed HIT solar cell is shown in Fig. 1b. The build-in potential (eV_{bi}) in the HIT solar cell is defined as:

$$\begin{aligned} eV_{bi} &= E_{g(MoS_2)} + \Delta E_V + kT \ln\left(\frac{N_A N_D}{N_{V(Si)} N_{C(MoS_2)}}\right) \\ &= E_{g(Si)} - \Delta E_C + kT \ln\left(\frac{N_A N_D}{N_{V(Si)} N_{C(MoS_2)}}\right) \end{aligned} \quad (1)$$

In which e, E_g, ΔE_C and ΔE_V, k and T are the electric charge, bandgap, the conduction band offset and valence band offset, Boltzmann constant, and the temperature in Kelvin, respectively. N_A and N_D represent the acceptor and donor concentrations and N_C and N_V are the conduction band and valence band density.[53] In this HIT cell, a barrier for hole transportation is built at the interface between n-MoS_2 and the pristine graphene corresponding to the potential energy of eV_{bi} which can be expressed using the following equation:[54]

$$eV_{bi} = e(\Phi_{MoS_2} - \Phi_{Gr}) \quad (2)$$

In which, Φ_{MoS_2} and Φ_{Gr} are the work functions of n-MoS_2 and pristine graphene respectively. As reported in previously published literature,[16and55] the Φ of a six layer MoS_2 (~5.4 eV) is greater than that of pristine graphene (~4.1 eV).

3. Results and discussion

3.1 Consequences of the n-MoS_2 layer

In this section, we have optimized the donor N_D of the n-MoS_2 layer in the range from 10^{12} to 10^{18} cm^{-3}, and other parameters such as the thickness, N_C/N_V, E_g, χ and ε_r are kept at feasible values as shown in Fig. 2a and b. In Fig. 2b, it can be seen that as

Open Access Article. Published on 23 June 2020. Downloaded on 6/24/2020 4:25:19 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

N_D was increased from 10^{12} cm^{-3} to 10^{16} cm^{-3}, the V_{OC} remained constant at the value of 636.7 mV. At $N_D > 10^{16}$ cm^{-3}, the V_{OC} increased to 646.1 mV as N_D was increased to 10^{18} cm^{-3}. At a low value of N_D, the consistency of V_{OC} indicates that it does not depend on N_D unless the condition $\Delta n_o \gg N_D$ is followed, as reported previously,[56] in which Δn_o represents the excess electron–hole pairs. At $N_D > 10^{16}$ cm^{-3}, the V_{OC} increases owing to enhancement in the eV_{bi} according to eqn (2). As the mobility of the photo-generated carriers remains almost constant even with an increase of N_D, this might keep J_{SC} constant at 32.33 mA cm^{-2}. Similarly, the FF was found to be constant at 76.71% until the N_D reached 10^{16} cm^{-3} and then the FF slightly reduced to 75.73%. This is due to an increase in the recombination rate at a higher N_D (10^{18} cm^{-3}). As we have observed, the N_D controls the V_{OC}, thus the η of the cell remained constant at 15.79% as the N_D increases, and a maximum of $\eta = 15.82\%$ was obtained at $N_D = 10^{18}$ cm^{-3}.

The N_C/N_V is related to the effective mass of the electrons (m_e^*) and holes (m_h^*) as follows: $N_{C/V} \propto 2m_{e/h}^*$. Furthermore, $m_{e/h}^*$ is related to the mobility ($\mu_{e/h}$) as follows: $\mu_{e/h} = \frac{e\tau}{m_{e/h}^*}$. From these two relationships, it can be concluded that with the change in mobility ($\mu_{e/h}$) the conduction and valence band density (N_C/N_V) also changes. It has been reported that the mobility of the charge carriers in MoS_2 changes with the change in the number of layer. MoS_2 has a mobility of 200 cm^2 V^{-1} s^{-1} in the monolayer and approximately 517 cm^2 V^{-1} s^{-1} in few layers.[11] Therefore, after optimization of N_D to 10^{18} cm^{-3}, the N_C/N_V was optimized in the range 3×10^{18} to 9×10^{20} cm^{-3} as shown in Fig. 2c and d. It can be seen in Fig. 2d that the V_{OC} increased from 646.1 to 668 mV with an increase of the N_C/N_V

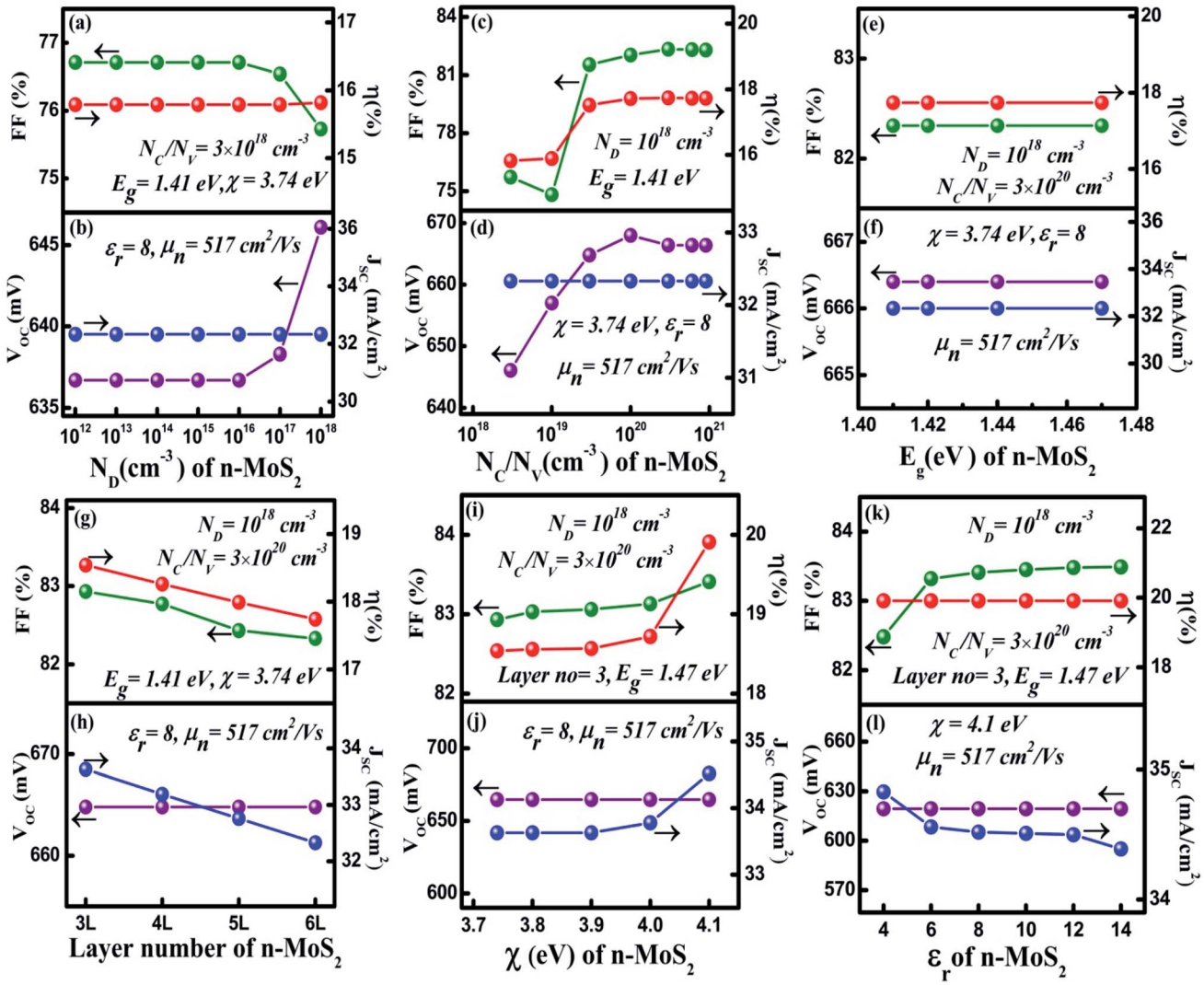


Fig. 2 Optimization of the n-MoS_2 layer: (a) and (b) donor concentration N_D (cm^{-3}), (c) and (d) effective conduction band and valance band density N_C/N_V (cm^{-3}), (e) and (f) the bandgap energy E_g (eV), (g) and (h) the number of layers of MoS_2, (i) and (j) the electron affinity χ (eV), and (k) and (l) the dielectric constant ε_r. The p-cSi parameters were maintained as: thickness = 100 μm, $N_A = 1 \times 10^{16}$ cm^{-3}, $N_C/N_V = 3 \times 10^{19}$ cm^{-3} and $\chi = 4.05$ eV, and the graphene parameters were maintained as: number of layers = 5, reflectance = 0.092 and work function = 4.31 eV. The other parameters are given in Tables 1 and 2.

Open Access Article. Published on 23 June 2020. Downloaded on 6/24/2020 4:25:19 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

from 3×10^{18} cm^{-3} to 10^{20} cm^{-3}. The increment in V_{OC} can be understood by considering the equation: $V_{OC} = \frac{kT}{q} \ln\left(\frac{np}{{n_i}^2}\right)$ in which n_i is the carrier concentration, n is the intrinsic electron concentration and p represents the intrinsic hole concentration. Here, $n = n_o + \Delta n$ and $p = p_o + \Delta p$, in which n_o and p_o are the concentrations of the electrons and holes in the equilibrium condition and Δn, Δp are the excess photo-generated electron and hole concentrations, respectively. As, ${n_o}{p_o} = {n_i}^2$, $n = N_D$ and $\Delta n = \Delta p$, V_{OC} can therefore be rewritten as $V_{OC} = \frac{kT}{q} \ln\left(1 + \frac{\Delta p}{N_D} + \frac{\Delta p^2}{{n_i}^2} + \frac{\Delta p N_D}{{n_i}^2}\right)$.[57] We can assume that V_{OC} increased with N_C/N_V owing to an increase of Δp and Δn. As shown in Fig. 2d, the V_{OC} decreased a little bit to 666.4 mV as the N_C/N_V increased to 9×10^{20} cm^{-3}, which is due to the decrement of eV_{bi} according to eqn (1). The J_{SC} was found to remain constant at 32.33 mA cm^{-2} with an increase of N_C/N_V from 3×10^{18} to 9×10^{20} cm^{-3}. Although the generation rate (G) of the photo-generated carriers is reported to increase with higher values of N_C/N_V.[58] The constant J_{SC} indicates that the photo-generated carriers either recombine or are trapped close to the metal/semiconductor interface. As with the V_{OC}, the FF is also found to increase and a maximum value of 82.33% was obtained at $N_C/N_V = 3 \times 10^{20}$ cm^{-3}. Therefore, η was found to be enhanced from 15.82% to a maximum value of 17.73%.

After optimization of the N_D and N_C/N_V to values of 10^{18} cm^{-3} and 3×10^{20} cm^{-3}, respectively we varied the bandgap (E_g) in the range 1.41–1.47 eV (ref. 39) and results are shown in Fig. 2e and f. As we can see, all of the cell parameters are constant and a value of $\eta = 17.74\%$ was obtained in this range of E_g.

For the simulation with different numbers of layers of n-MoS_2 as shown in Fig. 2g and h, the V_{OC} was found to remain constant at a value of 664.8 mV, whereas the J_{SC} reduced from 33.63 to 32.33 mA cm^{-2}, when the number of layers was increased from 3 to 6. This happens because the increase in the number of layers reduces the transmittance of photons to reach the junction, which results in a reduction of the electron–hole pair generation rate. The reduction of the transmittance with the increase in the n-MoS_2 layer number leads to a reduction of the FF. The η was found to decrease from 18.54% to 17.74% as the number of layers increased from 3 to 6. Hence, we used 3L as the optimized number of layers of MoS_2 for further simulation.

Fig. 2i and j show the cell parameters for the optimization of χ in the ranges 3.7 to 4.1 eV. Here we have considered E_g = 1.47 eV for 3L MoS_2.[42] The V_{OC} remained constant at a value of 664.8 mV when χ (eV) was increased from 3.7 to 4 eV. Upon a further increase in the χ to 4.1 eV, a small increase in the V_{OC} to approximately 691.4 eV was obtained owing to an increase in the eV_{bi}. A constant J_{SC} of 33.63 mA cm^{-2} was found in the range 3.74–3.90 eV of χ and this was enhanced to 34.16 mA cm^{-2} for χ = 4.1 eV. A small increase in the ΔE_C (see Fig. 1b) was observed with the increase in χ, which leads to a slightly higher V_{OC} and J_{SC} at $\chi = 4.1$ eV. The FF was observed to increase from 82.93% (@ $\chi = 3.74$ eV) to 83.41% (@ $\chi = 4.1$ eV) as a result of the reduction in the sheet resistance. Therefore, η was found to increase from 18.54% to 19.19%.

It has been found from electric transportation measurements that the dielectric response of the MoS_2 layers is sensitive to the substrate that is used and the dielectric effect relies on the thickness of the sample.[43] The effect of the dielectric constant (ε_r) on the cell parameters is depicted in Fig. 2k and l. It can be observed in Fig. 2l, that the V_{OC} does not depend on the variation of ε_r and it was found to be a constant at 691.4 eV. However, the J_{SC} shows a light decrement from 34.83 to 34.39 mA cm^{-2} with an increase of ε_r up to 14. In a p–n junction, $eV_{bi} \propto \frac{1}{\varepsilon_r}$, which affect the relative permittivity of the MoS_2 layer.[54,57] The generation of the e–h pair increases in the MoS_2 layer with an increase in the ε_r following Coulombs law as $F \propto \frac{1}{\varepsilon_r}$, in which F is the electrostatic force between the electron and holes, but at the same time the increase in ε_r reduces eV_{bi}. As a consequence of both of these effects, we obtained a constant V_{OC}. In contrast, under short circuit conditions, the photo-generated carriers experienced reduced eV_{bi}, which boosts the recombination of the carriers. Thus, the J_{SC} was found to decrease. The FF showed a minor increase from 82.43% to 83.41% when ε_r was increased from 4 to 6. After this the increase in FF slowed down and only reached 83.49% for $\varepsilon_r = 14$. In Fig. 2k, η was found to largely depend on the V_{OC} and remained constant at 19.91% with ε_r. The cell parameters of the proposed cell were found to be independent of the electron mobility (μ_n) of the n-MoS_2 layer, which has been varied in the range of 517–50 cm^2 V^{-1} s^{-1}.

3.2 Effect of the a-Si:H(i) layer thickness

In the HIT solar cell, an intrinsic layer of a-Si:H is inserted, mainly for passivation of the c-Si surface in the cell.[51] Moreover, the insertion of the thin intrinsic layer also contributes to the depletion region, as well as to the drift current and electric field strength.[59] Fig. 3a and b shows the dependence of the cell parameters on the thickness variation of the a-Si:H(i) layer is in the 3 to 10 nm range. It can be observed in Fig. 3b that along with the increase of a-Si:H(i) thickness from 3 to 4 nm, the V_{OC} was slightly enhanced from 691.4 to 693 mV owing to a minor

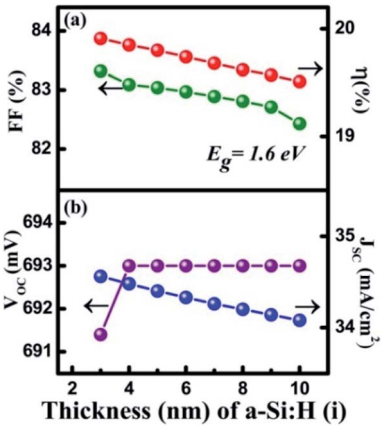


Fig. 3 Optimization of the a-Si:H(i) layer: (a) and (b) thickness (nm) in which E_g was maintained at 1.6 eV.

Open Access Article. Published on 23 June 2020. Downloaded on 6/24/2020 4:25:19 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

increase in the barrier height at the a-Si:H(i) layer and p-cSi interface.[43] The V_{OC} then became constant at 693 mV upon further increasing the thickness of a-Si:H(i). The J_{SC} was found to decrease linearly from 34.56 to 34.08 mA cm^{-2}. It is reported that a thicker intrinsic layer increases the series resistance in the cell.[50] Therefore, we observed a tiny change in the J_{SC}. Thus, the FF was found to reduce from 83.32% to 82.43%. Consequently, the η obtained reduced from 19.91% to 19.51%. To observe the cell performance for the optimized bandgap energy of a-Si:H(i), a further simulation was carried out by varying the E_g a-Si:H(i) in the range 1.6 to 2 eV. However, the maximum η = 19.19% was only obtained at E_g = 1.6 eV. This might be because the absorption of the photons by the MoS_2 layer only corresponds to this E_g of the a-Si:H(i) layer.

3.3 Optimization of the p-cSi layer

After optimizing the n-MoS_2 and a-Si:H(i) parameters to their optimal values, the simulation was further performed to optimize the parameters of the p-cSi layer, while keeping the parameters of n-MoS_2 and a-Si:H(i) at their optimal values, as discussed in the previous section. The results of the cell performance against variation in the N_A variation of p-cSi are shown in Fig. 4a and b. It is observed in Fig. 4b that when the N_A is increased from 10^{15} to 10^{17} cm^{-3}, the V_{OC} also increased from 691.2 to 699.2 mV owing to the higher eV_{bi} with N_A according to eqn (1). Similarly, J_{SC} was also found to increase from 34.23 to 39.15 mA cm^{-2} as the increased eV_b pushes the photo-generated carriers effectively from the depletion region into the contact region without experiencing a recombination effect or the charge carrier trapped effect in the neutral region. This resulted in an increment in the J_{SC}.[57] The increment in the FF from 81.48% to 82.76% can be attributed to the reduction of the sheet resistance as J_{SC} is also increased. These changes lead to an improvement in η from 19.46% to 21.88%. Hence, $N_A = 10^{17}$ cm^{-3} with a maximum η of 21.88% is considered to be the optimal value for the c-Si layer.

To understand the effect of N_C and N_V on the performance of the HIT solar cell, we varied these parameters within the range of 3 × 10^{19} to 1 × 10^{21} cm^{-3} for the p-cSi layer. It can be observed from Fig. 4c and d, with the increase in the N_C/N_V, the V_{OC} decreased linearly from 699.2 to 560.2 mV. According to eqn (1), with the increase in the N_V of p-cSi layer the eV_{bi} decreases and therefore V_{OC} also decreases. In contrast, the value of J_{SC} shows a small increase from 37.81 to 37.91 mA cm^{-2} with N_C/N_V. It has also been reported that N_C/N_V influences the bandgap as well as the absorption of photons.[58] It can be estimated that for N_C/N_V > 3 × 10^{19} cm^{-3}, the high absorption of photons in the active silicon layer increased the G of electron–

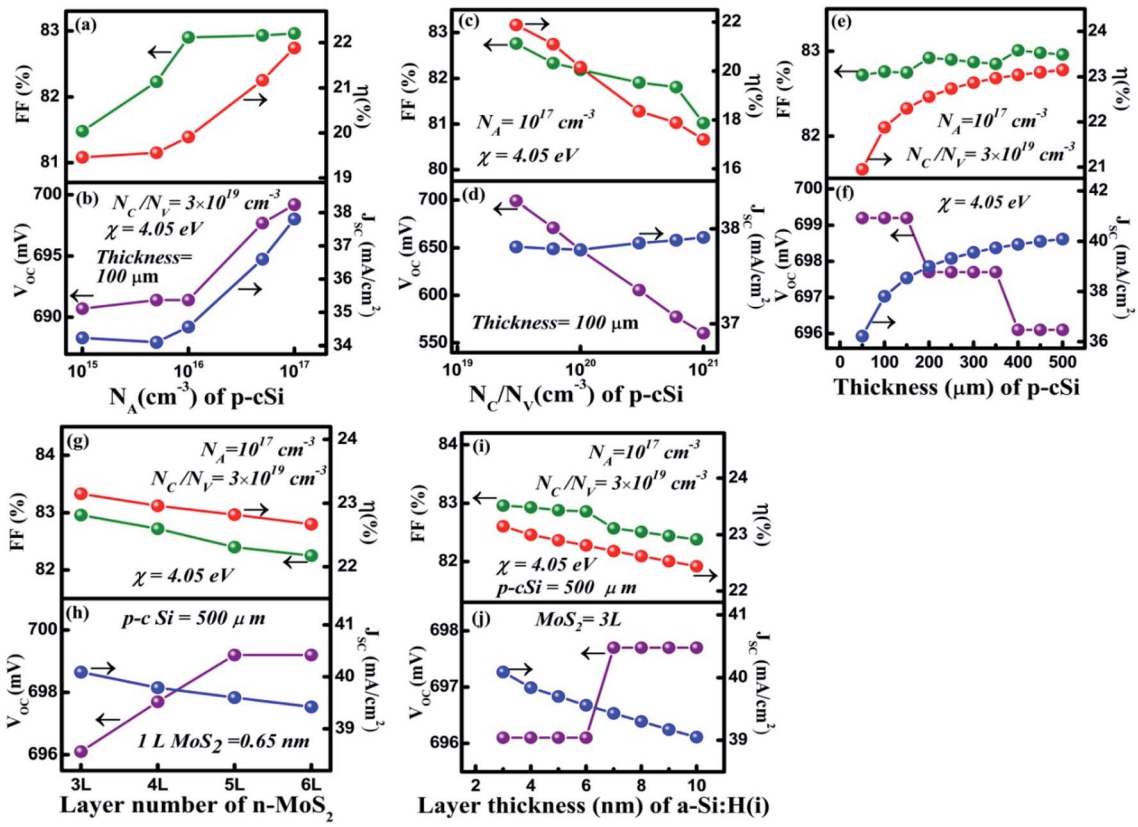


Fig. 4 Optimization of the p-cSi layer: (a) and (b) acceptor concentration N_A (cm^{-3}), (c) and (d) effective conduction band and valance band density N_C/N_V (cm^{-3}), (e) and (f) thickness (μm), (g) and (h) number of layers of MoS_2 and (i) and (j) layer thickness of a-Si:H(i). The optimized parameters of a-Si:H(i) were maintained as: thickness = 3 nm, E_g = 1.6 eV, and the parameters for n-MoS_2 were maintained as: $N_D = 10^{18}$ cm^{-3}, N_C/N_V = 3 × 10^{20} cm^{-3}, number of layers = 3, E_g = 1.47 eV, χ = 4.1 eV and ε_r = 6.

Open Access Article. Published on 23 June 2020. Downloaded on 6/24/2020 4:25:19 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

hole pairs. In Fig. 4d, the FF showed a deterioration from 82.76% to 81.02%, which resulted in the η dropping to 17.2% from 21.88%. The maximum value of $\eta = 21.88\%$ was obtained at $N_C/N_V = 3 \times 10^{19}$ cm^{-3}.

In solar cell technology, the thickness of the p-cSi wafer plays an important role in the fabrication of HIT solar cells. Fig. 4e and f represents the variation of cell parameters as a function of the p-cSi wafer thickness from 50 to 500 μm. In Fig. 4f, the V_{OC} remains constant at 699.2 mV. Upon a further increase in the thickness, the V_{OC} reduced sharply to 697.7 mV and again remained constant up to a thickness of 350 μm. Above this thickness, the V_{OC} was found to deteriorate to the value 696.1 mV and becomes independent of the change of thickness. This variation may be due to the variation in the recombination of the photo-generated carriers under open circuit conditions. The observed enhancement in the J_{SC} from 36.22 to 40.09 mA cm^{-2} can be ascribed to the boost in the generation of the charge carriers with the increase in the p-cSi thickness under short circuit conditions. The FF was observed to increase slowly from 82.72% to 82.96% owing to the reduction in the series resistance. These processes help to increase the η from 20.95% to 23.15% at a thickness of 500 μm.

To study the role of the number of layers of n-MoS_2 and the a-Si:H(i) layer thickness with the best optimized p-cSi parameters, the proposed cell was again simulated by varying the number of layers of MoS_2 and the thickness of a-Si:H and the results are depicted in Fig. 4g, h and i, j respectively. From Fig. 4g and h, it can be clearly seen that the V_{OC} increased from 696.1 to 699.2 mV when the number of layers of n-MoS_2 was increased from 3 to 5. Furthermore, the increase in the number of layers to 6V_{OC} remains constant at 699.2 mV. On the other hand, the J_{SC}, FF, and η were observed to decrease with the increase in the number of layers. The J_{SC} decreased from 40.09% to 39.42%, the FF from 82.96% to 82.25 and the η from 23.15% to 22.67%. In Fig. 4i and j, as the layer thickness of a-Si:H(i) was changed to 6 from 3 nm, the V_{OC} gave a constant value of 696.1 mV and increased slightly to 697.7 mV when the thickness was increased to 6 nm. The V_{OC} then became independent of the

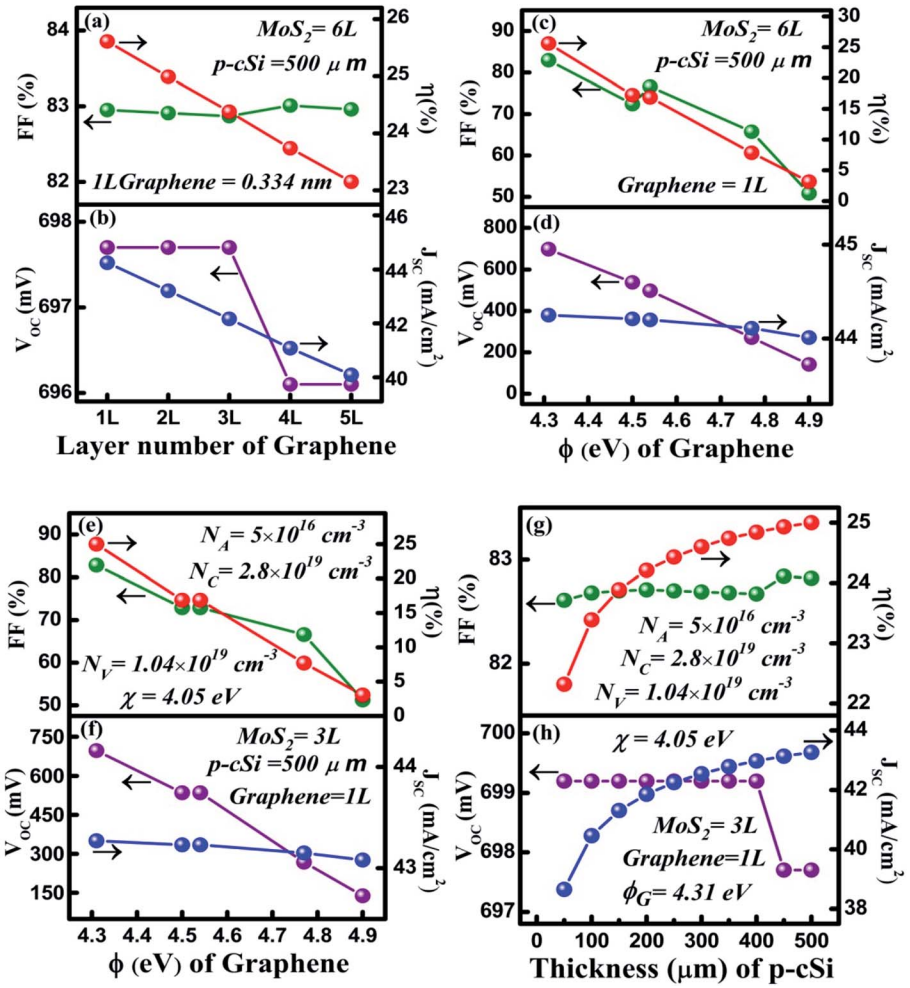


Fig. 5 Optimization of the graphene (front contact) layer: (a) and (b) optimization of the graphene layer number, (c) and (d) optimization of the Φ (eV) of the graphene layer in which the parameters of n-MoS_2 and p-cSi are maintained at their best-optimized values, (e) and (f) optimization of the Φ (eV) of the graphene layer and (g) and (h) optimization of the thickness of a commercially available silicon wafer with practically available p-cSi parameters.

Open Access Article. Published on 23 June 2020. Downloaded on 6/24/2020 4:25:19 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

thickness even when it was increased to 10 nm. The J_{SC} was found to decline from 40.09 to 39.05 mA cm^{-2} with the thickness of a-Si:H(i). The FF was found to decrease from 82.96% to 82.38% and a minimum η of 22.14% was achieved at 10 nm.

3.4 Simulation and optimization of the front contact (graphene) parameters

In a solar cell, contact plays a transparent conducting electrode role in the solar cell by allowing the light to transmit through and maintain a high electrical conductivity. Fig. 5a and b shows the dependence of the HIT solar cell parameters on the variation of the graphene layer (front contact) number. It is observed in Fig. 5b that with the increase in the number of graphene layers from 1 to 3, the V_{OC} remained constant at 697.7 mV. The reason behind this consistency might be the decrease in the recombination rate which limits the reduction of the charge carrier generation under open-circuit voltage conditions. However, upon further increasing of the graphene layer number to 4, the absorbed light might not be sufficient to produce an efficient electron–hole pair, which causes the V_{OC} to slightly decrease to 696.5 mV and then become constant with an increase in the layer number. On the other hand, the J_{SC} was observed to decline linearly from 44.25 to 40.09 mA cm^{-2}. When the number of layers of graphene was increased the optical transmittance reduced and thus the G of the charge carriers decreased and therefore the J_{SC} reduced too. Under short circuit conditions, the photo-generated carriers could be trapped close to the graphene/n-MoS_2 junction. Therefore, we have observed a linear decrease in the J_{SC}, which also, in turn, reduced the η from 25.61% to 23.15%. An insignificant change in the FF was found from 82.95% to 82.96%.

The presence of a strong covalent bond between the C and Mo atoms or the presence of a dangling bond of sulfur present in the interfaces affects the interfacial barrier height and a variation in the barrier height results in a variation in the contact resistance.[60,61] Usually, the contact resistance is a vital factor that affects the solar cell performance.[62] However, many experimental reports suggest that by controlling the Fermi level (*i.e.* the work function) of graphene, the contact resistance can be minimized in the graphene/MoS_2 interfaces.[63,64] Therefore, in this work, we have focused primarily on the work function of graphene. The depicted results shown in Fig. 5c and d, show the variation of the cell parameters with the variation of the graphene work function (Φ_{Gr}) from 4.31 to 4.9 eV. A large reduction in the V_{OC} was observed in Fig. 5d, in which it decreased from 697.7 to 141.4 mV with an increase of Φ_{Gr}. From eqn (2), it can be observed that with the increase in Φ_{Gr}, the potential barrier energy eV_{bi} of graphene/n-MoS_2 decreases. It has also been reported that when the Φ_{TCO} (contact) of Si HIT solar is low, the eV_{bi} of the TCO (transparent conducting oxide)/n-type emitter has the same direction as that of the eV_{bi} of the n-emitter/p-Si. On the other hand, when the Φ_{TCO} (contact) is high, the eV_{bi} of the TCO/n-type emitter contact will have an inverted direction to that of the eV_{bi} of the n-emitter/p-Si junction. Moreover, with the increase in the Φ of the TCO contact in the silicon HIT solar cell, the depletion region in the emitter region increases and widens and results in the overlap of the contact/n-type emitter region and the n-type emitter/p-type silicon junction region if the emitter is not sufficiently thin enough.[4] Thus, the higher Φ_{Gr} might increase the possibility of overlapping between the graphene/n-MoS_2 depletion region and the n-MoS_2/p-cSi depletion region, which has resulted in a reduction in the V_{OC}. In contrast, owing to the overlap of the graphene/n-MoS_2 depletion region and the n-MoS_2/p-cSi depletion region, the eV_{bi} of n-MoS_2/p-cSi reduces. This results in a reduction of the built-in electric field which is unable to push the photo-generated carriers effectively to their respective sides under short circuit conditions and thus the J_{SC} is observed to decline linearly from 44.25 to 40.09 mA cm^{-2} with an increase in the Φ_{Gr}. The overall reduction in the V_{OC} and J_{SC} reduced η from 25.61% to 23.15%. The maximum η = 25.61% was achieved at Φ_{Gr} = 4.31 eV of pristine graphene.

To validate our simulated results, a comparison has been performed by considering the best-optimized values of n-MoS_2 and a-Si:H(i) with the practically available p-cSi parameters by varying the Φ_{Gr} and p-cSi thickness within the range of 4.31–4.9 eV and the 50–500 μm range respectively, as depicted in Fig. 5e–h and Table 4, using the practically available data for p-cSi.[52,65] It can be seen from Fig. 5e and f, that the maximum η =

Table 4 Summary of the most optimized cell

Cell parameter	Most optimized cell (before optimization of the parameters of the graphene layers)	Graphene layer number optimized cell (after optimization of n-MoS_2 and p-cSi)	n-cSi wafer optimized cell for practically available silicon parameters
V_{OC} (mV)	696.1	697.7	697.7
J_{SC} (mA cm^{-2})	40.09	44.25	43.27
FF (%)	82.96	82.95	82.82
η (%)	23.15 [@ 3L n-MoS_2 at $N_D = 10^{18}$ cm^{-3}, $N_C/N_V = 3 \times 10^{20}$ cm^{-3}, E_g = 1.47 eV, χ = 4.47 eV, ε_r = 6, μ_n = 517 cm^2 V^{-1} s^{-1}, μ_p = 8.5 cm^2 V^{-1} s^{-1}, and @ 6 nm a-Si:H(i) at E_g = 1.6 eV and @ 500 μm p-cSi at $N_A = 1 \times 10^{17}$ cm^{-3}, $N_C/N_V = 3 \times 10^{19}$ cm^{-3}, χ = 4.05 eV and @ 5L graphene at Φ = 4.31 eV]	25.16 [@ 3L n-MoS_2 at $N_D = 10^{18}$ cm^{-3}, $N_C/N_V = 3 \times 10^{20}$ cm^{-3}, E_g = 1.47 eV, χ = 4.47 eV, ε_r = 6, μ_n = 517 cm^2 V^{-1} s^{-1}, μ_p = 8.5 cm^2 V^{-1} s^{-1}, and @ 6 nm a-Si:H(i) at E_g = 1.6 eV and @ 500 μm p-cSi at $N_A = 1 \times 10^{17}$ cm^{-3}, $N_C/N_V = 3 \times 10^{19}$ cm^{-3}, χ = 4.05 eV and @ 1L graphene at Φ = 4.31 eV]	25 [@ 3L n-MoS_2 at $N_D = 10^{18}$ cm^{-3}, $N_C/N_V = 3 \times 10^{20}$ cm^{-3}, E_g = 1.47 eV, χ = 4.47 eV, ε_r = 6, μ_n = 517 cm^2 V^{-1} s^{-1}, μ_p = 8.5 cm^2 V^{-1} s^{-1}, and @ 6 nm a-Si:H(i) at E_g = 1.6 eV and @ 500 μm p-cSi at $N_A = 5 \times 10^{16}$ cm^{-3}, $N_C/N_V = 2.8 \times 10^{19}/1.04 \times 10^{19}$ cm^{-3}, χ = 4.05 eV and @ 1L graphene at Φ = 4.31 eV]

Open Access Article. Published on 23 June 2020. Downloaded on 6/24/2020 4:25:19 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

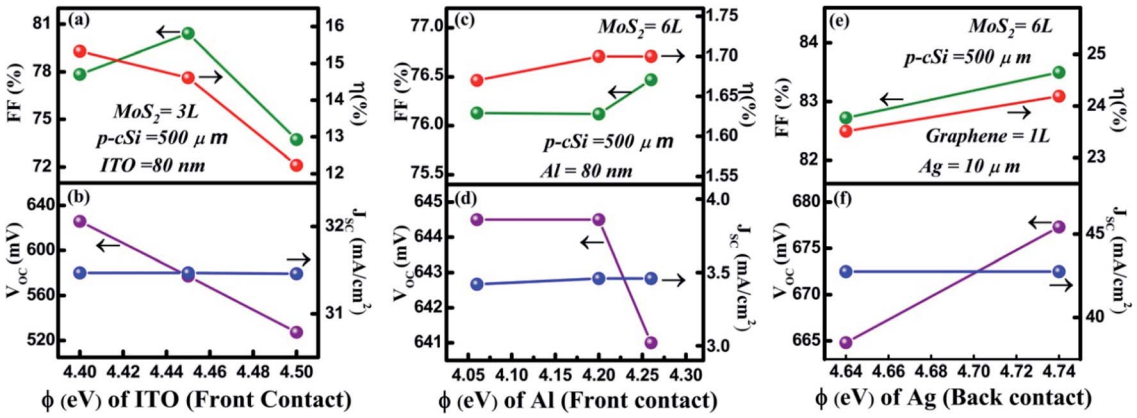


Fig. 6 Changing the front and back contact: (a) and (b) variation in the Φ (eV) of ITO, (c) and (d) variation in the Φ (eV) of Al, and (e) and (f) variation in the Φ (eV) of Ag.

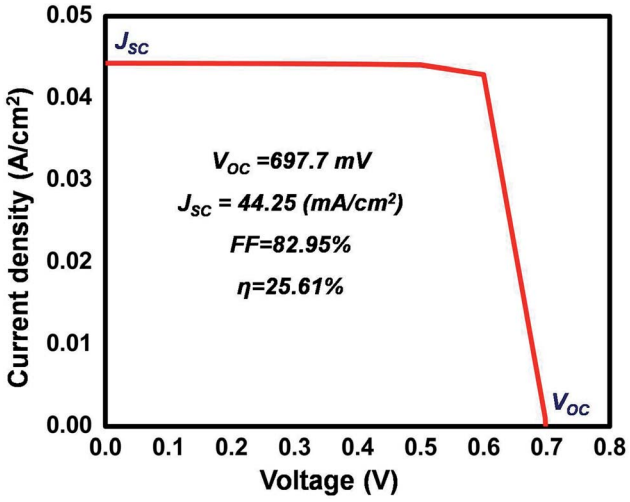


Fig. 7 *IV* characteristics of the most optimized cell.

25% was obtained at $\Phi_{Gr} = 4.31$ eV, which is close to our best-optimized results as shown in Fig. 5c and d. Similarly, in Fig. 5g and h, the same maximum $\eta = 25\%$ was achieved for a 500 μm thick p-cSi wafer.

3.5 Variation and simulation of the front contact and back contact material

In this section, a comparison of the cell parameters has been performed by changing the front contact material graphene to ITO and Al[19] and the back contact to Ag. As seen in Fig. 6a and b the thickness of ITO is considered to be 80 nm and the Φ_{ITO} has been varied from 4.40–4.50 eV.[34] The maximum values for V_{OC}, J_{SC}, FF and η were obtained and found to be 625.8 mV, 31.47 mA cm^{-2}, 80.41% and 15.33% respectively. Similarly, in Fig. 6c and d, when Φ_{Al} was varied, the maximum values for V_{OC}, J_{SC}, FF, and η were found to be 650.8 mV, 4.36 mA cm^{-2}, 82.79% and 2.353%, respectively. In both cases, η was significantly lower than that of η obtained using the graphene layer as a contact.

Fig. 6e and f shows the variation in the cell parameters when the back contact was changed from Au to Ag and the Φ of Ag was maintained from 4.64 to 4.74 eV. It was observed that in this range of Φ for Ag, the V_{OC} increased from 664.8 to 677.3 mV, whereas the J_{SC} remained constant at 42.76 mA cm^{-2}. The FF and η were observed to decrease from 82.72% to 83.5% and 23.51% to 24.19%, respectively. This η was close to our simulated η_{max} value, that is, 25.61% when we used Au as a contact. Here, the back contacts have been considered as the uniform thin film deposited on the backside of p-cSi.

3.6 maximum power and series resistance

The proposed solar cell as shown in Fig. 1a was simulated and characterized under a standard AM1.5G spectrum and forward voltage in the range of 0 to 0.7 V. The *IV* characteristics of the best optimized solar cell generated by the software are shown in Fig. 7. The maximum power delivered by the cell, as well as the

Table 5 Graphene and MoS_2 silicon-based solar cells

Type of study	Solar cell structure	V_{OC} (V)	J_{SC} (mA cm^{-2})	FF (%)	η (%)	Ref.
Experimental	ITO/graphene/MoS_2/n-cSi/(Ti/Ag)	0.59	36.8	73	15.8	66
Experimental	Au/graphene/MoS_2/n-Si/In	0.56	33.4	60	11.1	67
Experimental	Au/graphene/MoS_2/n-Si/Au	0.50	28.1	47	6.56	68
Experimental	Pd/n-MoS_2/i-SiO_2/p-Si/In	0.30	5.5	42	4.5	69
Experimental	Ni/graphene/MoS_2/p-csi/Al	0.51	—	—	2.58	70
Theoretical	TCO/MoS_2/SiGe:H/p-Si/Al	0.652	40.01	83.7	21.85	3
Theoretical	Graphene/n-MoS_2/a-Si:H/p-cSi/Au	0.697	44.25	82.95	25.61	This work

Open Access Article. Published on 23 June 2020. Downloaded on 6/24/2020 4:25:19 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

series resistance (R_s), was calculated from the *IV* characteristics and found to be 25.4 mW cm^{-2} and 2.5 Ω cm^2 was achieved respectively (see ESI†).

Finally, a comparison between the experimental and theoretical demonstration of graphene, MoS_2, and Si-based solar cells was performed and the results are detailed in Table 5. However, except for our work none of the studies used graphene as a contact material or TCE.

4. Conclusions

In order to develop an efficient HIT solar cell, a graphene/n-MoS_2/a-Si:H(i)/p-cSi/Au structure has been proposed and simulated using AFORS-HET v 2.5 software. The individual effects of the six parameters of the n-MoS_2 layer were analyzed individually. After optimizing the parameters of the n-MoS_2, a-Si:H(i), and p-cSi layers, an optimum efficiency of 23.15% was achieved. As the contact material in a solar cell has an adverse effect on the cell performance, a further simulation was performed to observe the effect of graphene on the proposed HIT solar cell by varying the number of layers and the work function. From this simulation, a maximum efficiency of 25.61% was achieved for single-layer graphene at a work function of 4.1 eV. On the other hand, by changing the front contact graphene to ITO and Al we achieved efficiencies of 15.33% and 2.35%. This confirms that graphene could be an efficient contact material for use in an n-MoS_2/a-Si:H(i)/p-cSi/Au HIT solar cell. Our simulated results establish that the graphene layer could be an effective contact TCE material and n-MoS_2 could be an efficient emitter layer in a single facial n-MoS_2 and p-cSi based HIT solar cell.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors are very thankful to Helmholtz-Zentrum Berlin for providing us with the AFORS-HET simulation software and the Department of Science and Technology (DST), Ministry of India for their financial support.

References

1 N. Dwivedi, S. Kumar, A. Bisht, K. Patel and S. Sudhakar, *Sol. Energy*, 2013, **88**, 31–41.
2 M. Belarbi, M. Beghdad and A. Mekemeche, *Sol. Energy*, 2016, **127**, 206–215.
3 S. Xu, X. Zeng, W. Wang, G. Zhou, Y. Hu and S. Wu, *J. Miner. Mater. Char. Eng.*, 2017, **5**, 323–338.
4 L. Zhao, C. L. Zhou, H. L. Li, H. W. Diao and W. J. Wang, *Sol. Energy Mater. Sol. Cells*, 2008, **92**, 673–681.
5 K. Masuko, M. Shigematsu, T. Hashiguchi, D. Fujishima, M. Kai and N. Yoshimura, *IEEE J. Photovolt.*, 2014, **4**, 1433–1435.
6 I. Khrapach, F. Withes, T. H. Bointon, D. K. Polyushkin, W. L. Barnes, S. Russo and M. F. Craciun, *Adv. Mater.*, 2012, **24**, 2488–2849.
7 P. Lazar, R. Zboril, M. Pumera and M. Otyepka, *Phys. Chem. Chem. Phys.*, 2014, **16**, 14231–14235.
8 J. H. Kim, J. H. Hwang, J. Suh, S. Tongay, S. Kwon, C. C. Hwang, J. Wu and J. Y. Park, *Appl. Phys. Lett.*, 2013, **103**, 171604.
9 A. S. Mayorov, R. V. Gorbachev, S. V. Morozov, L. Britnell, R. Jalil, L. A. Ponomarenko, P. Blake, K. S. Novoselov, K. Watanbe, T. Taniguchi and A. K. Geim, *Nano Lett.*, 2011, **11**, 2396–2399.
10 X. Zhong, W. Zhou, Y. Peng, Y. Zhou, F. Zhou, Y. Yin and D. Tang, *RSC Adv.*, 2015, **5**, 45239.
11 B. Radisavljevic, A. Radenovic, J. Brivio, V. Giacometti and A. Kis, *Nat. Nanotechnol.*, 2011, **6**, 147.
12 B. Radisavljevic and A. Kis, *Nat. Mater.*, 2013, **12**, 815.
13 D. S. Tsai, K. K. Liu, D. H. Lien, M. L. Tsai, C. F. Kang and C. A. Lin, *ACS Nano*, 2013, **7**, 3095–3911.
14 X. Yang, W. Fu, W. Liu, J. Hong, Y. Cai, C. Jin, M. Xu, H. Wang, D. Yang and H. Chen, *J. Mater. Chem. A*, 2014, **2**, 7727.
15 J. M. Yun, Y. J. Noh, J. S. Yeo, Y. J. Go, S. I. Na, H. G. Jeong, J. Kim, S. Lee, S. S. Kim, H. Y. Kim and D. Y. Kim, *J. Mater. Chem. C*, 2013, **1**, 3777.
16 S. Choi, Z. Shaolin and W. Yang, *J. Korean Phys. Soc.*, 2014, **64**, 1550–1555.
17 J. H. Li, D. Lee, D. Qu, X. Liu, J. Ryu and A. Seabaugh, *Nat. Commun.*, 2015, **6**, 6564.
18 Y. Du, H. Liu, A. T. Neal, M. Si and P. D. Ye, *IEEE Electron Device Lett.*, 2013, **34**, 1328–1330.
19 M. L. Tsai, S. H. Su, J. K. Chang, D. S. Tsai, C. H. Chen and C. I. Wu, *ACS Nano*, 2014, **8**, 8317–8322.
20 L. Hao, Y. Liu, W. Gao, Z. Han, Q. Xue and H. Zeng, *J. Appl. Phys.*, 2015, **117**, 114502.
21 D. D. Tune, N. Mallik, H. Fornasier and B. S. Flavel, *Adv. Energy Mater.*, 2020, **10**, 1903261.
22 M. Bat-Erdene, M. Batmunkh, S. A. Tawfik, M. Fronzi, M. J. Ford, C. J. Shearer, L. Yu, M. Dadkhah, J. R. Gascooke, C. T. Gibson and J. G. Shapter, *Adv. Funct. Mater.*, 2017, 1704488.
23 L. Yu, A. S. R. Bati, T. S. L. Grace, M. Batmunkh and J. G. Shapter, *Adv. Energy Mater.*, 2019, 1901063.
24 S. Lin, P. Wang, X. Li, Z. Wu, Z. Xu and S. Zhang, *Appl. Phys. Lett.*, 2015, **107**, 153904.
25 Z. Qiao, J. L. Ji and Y. L. Zhang, *Chin. Phys. B*, 2017, **26**, 068802.
26 L. Fan, F. Wang, J. Liang, X. Yao, J. Fang and D. Zhang, *J. Semicond.*, 2017, **38**, 014003.
27 H. Keum, H. J. Chung and S. Kim, *ACS Appl. Mater. Interfaces*, 2013, **5**, 6061–6065.
28 *Introduction to Quantum Transport in Low-Dimensional Systems*, http://www.hri.res.in/∼nanotr16/notes/KJThomas-1.pdf, accessed July 2018.
29 S. D. Sarma, S. Adam, E. H. Hwang and E. Rossi, *Rev. Mod. Phys.*, 2011, **83**, 407.

 This journal is © The Royal Society of Chemistry 2020

Open Access Article. Published on 23 June 2020. Downloaded on 6/24/2020 4:25:19 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

30 J. Suh, T. E. Park, D. Y. Lin, D. Fu, J. Park and H. J. Jung, *Nano Lett.*, 2014, **14**, 6976–6982.
31 Z. Lin, B. R. Carvalho, E. Kahn, R. Lv and R. Rao, *2D Matter.*, 2016, **3**, 022002.
32 A. I. Pantazi and V. Yannopapas, *Opt. Commun.*, 2016, **374**, 24–28.
33 K. Patel and P. K. Tyagi, *Carbon*, 2017, **116**, 744–752.
34 Y. Park, V. Choong, Y. Gao, B. R. Hsieh and C. W. Tang, *Appl. Phys. Lett.*, 1996, **68**, 2699–2701.
35 A. Kahn, Fermi level, work function and vacuum level, *Mater. Horiz.*, 2016, **3**, 7–10.
36 R. M. Eastment and C. H. B. Mee, *J. Phys. F: Met. Phys.*, 1973, **3**, 1738–1745.
37 A. W. Dweydari and C. H. B. Mee, *Phys. Status Solidi A*, 1975, **27**, 223–230.
38 Z. H. Ni, H. M. Wang, J. Kasim, H. M. Fan, T. Yu and Y. H. Wu, *Nano Lett.*, 2007, **7**, 2758–2763.
39 F. J. Nelson, V. K. Kamineni, T. Zhang, E. S. Comfort, J. U. Lee and A. C. Diebold, *Appl. Phys. Lett.*, 2010, **97**, 253110.
40 C. P. Lu, G. Li, J. Mao, L. M. Wang and E. Y. Andrei, *Nano Lett.*, 2014, **14**, 4628.
41 B. Radisavljevic and A. Kis, *Nat. Mater.*, 2013, **12**, 815–820.
42 K. F. Mak, C. Lee, J. Hone, J. Shan and T. Y. Heinz, *Phys. Rev. Lett.*, 2010, **105**, 136805.
43 E. J. G. Santos and E. Kaxiras, *ACS Nano*, 2013, **7**, 10741–10746.
44 S. L. Howell, D. Jariwala, C. C. Wu, K. S. Chen, V. K. Sangwan and J. Kang, *Nano Lett.*, 2015, **15**, 2278–2284.
45 O. Salehzadesh, N. H. Tran, X. Liu, I. Shihand and Z. Mi, *Nano Lett.*, 2014, **14**, 125–4130.
46 H. Zhang, Y. Ma, Y. Wan, X. Rong, Z. Xie and W. Wang, *Sci. Rep.*, 2015, **5**, 8440.
47 W. Zhou, X. Zou, S. Naijmaei, Z. Liu, Y. Shi and J. Kong, *Nano Lett.*, 2013, **13**, 2615–2622.
48 X. Garros, G. Reimbold, J. Cluzel, D. Muñoz and P.-J. Ribeyron, *Microelectron. Eng.*, 2011, **88**, 1247–1250.
49 L. Qin, Y. Xiao-Jun, L. Cheng and C. Ming-bo, *Optoelectron. Lett.*, 2010, **6**, 0108–0111.
50 W. Lisheng and C. Fengxiang, *J. Optoelectron. Adv. Mater.*, 2011, **13**, 81–88.
51 A. Rawat, M. Sharma, D. Chaudhary, S. Sudhakar and S. Kumar, *Sol. Energy*, 2014, **110**, 691–703.
52 L. Zhao, H. L. Li, C. L. Zhou, H. W. Diao and W. J. Wang, *Sol. Energy*, 2009, **83**, 812–816.
53 F. Rana, Semiconductor Heterostructure, in *Semiconductor Optoelectronics*, 2017, ch. 2, https://courses.cit.cornell.edu/ece533/Lectures/handout2.pdf.
54 J. Singh, *Semiconductor devices: Basic Principles*, Wiley India (P) Ltd, New Delhi, 2004.
55 V. K. Thakur and M. K. Thakur, *Chemical Functionalization of Carbon Nanomaterials: Chemistry and Applications*, CRC Press, New York, 2016, p. 918.
56 A. V. Sachenko, V. P. Kostylyov, M. V. Gerasymenko, R. M. Korkishko, M. R. Kulish and M. I. Slipchenko, *Quantum Electronics & Optoelectronics*, 2016, **19**, 67–74.
57 Y. Zhao, C. Liang, M. Sun, Q. Liu, F. Zhang and D. Li, *J. Appl. Phys.*, 2014, **116**, 154506.
58 Y. Zhou and G. Long, *J. Phys. Chem. C*, 2017, **121**, 1455–1462.
59 W. Lisheng, C. Fengxiang and A. Yu, *J. Phys.: Conf. Ser.*, 2011, **276**, 012177.
60 N. Jensen, R. M. Hausne, R. B. Bergmann, J. H. Werner and U. Rau, *Prog. Photovoltaics Res. Appl.*, 2002, **10**, 1–13.
61 X. Liu, J. Gao, G. Zhang and Y. W. Zhang, *Nano Res.*, 2017, **10**, 2944–2953.
62 M. Houssa, K. Iordanidou, A. Dabral, A. Lu, R. Meng, G. Pourtois, V. V. Afanas'ev and A. Stesmans, *Appl. Phys. Lett.*, 2019, **114**, 163101.
63 D. K. Schroder and D. L. Meier, *IEEE Trans. Electron Devices*, 1984, **31**(5), 637–647.
64 Y.-J. Yu, Y. Zhao, S. Ryu, L. E. Brus, K. S. Kim and P. Kim, *Nano Lett.*, 2009, **9**, 3430–3434.
65 H. Wang, L. Yu, Y.-H. Lee, Y. Shi, A. Hsu, M. L. Chin, L.-J. Li, M. Dubey, J. Kong and T. Palacios, *Nano Lett.*, 2012, **12**, 4674–4680.
66 J. Ma, H. Bai, W. Zhao, Y. Yuan and K. Zhang, *Sol. Energy*, 2018, **160**, 76–84.
67 Y. Tsuboi, F. Wang, D. Kozawa, K. Funahashi, S. Mouri, Y. Miyauchi, T. Takenobub and K. Matsuda, *Nanoscale*, 2015, **7**, 14476–14482.
68 K. Jiao, C. Duan, X. Wu, J. Chen, Y. Wang and Y. Chen, *Phys. Chem. Chem. Phys.*, 2015, **17**, 8182.
69 L. Z. Hao, W. Gao, Y. J. Liu, Z. D. Han, Q. Z. Xue, W. Y. Guo, J. Zhu and Y. R. Li, *Nanoscale*, 2015, **7**, 8304–8308.
70 W. Shi and M. Xiying, *Coatings*, 2018, **8**, 2.

Received: 28 October 2019 | Revised: 23 February 2020 | Accepted: 21 March 2020

DOI: 10.1002/ctpp.201900175

Check for updates

Contributions to Plasma Physics

ORIGINAL ARTICLE

Theoretical analysis for transmission of Gaussian and sine time irradiance of electromagnetic beam in collisional dusty plasmas

Ruchi Sharma | **Suresh C. Sharma**

Department of Applied Physics, Delhi Technological University (DTU), Shahbad Daulatpur, Bawana Road, Delhi, India

Correspondence
Suresh C. Sharma Department of Applied Physics, Delhi Technological University (DTU), Shahbad Daulatpur, Bawana Road, Delhi 110042, India.
Email: suresh321sharma@gmail.com

Abstract

A theoretical model of propagation of Gaussian and Sine time irradiance of an electromagnetic beam in collisional dusty plasma has been done in the present analysis. It contains equilibrium of dust charge, particle density, and energy of plasma ingredients having charge neutrality. Ionization of neutral particles, recombination of free electrons with ions, adsorption and emission of electrons from dust grain surface, and binary collisions between plasma components are also considered in this treatment. Time varying behaviour of modified electron temperature and collision frequency has been illustrated numerically as a function of dust densities. Also, the comparative analyses of variation of beam waist parameter with the dimensionless length of transmission for both the Gaussian and Sine time irradiance are involved in this model as a function of distinguishable time width, collision frequencies, and dust densities under the condition that the size of dust nebulous is greater than the electrons mean free path for the adsorption on the dust grain surface. The observed results are significant for the applications in industry and astrophysics.

KEYWORDS

dusty plasma, electromagnetic beam, time irradiance, transmission

1 | INTRODUCTION

Several research[1–11] studies have been performed to show the significant effect of dust grains in plasma because of its exciting applications[12–18] in space, industries, and laboratories plasmas. Particularly, the self-focusing behaviour of an electromagnetic beam propagating in dusty plasma becomes an interesting domain of research promoted by the various investigators.[19–25] Mishra et al.[26] have studied the transportation of complex plasma featured by Mathis, Rumpl, and Nordsieck (MRN) power law grain size distribution with and without electric field. They observed the larger plasma transparency with larger dust grains in dark plasma while the case is reversed in illuminated plasma where the photoemission is predominant. Sharma and Sharma[27] have discussed the kinetics of nonuniform transmission of a Gaussian electromagnetic beam in strongly ionized dusty plasma considering the elastic and charging collisions with photoemission and electrons adsorption on the boundary of dust grains. Jia et al.[28] used the Boltzmann distribution law to study the transportation characteristics of an electromagnetic wave propagating in weakly and strongly ionized dusty plasma. Mishra et al.[29] have used the paraxial ray approximation to study the propagation behaviour of a Gaussian irradiance

 © 2020 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

of an electromagnetic beam in dusty plasma under the effect of elastic and inelastic collisions. Wang et al.[30] have studied the electromagnetic wave propagation characteristics of inhomogeneous dusty plasma. They developed the models of double exponential and hyperbolic electron number density distribution for dusty plasma and also studied the power transmission and reflection coefficient on the basis of their analytical study. Sodha et al.[31] also analysed the cases of thermionic emission, photoemission, and emissionless processes to formulate the problem of fluctuation of charge on account of presence of dust grains.

Several investigations on temporal profile of electromagnetic wave propagating in dust-free plasma have been carried out by various researchers[32–34] but in this paper; analytical model is evolved to access the transmission kinetics of both Gaussian and Sine time irradiance of electromagnetic wave owing to mutual interactions of dust and other plasma components. It is familiar that the electrons in dusty plasma get heated up because of highly irradiated electromagnetic wave to the stationary state temperature of the order of relaxation time $\tau' = 1/(\delta_{comp}\nu_{ef})$, where δ_{comp} is the fractional loss of energy of electrons due to collisions and ν_{ef} is the effective collision frequency under the influence of dust grains. Consequently, with the increasing electron temperature, electron density attains a steady state value during the relaxation time τ'. For the pulse duration $(t_0) < \tau'$, $n_e/n_{e0} = 1$ and for $t_0 > \tau'$, n_e/n_{e0} attains a steady value. Hence, the electron density is considered as a function of temperature. In this communication, for $t_0 > \tau'$, the steady-state electron density and temperature in dusty plasma is obtained on account of ionization of neutrals, recombination of free electrons with ions, net adsorption on dust grain surface, and diffusion due to space charge field and temperature gradient which are ignored in other analytical studies.[19, 22] Power loss due to ohmic heating and collisions between the constituent species have also been taken into account to predict the electron temperature, which causes the nonlinearity. Ohmic heating of ions has been ignored due to their heavier mass. The significance of thermal conduction has been neglected in the present formalism.

For the computation, metallic dust cloud of stainless steel having work function of 7.8 eV as in ionosphere has been considered so that photoemission can take place for spectrum of Lyman α radiations (1215.7 Å) in the extreme UV region.

The left over manuscript is schemed as follows: In Section 2, the coupled sets of differential equations of charge neutrality, electron density, and electron temperature are solved considering the paraxial ray approximation to structure the analytical model for formulating the modified dielectric permittivity under the influence of dust grains. Section 3 includes the solution of propagation of electromagnetic beam for the both Gaussian and Sine time irradiance in dusty plasma following the Akhmanov's et al.[35] and its extension by Sodha et al.[36] in respect of modification of the beam waist parameter f with the dimensionless length of transmission ζ for distinguishable dust parameters. Consequences of the existing model are discussed in Section 4, and, in the end, we reach to the conclusions in Section 5.

2 | ANALYTICAL MODEL

Assume dusty plasma to be made up of dust grain spheres with charge $-Z_d e$ and radius r_d, ions, electrons, and neutral atoms, where Z_d is the state of charge on dust and e is the charge on an electron. The kinetics of dusty plasma include ambipolar diffusion, neutral ionization, neutralization of the ions, adsorption and photoemission of electrons, collisions between plasma constituents, and Ohmic heating caused by nonuniform irradiance of the beam.

The following equations demonstrate the kinetics of dusty plasma as:

2.1 | Ambipolar diffusion term

The ambipolar diffusion term due to density decay of constituent species is given by

$$\prod_e = -n_e\mu_e U - D_{te} n_e(\nabla T_e/T_e) - D_{ne}\nabla n_e, \tag{1a}$$

$$\prod_i = n_i\mu_i U - D_{ti} n_i(\nabla T_i/T_i) - D_{ni}\nabla n_i, \tag{1b}$$

$$\prod_d = -n_d\mu_d U - D_{nd}\nabla n_d, \tag{1c}$$

where Π_g, n_g, U, T_g, $D_{tg}(=5k_BT_g/2m_g\nu_{efg})$, $D_{ng}(=k_BT_g/m_g\nu_{efg})$, $\mu_g(=eD_{ng}/k_BT_g)$, are the flux, number density, space charge field, temperature, thermal diffusion coefficient, diffusion coefficients, and mobility; the subscript g stands for electrons, ions, and dust grains, respectively.

The net rate of diffusion of both the charges per unit area must be balanced in stationary state, so we get

$$\prod_i = \prod_e + \prod_d. \tag{2a}$$

The mobility and diffusion of dust particles are supposed to be zero because of their heavy mass and hence their flux comes out to be zero. Thus, the Equation (2a) gives

$$\prod_i = \prod_e. \tag{2b}$$

Using the Equations (1a), (1b), and (2b), eliminating U, net particle flux or density decay comes out to be

$$\prod_p = -D_{te} n_e (\nabla T_e/T_e) - D_{ne} \nabla n_e - n_e \mu_e \left(\frac{D_{ni} \nabla n_i - D_{ne} \nabla n_e + D_{ti} n_i (\nabla T_i/T_i) - D_{te} n_e (\nabla T_e/T_e)}{n_e \mu_e + n_i \mu_i} \right), \tag{3a}$$

$$\prod_p = - \left(\frac{D_{ni}[c_2 n_e \nabla n_i + c_1 n_i \nabla n_e] + D_{ti} n_e n_i [c_2 (\nabla T_i/T_i) + c_3 (\nabla T_e/T_e)]}{n_e c_2 + n_i} \right), \tag{3b}$$

where $c_1 = D_{ne}/D_{ni}$, $c_2 = \mu_e/\mu_i$ and $c_3 = D_{te}/D_{ti}$.

In this article, axial parameters of diffusion coefficients and mobility are taken into consideration.

2.2 | Charge neutrality

Neutrality of charge may be forwarded as

$$Z_d n_d + n_i - n_e = 0, \tag{4}$$

where dust grains number density is represented by n_d, and n_i and n_e are the ionic and electronic densities under the influence of dust grains, respectively.

2.3 | Dust charge balance

The net charge on dust grains is balanced by charge flux through the surfaces of dust grains, thus

$$\frac{dZ_d}{dt} = n_{ph}(Z_d, T_d) + n_{ic}(Z_d, T_i) - \Upsilon n_{ec}(Z_d, T_e), \tag{5}$$

where $n_{ph}(= \pi r_d^2 n_p)$ is the rate of photoelectric emission by surfaces of charged dust grains, where[37]

$$n_p = \Lambda(\nu)\chi(\nu) = \Lambda(\nu)(4\pi m_e k_B^2 T_d^2/\mathrm{h}^3)\mu(\nu)\Phi_{\mathrm{ph}}(\eta),$$

$\Phi_{ph}(\eta)$ is tabulated integral(Fowler 1936) stated as

$$\Phi_{ph}(\eta) = \int_0^{\exp(\eta)} \frac{\ln(1+\Omega)}{\Omega} d\Omega,$$

$$\eta = (\mathrm{h}\nu - \phi_0)/k_B T_d.$$

n_p indicates the rate of photon density emitted from surface of neutral or negatively charged dust grains, $\Lambda(\nu)$ represent the photon density incident normally to the dust surface, $\chi(\nu)$ is photoelectric efficiency. h represents the Planck's constant, ν indicates the frequency of incident beam, ϕ_0 shows the work function of material of dust grain, T_d represents the temperature of dust grains considered as constant due to large thermal capacity, $\mu(\nu)$ is the probability of absorption of a photon by an electron striking the surface from inside.[38] $n_{ic} = \pi r_d^2 (8k_B T_i/\pi m_i)^{1/2} n_i [1 - Z_d \alpha_i]$ and $n_{ec} = \pi r_d^2 (8k_B T_e/\pi m_e)^{1/2} n_e \exp[Z_d \alpha_e]$ are the ionic and electronic flux charging the dust grain surface, respectively,[29]

$\alpha_e = (e^2/r_d k_B T_e)$, $\alpha_i = (e^2/r_d k_B T_i)$, Υ indicates the adsorption coefficient of electrons on surface of dust grains, r_d is the dust grain radius, k_B is the Boltzmann's constant, T_i and T_e are the ions and electrons temperature, m_i and m_e represent the masses of ions and electrons, respectively.

2.4 | Neutral atoms, electrons, and ion number balance

$$n_i + n_0 = n_{i0} + n_{00} = \mathrm{N}_t, \tag{6}$$

where n_0 is neutral particle density under the influence of dust grains, n_{00}, n_{i0} are the densities of neutral atoms and ions in plasma not having dust grains, and N_t is an invariable, respectively.

$$\frac{dn_e}{dt} = (\beta_i n_0 - \alpha_r n_e n_i) - n_d(\Upsilon n_{ec} - n_{ph}) - \nabla \prod_{\mathrm{p}}, \tag{7}$$

and

$$\frac{dn_i}{dt} = (\beta_i n_0 - \alpha_r n_e n_i) - n_d n_{ic} - \nabla \prod_{\mathrm{p}}, \tag{8}$$

where β_i represent the coefficient of ionization, $\alpha_r = (4.5 \times 10^{-7}(250/T_e)^{-0.7})\ \mathrm{cm^3 s^{-1}}$ is the recombination coefficient of electrons with ions corresponding to O^+ present in ionosphere[1] upto 250 km height, the dominant ionic composition is O^+. Equations (7) and (8) represent the rate of growth of number densities of electrons and ions in plasma under the influence of dust grains, respectively. The initial terms on RHS of Equations (7) and (8) indicate the gain of electrons and ions per unit time per unit volume on account of ionization of neutrals. The succeeding pair of terms in both the equations corresponds to number of neutral atoms formed due to recombination of free electrons into ions, and the net adsorption of electrons and ions on the surface of dust grains.

2.5 | Electrons and ions energy balance

$$\frac{d}{dt}\left(\frac{3}{2} n_e k_B T_e\right) = \beta_i n_0 \varepsilon_e - \alpha_r n_e n_i \left(\frac{3}{2} k_B T_e\right) - n_d(n_{ec}\varepsilon_{ec} - n_{ph}\varepsilon_{ph}) + \frac{n_e e^2 \nu_{ef} U_0^2}{4 m_e(\omega_d^2 + \nu_{ef}^2)} - \nu_{em}\delta_{em}\frac{3}{2} n_e k_B (T_e - T_0) + \nu_{ei}\delta_{ei}\frac{3}{2} n_e k_B (T_e - T_i) \tag{9}$$

and

$$\frac{d}{dt}\left(\frac{3}{2} n_i k_B T_i\right) = \beta_i n_0 \varepsilon_i - \alpha_r n_e n_i \left(\frac{3}{2} k_B T_i\right) - n_d n_{ic}\varepsilon_{ic} - \nu_{in}\delta_{in}\frac{3}{2} n_i k_B (T_i - T_0) + \nu_{ei}\delta_{ei}\frac{3}{2} n_e k_B (T_e - T_i) \tag{10}$$

where ε_e and ε_i are the mean energy of electrons and ions on account of ionization of neutral atoms, $\varepsilon_{ec}(=2k_B T_e - Z_d(e^2/r_d))$ and $\varepsilon_{ic}(=(2 - Z_d\alpha_i)/(1 - Z_d\alpha_i))$ indicate the average energy of adsorption of electrons and ions on the surface of dust grains, $\varepsilon_{ph} = \left(\frac{k_B T_d}{\Theta(\eta - \zeta)} \int_0^\infty 2\zeta \ \ln(1 + \exp(\eta - \zeta))d\zeta\right) - Z_d\alpha_d$ is related to photoemission energy, where Θ $(=Z_d e/r_d)$ is the charged dust grain surface potential, $\alpha_d = (e^2/r_d k_B T_d)$, $\nu_{ef}(=\nu_{ei} + \nu_{ed} + \nu_{en} + \nu_{re} + \nu_{ac})$ is the modified collision frequency owing to ions, dust grains, neutral atoms elastic collisions with electrons and inelastic collisions by virtue of recombination of electrons into ions, and adsorption of electrons on the surface of dust grains, given by $\nu^{[1]}{}_{ei} = \nu_{ei0}(n_i/n_{e0})(T_e/T_{e0})^{-3/2}$, $\nu^{[23]}{}_{\mathrm{ed}} = \nu_{\mathrm{ed0}}(\mathrm{T_e}/\mathrm{T_{e0}})^{-3/2}\mathrm{Z_d^2}$, $\nu^{[1]}{}_{\mathrm{en}} = \nu_{\mathrm{en0}}(\mathrm{n_0}/\mathrm{n_{00}})(\mathrm{T_e}/\mathrm{T_{e0}})^{1/2}$, $\nu_{\mathrm{in}} = \nu_{\mathrm{in0}}(\mathrm{n_0}/\mathrm{n_{00}})(\mathrm{T_i} + \mathrm{T}_d/\mathrm{T_{i0}} + \mathrm{T}_d)^{1/2}$, $\nu_{\mathrm{re}}(=\alpha_{\mathrm{r}}\mathrm{n_i})^{26}$, $\nu_{\mathrm{ac}} = \mathrm{n_d}(\mathrm{n_{ec}}/\mathrm{n_e})$, where $\nu_{ei0} = (5.5 n_{e0}/T_{e0}^{3/2})\ln(220 T_{e0}/n_{e0}^{1/3})$, $\nu_{ed0} = (2.9 \times 10^{-6}) n_{d0} T_{e0}^{-3/2} \ln\Lambda_e$ [$\ln\Lambda_e$(coulomb logarithm) = 10–13], $\nu_{en0} =$

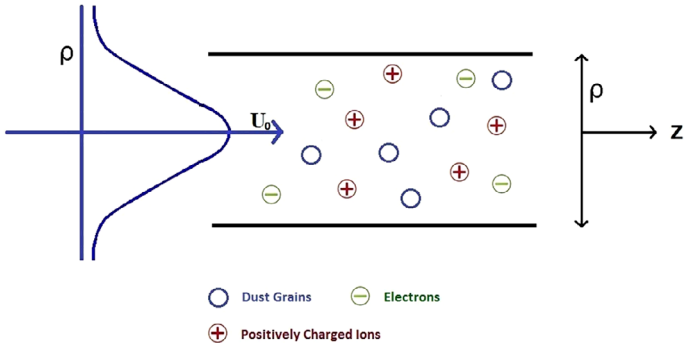


FIGURE 1 Scheme of irradiance of Gaussian electromagnetic beam in dusty plasma

$(8.3\times10^5)\sigma_0 n_{00}T_{e0}^{1/2}$, $\nu_{in0} = (16/3)(2K_BT_d/\pi)^{1/2}\sigma_0 n_{00}(T_{i0}+T_d)^{1/2}$ are the electrons-ions/dust and ion-neutral particles collisions frequencies without dust grains, where n_{e0} is the number density for electrons, T_{e0} is the temperature of electrons without dust grains. $\sigma_0(=6.6\times10^{-15}[(T_e/4)-0.1/\{1+(T_e/4)^{1.6}\}])$ is the electrons and neutral particles collisional cross-section area, where σ_0 is in cm^2 and T_e is in eV. The initial pair of terms of Equations (9) and (10) correspond to average energy gain on account of ionization and recombination processes. The third term in both the equations refers to net energy loss due to adsorption process, and the succeeding term in Equation (9) belongs to power loss due to process of photoemission from the outer layer of dust grains. The fifth term in Equation (9) corresponds to electrons energy gain from the field due to Ohmic heating effect. Ohmic heating of heavier ions can be ignored. The last term in the two describes the rate of energy loss per unit volume attributed to collisions between dusty plasma components. The term regarding thermal conduction can be ignored on account of predominant collisional effect.

2.6 | Evaluation of ionization coefficient and average energy of electrons and ions

By applying electron density and energy balance equations in dust-free plasma, ionization coefficient and average energies of electrons and ions can be evaluated. The initial condition for them are given as

$$\beta_i n_{00} = \alpha_r n_{e0} n_{i0} = \alpha_r n_{e0}^2, \tag{11}$$

$$\varepsilon_e = \frac{d}{dt}\left(\frac{3}{2}k_BT_{e0}\right) + (\alpha_{r0}n_{e0})^{-1}\left[\nu_{en}\delta_{en}\frac{3}{2}k_B(T_{e0}-T_0) + \nu_{ei0}\delta_{ei}\frac{3}{2}k_B(T_{e0}-T_{i0})\right], \tag{12a}$$

and

$$\varepsilon_i = \frac{d}{dt}\left(\frac{3}{2}k_BT_{i0}\right) + (\alpha_{r0}n_{e0})^{-1}\left[\nu_{in}\delta_{in}\frac{3}{2}k_B(T_{i0}-T_0) + \nu_{ei0}\delta_{ei}\frac{3}{2}k_B(T_{e0}-T_{i0})\right]. \tag{12b}$$

2.7 | Paraxial approximation approach

Let us suppose that the electromagnetic beams with Gaussian and Sine time irradiance spreading along their wave front. Cylindrical coordinate system is suitable to study for their propagation, which may be considered along the z-axis. Propagation of Gaussian irradiance is shown in Figure 1.

The spatial irradiance of the beam along z = 0 is described by

$$U^2 = U_0^2(t)e^{(-\rho^2/\rho_0^2)}\ , \tag{13}$$

where U^2 is the electric field intensity of the Gaussian beam at a length ρ from axis, U_0^2 is the axial intensity at $\rho = 0$, whereρ is the radial coordinate of cylindrical system. Major analysis of the theory assumes the time irradiance of Gaussian function of the form.

$$U_0^2(t) = U_{00}^2F'(t/t_0), \tag{13a}$$

where t_0 is the beam time width.

$$F'(t/t_0) = \begin{cases} Sin(\pi - t/t_0) & [0 < t/t_0 < \pi] \quad \text{for Sine beam} \\ 0, \text{orelse} \end{cases}, \tag{13b}$$

$$F'(t/t_0) = \exp(-t^2/t_0^2) \ [-\infty < t/t_0 < \infty] \text{ for Gaussian beam} \tag{13c}$$

For a defined value of z, the irradiance in paraxial approximation is represented as

$$U_0^2 = \frac{U_{00}^2(z,t)}{f^2(z,t)} e^{(-\rho^2/\rho_0^2 f^2(z,t))} F'(t/t_0), \tag{14}$$

where $f(z, t)$ is the beam waist parameter, which depends upon the length of transmission, z and time t, and $\rho_0 f$ represent the spot of the beam.

From above statement, it is suggested that the major intensity of beam survives along its principal axis. Therefore, Gaussian beam can be analysed by the paraxial approximation approach i.e., $\rho/\rho_0 f \ll 1$. So, all the applicable variables are expanded along the principal part ($\rho = 0$) only up to the quadratic expression ρ^2/ρ_0^2 as

$$P(\rho, z, t) = P_a(z, t) - (\rho^2/\rho_0^2) P_\rho(z, t), \tag{15}$$

where P_a and P_ρ are the parts of all the accordant parameters characterized by P(ρ,z), along axis and radius, respectively. Although the spatial irradiance of the Gaussian beams do not modify during its propagation in dusty plasma, by introducing the accordant parameters and the dimensionless variable $t/t_0 = \tau$ in kinetic Equations (5), (7)–(10), and comparing the coefficients of ρ^0 and ρ^2/ρ_0^2 terms, we can write the above equations as

2.8 | Dust charge balance

$$\frac{dZ_{da}}{d\tau} = t_0[n_{ica}(Z_d, T_i) + n_{pha}(Z_d, T) - \Upsilon n_{eca}(Z_d, T_e)] \tag{5a}$$

and

$$\frac{dZ_{d\rho}}{d\tau} = t_0[n_{ic\rho}(Z_d, T_i) + n_{ph\rho}(Z_d, T) - \Upsilon n_{ec\rho}(Z_d, T_e)]. \tag{5b}$$

2.9 | Electrons and ions number balance

$$\begin{aligned} \frac{dn_{ea}}{d\tau} &= t_0[\ \beta_{ia} n_0 - \alpha_{ra} n_{ea} n_{ia} - n_d(\Upsilon n_{eca} - n_{pha})] \\ &\quad - \frac{(2/\rho_0^2)\, t_0}{n_{ia} + c_2 n_{ea}} [D_{1i}(c_1 n_{ia} n_{e\rho} + c_2 n_{ea} n_{i\rho}) + D_{2i} n_{ea} n_{ia}(c_2 T_{ia}^{-1}\ T_{i\rho} + c_3 T_{ea}^{-1}\ T_{e\rho})] \end{aligned} \tag{6a}$$

and

$$\begin{aligned} \frac{dn_{e\rho}}{d\tau} &= t_0[\beta_{i\rho} n_0 - \alpha_{r\rho} n_{ea} n_{ia} - n_d(\Upsilon n_{ec\rho} - n_{pha})] \\ &\quad - \frac{(6/\rho_0^2) t_0}{n_{ia} + c_2 n_{ea}} \begin{bmatrix} D_{1i}(c_1 + c_2) n_{i\rho} n_{e\rho} - \frac{(c_1 n_{ia} n_{e\rho} + c_2 n_{ea} n_{i\rho})(n_{i\rho} + c_2 n_{e\rho})}{n_{ia} + c_2 n_{ea}} \\ + D_{2i}\left[(n_{ea} n_{i\rho} + n_{ia} n_{e\rho})(c_2 T_{ia}^{-1}\ T_{i\rho} + c_3 T_{ea}^{-1}\ T_{e\rho}) \right. \\ \left. - n_{ea} n_{ia}(c_2 T_{ia}^{-2}\ T_{i'\rho}^{2} + c_3 T_{ea}^{-2}\ T_{e\rho}^{2}) \right. \\ \left. - \frac{n_{na} n_{ia}(c_2 T_{ia}^{-1}\ T_{i\rho} + c_3 T_{ea}^{-1}\ T_{e\rho})(n_{i\rho} + c_2 n_{e\rho})}{n_{ia} + c_2 n_{ea}} \right] \end{bmatrix}. \end{aligned} \tag{6b}$$

$$\frac{dn_{ia}}{d\tau} = t_0[\ \beta_{ia}n_0 - \alpha_{ra}n_{ea}n_{ia} - n_d n_{ica}] - \frac{(2/\rho_0^2)}{n_{ia}+c_2n_{ea}}[D_{1i}(c_1 n_{ia}n_{e\rho} + c_2 n_{ea}n_{i\rho}) + D_{2i}n_{ea}n_{ia}(c_2 T_{ia}^{-1}\ T_{i\rho} + c_3 T_{ea}^{-1}\ T_{e\rho})] \quad (7a)$$

and

$$\frac{dn_{i\rho}}{d\tau} = t_0[\beta_{i\rho}n_0 - \alpha_{r\rho}n_{ea}n_{ia} - n_d n_{ic\rho}] - \frac{(6/\rho_0^2)\ t_0}{n_{ia}+c_2n_{ea}}\begin{bmatrix} D_{1i}(c_1+c_2)n_{i\rho}n_{e\rho} - \frac{(c_1n_{ia}n_{e\rho}+c_2n_{ea}n_{i\rho})(n_{i\rho}+c_2n_{e\rho})}{n_{i'a}+c_2n_{e'a}} \\ +D_{2i}\left[(n_{ea}n_{i\rho}+n_{ia}n_{e\rho})(c_2T_{ia}^{-1}\ T_{i\rho}+c_3T_{ea}^{-1}\ T_{e\rho})\right. \\ -n_{ea}n_{ia}(c_2T_{ia}^{-2}\ T_{i\rho}^{2}+c_3T_{ea}^{-2}\ T_{e\rho}^{2}) \\ \left. -\frac{n_{ea}n_{ia}(c_2T_{ia}^{-1}\ T_{i\rho}+c_3T_{ea}^{-1}\ T_{e\rho})(n_{i\rho}+c_2n_{e\rho})}{n_{ia}+c_2n_{ea}}\right] \end{bmatrix}. \quad (7b)$$

2.10 | Electrons and ions energy balance

$$\begin{aligned}\frac{d}{d\tau}\left(\frac{3}{2}n_{ea}k_BT_{ea}\right) &= t_0\left[\beta_{ia}n_0\varepsilon_e - \alpha_{ra}n_{ea}n_{ia}\left(\frac{3}{2}k_BT_{ea}\right) - n_d(\Upsilon n_{eca}\varepsilon_{eca} - n_{pha}\varepsilon_{pha})\right] \\ &\quad - \delta_{en}\nu_{ena}t_0\frac{3}{2}n_{ea}k_B(T_{ea}-T_0) + \delta_{ei}\nu_{eia}t_0\frac{3}{2}n_{ea}k_B(T_{ea}-T_{ia}) \\ &\quad + \delta_{comp}\nu_{efa}t_0\left(\frac{3}{2}k_BT_0\right)\frac{\beta_0U_{00}^2}{f^2}n_{ea}\end{aligned} \quad (9a)$$

and

$$\begin{aligned}\frac{d}{d\tau}\left(\frac{3}{2}k_B(n_{ea}T_{e\rho}+n_{e\rho}T_{ea})\right) &= t_0\begin{bmatrix}\beta_{i\rho}n_0\varepsilon_e - \frac{3}{2}k_B(\alpha_{r\rho}n_{ea}n_{ia}T_{ea} + \alpha_{ra}(n_{ea}n_{ia}T_{e\rho}+n_{ea}n_{i\rho}T_{ea}+n_{e\rho}n_{ia}T_{ea})) \\ -n_d(\Upsilon(n_{ec\rho}\ \varepsilon_{eca}+n_{eca}\varepsilon_{ec\rho}) - (n_{pha}\ \varepsilon_{ph\rho}+n_{ph\rho}\varepsilon_{pha}))\end{bmatrix} \\ &\quad - \frac{3}{2}k_B\delta_{en}t_0[(n_{e\rho}\nu_{ena}+n_{ea}\nu_{en\rho})(T_{ea}-T_0)+n_{ea}\nu_{ena}T_{e\rho}] \\ &\quad - \frac{3}{2}k_B\delta_{ei}t_0[(n_{e\rho}\nu_{eia}+n_{ea}\nu_{ei\rho})(T_{ea}-T_{ia})+n_{ea}\nu_{eia}(T_{e\rho}-T_{i\rho})] \\ &\quad + \left(\frac{3}{2}\delta_{comp}t_0k_B\ T\right)\frac{\beta_0U_{00}^2}{f^2}\left(\frac{n_{ea}\nu_{efa}}{f^2}+(n_{e\rho}\ \nu_{efa}+n_{ea}\nu_{ef\rho})\right)\end{aligned} \quad (9b)$$

where $\beta_0 = \frac{e^2}{3m_e\delta_{comp}k_BT\omega^2}$ and $\delta_{comp}\ (= \nu_{ac}/\nu_{ef})$[22, 39] is the fractional loss of collisional energy of electron with the dusty plasma components,

$$\begin{aligned}\frac{d}{d\tau}\left(\frac{3}{2}n_{ia}k_BT_{ia}\right) &= t_0\left[\beta_{ia}n_0\varepsilon_i - \alpha_{ra}n_{ea}n_{ia}\left(\frac{3}{2}k_BT_{ia}\right) - \Upsilon n_d n_{ica}\varepsilon_{ica}\right] \\ &\quad - \left[\delta_{in}\nu_{ina}t_0\frac{3}{2}n_{ia}k_B(T_{ia}-T_0) + \delta_{ei}\nu_{eia}t_0\frac{3}{2}n_{ea}k_B(T_{ea}-T_{ia})\right]\end{aligned} \quad (10a)$$

and

$$\begin{aligned}\frac{d}{d\tau}\left(\frac{3}{2}k_B(n_{ia}T_{i\rho}+n_{i\rho}T_{ia})\right) &= t_0\begin{bmatrix}\beta_{i\rho}n_0\varepsilon_i - \frac{3}{2}k_B(\alpha_{r\rho}n_{ea}n_{ia}T_{ia} + \alpha_{ra}(n_{ea}n_{ia}T_{i\rho}+n_{ea}n_{i\rho}T_{ia}+n_{e\rho}n_{ia}T_{ia})) \\ -n_d(\Upsilon(n_{ic\text{æ}}\ \varepsilon_{ica}+n_{ica}\varepsilon_{ic\text{æ}}))\end{bmatrix} \\ &\quad - \frac{3}{2}k_B\delta_{in}t_0[(n_{i\rho}\nu_{ina}+n_{ia}\nu_{in\rho})(T_{ia}-T_0)+n_{ia}\nu_{ina}T_{i\rho}] \\ &\quad - \frac{3}{2}k_B\delta_{ei}t_0[(n_{e\rho}\nu_{eia}+n_{ea}\nu_{ei\rho})(T_{ea}-T_{ia})+n_{ea}\nu_{eia}(T_{e\rho}-T_{i\rho})].\end{aligned} \quad (10b)$$

The simultaneous solution of Equations (5)–(10) for the stationary state (i.e., at $\tau \to \infty$) collectively provide the variation of different dusty plasma components along axis and radius, respectively, for both Gaussian and Sine time irradiance.

2.11 | Modified dielectric permittivity

Since pressure gradient of electrons and ionic gas is observed on account of nonuniform distribution of electrons and ionic temperature due to nonuniform irradiance of electromagnetic beam. Therefore, these gradients are balanced by space charge field in steady state. Thus, one can obtain

$$\frac{n_e}{n_{e0}} = \left(\frac{2T_{e0}}{T_e + T_{e0}}\right). \tag{16}$$

the modified dielectric permittivity in dust accompanying plasma may be given by[35,36,40]

$$\varepsilon_{ef} = \left(\frac{\omega_0^2}{\omega_d^2}\right)\frac{\omega_d^2}{(\nu_{ef}^2 + \omega_d^2)}\left(\frac{n_e}{n_{e0}}\right) = \varepsilon_a - \left(\rho^2/\rho_0^2\right)\varepsilon_\rho, \tag{16a}$$

$$\varepsilon_{ef} = \left(\frac{\omega_0^2}{\omega_d^2}\right)\frac{\omega_d^2}{(\nu_{ef}^2 + \omega_d^2)}\left(\frac{2T_{e0}}{T_e + T_{e0}}\right) = \varepsilon_a - \left(\rho^2/\rho_0^2\right)\varepsilon_\rho. \tag{16b}$$

Substituting T_e and ν_{ef} corresponding to the Equation (15) and comparing the coefficient of ρ^0 and ρ^2/ρ_0^2 terms, one obtain,

$$\varepsilon_a = 1 - \left(\frac{\omega_0^2}{\omega_d^2}\right)\left(1 - \frac{\nu_{efa}^2}{\omega_d^2}\right)\left(\frac{2T_{e0}}{T_{ea} + T_{e0}}\right) \tag{16c}$$

and

$$\varepsilon_\rho = \left(\frac{\omega_0^2}{\omega_d^2 + \nu_{efa}^2}\right)\left(\frac{2T_{e0}}{T_{ea} + T_{e0}}\right)\left(\frac{\nu_{efa}\nu_{ef\rho}}{\omega_d^2 + \nu_{efa}^2} - \frac{T_{e\rho}}{T_{ea} + T_{e0}}\right), \tag{16d}$$

where ε_a, ε_ρ, ν_{efa}, $\nu_{ef\rho}$ and T_{ea}, $T_{e\rho}$ are the modified dielectric permittivities, collision frequencies, and temperatures along axis and radius, respectively, under the influence of dust grains. Plasma frequencies represented by the term ω_0 in the absence of dust grains, whereas ω_d represents the vice versa. On introducing ν_{efa}, $\nu_{ef\rho}$, T_{ea}, *and* $T_{e\rho}$ obtained by the coupled solution of Equations (5)–(10), we get the ε_a and ε_ρ as a function of $\beta_0 U_{00}^2$.

3 | PROPAGATION OF ELECTROMAGNETIC BEAM WITH TIME IRRADIANCE IN DUSTY PLASMA

The irradiance of the Gaussian EM-beam in space and time is given by

$$\vec{U} = \hat{j}\, U_0(\rho, z, t)e^{i\omega_d t}, \tag{17}$$

where z is the direction of propagation and y is the direction of polarization represented by $\hat{j}$. The analogous equation corresponds to Equation (13),

$$\nabla^2\vec{U} - \nabla(\nabla\cdot\vec{U}) + \frac{\varepsilon_{ef}}{c^2}\frac{\partial^2\vec{U}}{\partial t^2} = 0. \tag{18}$$

As the transverse component of a beam has zero divergence, the Equation (18) reduces to

$$\nabla^2 U + (\omega_d^2/c^2)\varepsilon_{ef}U = 0. \tag{19}$$

Equation (19) can be represented in the following manner because of cylindrically symmetric nature of beam

$$\frac{\partial^2 U}{\partial z^2} + \frac{1}{\rho}\frac{\partial U}{\partial \rho} + \frac{\partial^2 U}{\partial \rho^2} = -\frac{\omega_d^2\varepsilon_{ef}}{c^2}U. \tag{20}$$

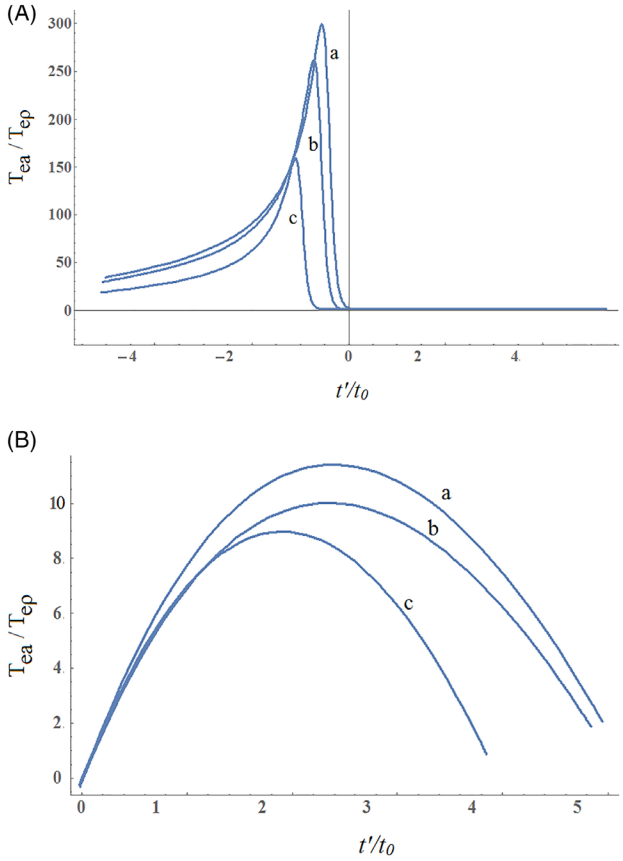


FIGURE 2 (a, b) Dependence of dimensionless electron temperature $T_{ea}/T_{e\rho}$ on dimensionless time width t'/t_0 for distinguishable dust densities for both Gaussian and Sine profile, respectively. The dust densities $n_d = 10, 100, 1000$ in cm^{-3} are labelled by a, b, and c, respectively

The solution of Equation (20) by Akhmanov et al.[35] can be described as

$$U(\rho, z, t) = A_0(\rho, z, t)e^{i(\omega_d t - kz)}, \tag{21}$$

$$A_0(\rho, z, t) = A_{00}(\rho, z, t)e^{-ik\mathbb{S}(\rho,z,t)}, \tag{22}$$

where $k = (\omega_d/c)\,\varepsilon_a^{1/2}$, A_{00} and $\mathbb{S}$ are real quantities, where $\mathbb{S}$ is called eikonal which represents the wave phase.

Measuring the real and imaginary parts of the resulting equations, obtained from putting the values of U and A_0, respectively, from Equations (21) and (22) in Equation (20) and pursuing Sodha et al.,[36] we get

$$2\frac{\partial \mathbb{S}}{\partial z} + \left(\frac{\partial \mathbb{S}}{\partial \rho}\right)^2 + \frac{2}{V_g}\frac{\partial \mathbb{S}}{\partial t} = \frac{1}{\varepsilon_a}\Phi(t) + \frac{1}{k^2 A_{00}}\left\{\frac{\partial^2 A_{00}}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial A_{00}}{\partial \rho}\right\}, \tag{23}$$

and

$$\frac{\partial A_{00}^2}{\partial z} + \frac{\partial \mathbb{S}}{\partial \rho}\frac{\partial A_{00}^2}{\partial \rho} + A_{00}^2\left(\frac{\partial^2 \mathbb{S}}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial \mathbb{S}}{\partial \rho}\right) + \frac{1}{V_g}\frac{\partial A_{00}^2}{\partial t} = 0, \tag{24}$$

where

$$V_g = c{\varepsilon_a}^{1/2}/(\varepsilon_a + \omega_d/2(\partial\varepsilon_a/\partial\omega_d)) \tag{24a}$$

is the group velocity of electromagnetic beam in dusty plasma,
and

$$\Phi(t) = \frac{\varepsilon_\rho e^{-t/\tau'}}{\tau'}\int_{-\infty}^{t} e^{t/\tau'} AA * dt. \tag{24b}$$

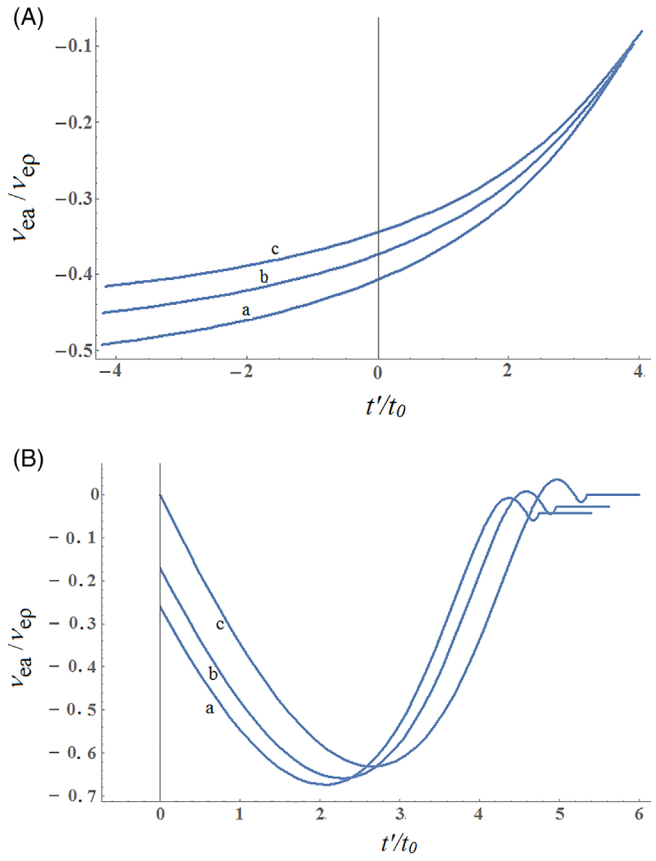


FIGURE 3 (a, b) Dependence of dimensionless electron frequency $\nu_{ea}/\nu_{e\rho}$ on dimensionless time width t'/t_0 for distinguishable dust densities for Gaussian and Sine time profile, respectively. The dust densities, $n_d = 10, 100, 1000$ in cm^{-3} are labelled by a, b, and c, respectively

On transforming the variable as $z' = z$ and $t' = t - z/V_g$, Equations (23) and (24) reduce to the form

$$2\frac{\partial \mathbb{S}}{\partial z'} + \left(\frac{\partial \mathbb{S}}{\partial \rho}\right)^2 = \frac{1}{\varepsilon_a}\Phi(t') + \frac{1}{k^2 A_{00}}\left\{\frac{\partial^2 A_{00}}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial A_{00}}{\partial \rho}\right\}, \tag{25}$$

$$\frac{\partial A_{00}^2}{\partial z'} + \frac{\partial \mathbb{S}}{\partial \rho}\frac{\partial A_{00}^2}{\partial \rho} + A_{00}^2\left(\frac{\partial^2 \mathbb{S}}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial \mathbb{S}}{\partial \rho}\right) = 0. \tag{26}$$

Supporting the earlier observations,[35,40,41] the solution of Equations (25) and (26) can be expressed as

$$A_{00}^2 = \frac{U_{00}^2(t')}{f^2(z', t')} e^{(-\rho^2/\rho_0^2 f^2)} F(t'/t_0), \tag{27}$$

Substituting the expression of $\Phi(t')$ and A_{00}^2 from Equation (24b) and Equation (27), respectively, in Equation (25), we obtain

$$2\frac{\partial \mathbb{S}}{\partial z'} + \left(\frac{\partial \mathbb{S}}{\partial \rho}\right)^2 = \frac{1}{2}\frac{\Phi(t')}{\varepsilon_a} - \frac{2}{k^2\rho_0^2 f^2} + \frac{\rho^2}{k^2\rho_0^4 f^4}. \tag{28}$$

The eikonal function $\mathbb{S}(\rho, z', t')$ in paraxial ray approximation approach, may be expressed as

$$\mathbb{S}(\rho, z', t') = \frac{\rho^2}{2}\mathrm{B}(z', t') + \Psi(z', t'), \tag{29}$$

where $\Psi(z', t')$ is any function of z'.

and t' and $\mathrm{B}(z', t') = \frac{1}{f(z',t')}\frac{\partial f(z',t')}{\partial z'}$.

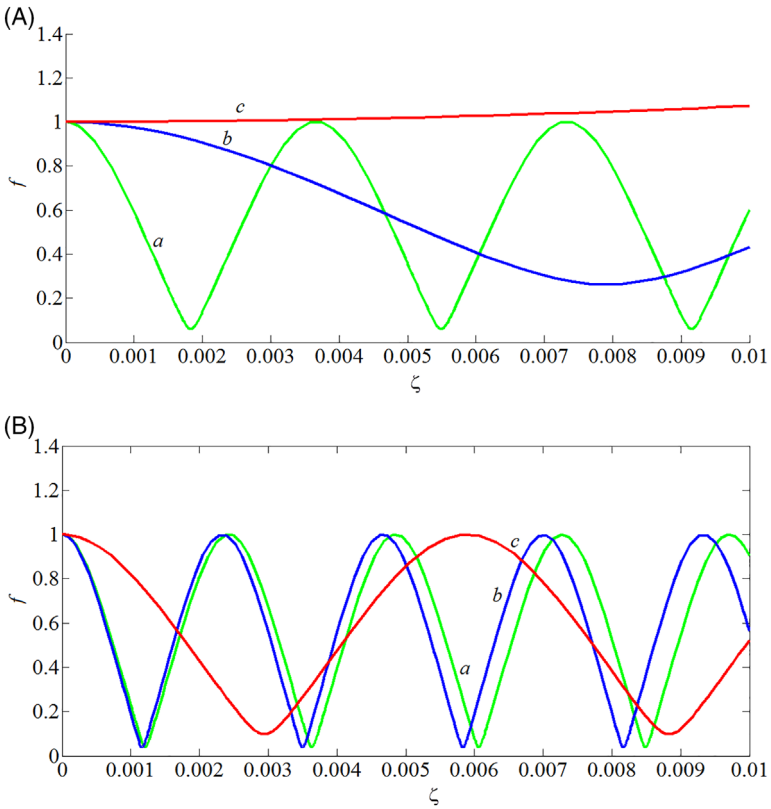


FIGURE 4 (a, b) Relying beam waist parameter f with the dimensionless length of transmission $\zeta(=z'/R_d)$ for the Gaussian and Sine beam for different time width t'/t_0. The labels a, b, and c indicate the dimensionless time width $t'/t_0 = 1$, 2, and 3, respectively

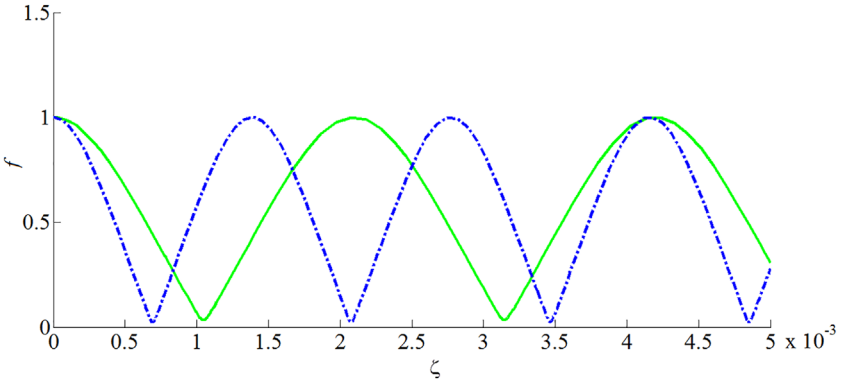


FIGURE 5 Relying beam waist parameter f on the dimensionless length of transmission $\zeta(=z'/R_d)$ for the Gaussian (solid) and Sine (dotted) beam as a function of maximal irradiance at time width $t'/t_0 = -0.5$, for Gaussian and $t'/t_0 = 0.8\pi$ for Sine beam

When beam propagates through dusty plasma, the solution of A_{00}^2 preserves the power and shape of the beam. Employing the expression for B in Equation (29) and substituting $\mathbb{S}$ from Equation (29) in Equation (28) and equating the co-efficient of ρ^2/ρ_0^2 the resulting equation (using the paraxial ray approximation), variation of beam waist parameter with the dimensionless length of transmission can be represented by the equation for both the profiles as

$$\frac{1}{f}\frac{\partial^2 f}{\partial z'^2} = \frac{1}{R_d^2}\frac{1}{f^4} - \frac{1}{2\rho_0^2}\frac{\varepsilon_\rho e^{-t'/\tau'}}{\varepsilon_a}\int_{-\infty}^{t'}\frac{U_{00}^2(t')e^{t'/\tau'}}{f^4}F'(t'/t_0)dt'. \tag{30}$$

Integral sign may be eliminated in the Equation (30), if variation of $F'(t'/t_0)$ is of the order of relaxation time τ, which causes small nonlinearity. Then the Equation (30) can be simplified to

$$\frac{d^2 f}{dz'^2} = \frac{1}{R_d^2}\frac{1}{f^3} - \frac{1}{2\rho_0^2}\frac{\varepsilon_\rho}{\varepsilon_a}\frac{U_{00}^2(t')}{f^3}F'(t'/t_0) \tag{31}$$

or

$$\frac{d^2f}{d\zeta^2} = -\frac{1}{R_{sl}^2}\frac{1}{f^3} + \frac{1}{f^3}, \tag{32}$$

where $R_{sl}\left(=\frac{2\varepsilon_a\rho_0^2}{\varepsilon_\rho U_{00}^2 F'(t'/t_0)}\right)^{1/2}$ is characterized by self-focusing length, $\zeta = \frac{z'}{R_d}$ and $R_d = k\rho_0^2$. On the right-hand side of Equation (32), the initial term corresponds to the nonlinear self-focusing of the beam whereas the succeeding term indicates the divergence phenomenon.

4 | RESULTS AND DISCUSSION

In this analytical model, we present the comparative analysis of both Gaussian and Sine time irradiance of the beam propagating in plasma under the influence of dust grains in collisional plasma. Neutral ionization, recombination of electrons and ions with neutrals, elastic and charging collisions, charging flux via adsorption and photoemission, and ohmic heating due to electromagnetic beam propagating in dusty plasma are the basic energy exchange methods considered in this analysis. Numerical calculations have been carried out in accordance to the metallic dust cloud of stainless steel having high work function of 7.8 eV within close range of space where the photoemission is the dominant mechanism, which corresponds to Lyman alpha radiations with the spectrum of 1215.7 $\mathring{A}$. The simultaneous solution of the coupled Equations (5)–(10) for both the Gaussian and Sine character of electromagnetic beam gives the dependence of modified electron temperature and collision frequency on time for distinguishable dust densities using MATLAB and MATHEMATICA software. Employing the relevant parameter to get ε_a and ε_ρ. Also, at the boundary condition f = 1, $\frac{df}{d\zeta} = 0$ at $\zeta = 0$, we obtain dependence of beam waist parameter upon dimensionless length of transmission, $\zeta(=z'/R_d)$ by substituting ε_a and ε_ρ, for different values of time width, dust densities, and collision frequency for both Gaussian and Sine time irradiance. The set of variables used in this analysis is listed below:$n_{e0} = n_{i0} = 10^6 cm^{-3}$, $n_0 = n_{00} = 2.5\times10^9 cm^{-3}$, $n_d = 10^3 cm^{-3}$, $k_B = 1.38\times10^{-16}\text{ergsk}^{-1}$, $T_{e0} = 1700K$, $m_e = 9.1\times10^{-28}g$, $T_{i0} = 1300K$, $m_i = 16\times1.6\times10^{-24}g$, $\rho_0 = 5.5\times10^{-2}\text{cm}$, $\omega_d = 1.1\times10^8\text{rads}^{-1}$, $e = 4.8\times10^{-10}\text{statcoloumb}$, $T_d = 275\text{K}$, $r_d = 5\times10^{-4}\text{cm}$, $\sigma_0 = 1.5\times10^{-16}\text{cm}^2$, $U_{00} = 10^6$ stat volt cm^{-1}.

The time sensitivity of dimensionless electron temperature has been shown in Figure 2a,b for variant dust densities for both Gaussian and Sine time characteristics, respectively. Peak of temperature corresponding to extremum irradiance at $t'/t_0 = (-0.5)$ for Gaussian pulse and 0.8π for Sine pulse is observed. A shift is observed from the dust-free environment where the maxima is obtained at $t'/t_0 = 0$ (red shift) for Gaussian pulse and $t'/t_0 = 0.5\pi$ (blue shift) for Sine pulse[34] on account of increase in collision frequency due to the presence of dust. Also the maximal temperature decreases with increase in dust density on account of faster electron energy decay due to increase in collisions with the density of dust grains as shown in Figure 3a,b, which represent the variation of dimensionless modified collision frequency with time width for both the Gaussian and Sine time profiles for different dust densities, respectively.

Figure 4a,b show the comparative analysis of propagation of beam waist parameter with the dimensionless length at the same time width for both the Gaussian and Sine time irradiance of the beam. Self-focusing enhances with lesser time width for both the profiles but beam get self-focused faster in sinusoidal irradiance due to boost rate of nonlinearity under the influence of dust grains.

Figure 5 show the reliance of beam waist parameter with the dimensionless length of transmission at the same modified collision frequency and at the point of time width of maximal irradiance under the influence of dust grains shown by the solid (Gaussian) and dotted (Sine) lines. Both the exponential and Sine profiles show the oscillatory convergence but the frequency of Sine profile is greater than the Gaussian one on account of different rate of nonlinearity for both the profiles under the influence of dust grains. Even the waist of the beam becomes narrower as compared to the dust-free environment.[34]

Figures 6a,b describe reliance of beam waist parameter on the dimensionless length of transmission for distinguishable dust densities. Since the collision frequency increases with dust density which further enhances the nonlinearity of the beam causing the enhanced self-focusing for both the profiles of the beam with the dust density but the Sine profile self-focused at faster rate than Gaussian profile.

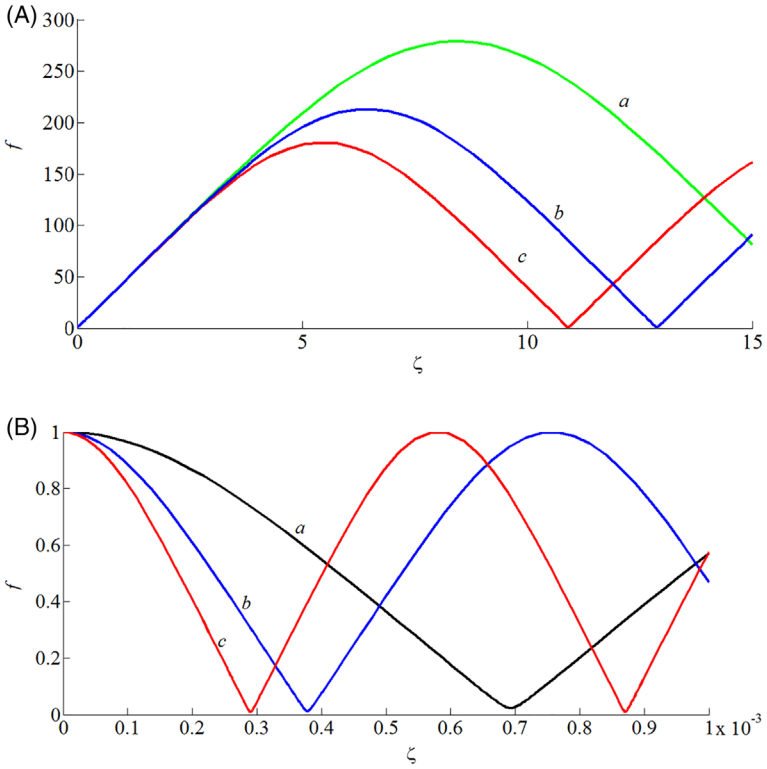


FIGURE 6 (a, b) Relying beam waist parameter f on the dimensionless length of transmission $\zeta(=z'/R_d)$ for the Gaussian and Sine beam for distinguishable dust densities, respectively. The dust densities, $n_d = 10$, 100, 1000 in cm^{-3} are labelled by a, b, and c, respectively

5 | CONCLUSIONS

This analytical model investigates the dust grain impact on self-focusing of both the Gaussian and Sine time irradiance of the beam transmitting through dusty plasma. A comparative analysis of growth of plasma parameters like temperature and collision frequencies modified under the influence of dust grains have been studied numerically to calculate the modified dielectric permittivity for both the profiles of the beam considering the paraxial ray approximation, neutrality of charge, ionization of neutrals, recombination of electrons with ions, charging flux via adsorption and photoemission of electrons, and diffusion. Also, the comparative growth of beam waist parameter along the dimensionless transmission length as a function of time width, point of maximal irradiance, and dust densities have been analysed for both the profiles in the impact of dust grains and compared with the dust-free environment. The consequences of this model are beneficial for the domain of dust particles like astrophysics, manufacturing, fabrication, plasma processing etc.

ACKNOWLEDGMENTS

The authors acknowledge to Prof. M. S. Sodha for his valuable discussions. The author (Ruchi Sharma) is thankful to her colleagues for their supportive approach to accomplish the research article.

REFERENCES

[1] A. V. Gurevich, Nonlinear Phenomena in the Ionosphere, Springer, New York **1978**.
[2] A. Gahlot, R. Walia, S. C. Sharma, R. P. Sharma, *J. Plasma Phys.* **2012**, *78*, 33.
[3] M. S. Sodha, A. Dixit, S. Srivastava, *Appl. Phys. Lett.* **2009**, *94*, 251501.
[4] M. S. Sodha, A. Dixit, *Appl. Phys. Lett.* **2009**, *95*, 101502.
[5] M. S. Sodha, S. Misra, S. K. Mishra, *Phys. Plasmas* **2009**, *16*, 123705.
[6] J. Goree, *Plasma Sources Sci. Technol.* **1994**, *03*, 400.
[7] M. Mikikian, L. Boufendi, A. Bouchoule, H. M. Thomas, G. E. Morfill, A. P. Nefedov, V. E. Fortov, PKE–Nefedov team, *New J. Phys.* **2003**, *5*, 19.
[8] I. Denysenko, J. Berndt, E. Kovacevic, I. Stefanovic, V. Selenin, J. Winter, *Phys. Plasmas* **2006**, *13*, 073507.
[9] D. Samsonov, J. Goree, *Phys. Rev. E* **1999**, *59*, 1047.
[10] P. K. Shukla, V. P. Silin, *Phys. Scr.* **1992**, *45*, 508.
[11] R. L. Merlino, A. Barkan, C. Thompson, N. D'Angelo, *Phys. Plasmas* **1998**, *5*, 1607.
[12] M. Bonitz, C. Henning, D. Block, *Rep. Prog. Phys.* **2010**, *73*, 6.

[13] S. V. Vladimirov, K. Ostrikov, A. A. Samarian, Physics and Applications of Complex Plasmas, Imperial College Press, London **2005**.
[14] V. N. Tsytovich, G. E. Morfill, S. V. Vladimirov, H. M. Thomas, Elementary Physics of Complex Plasmas, Springer, Berlin **2008**.
[15] O. Ishihara, *J. Phys. D: Appl. Phys.* **2007**, *40*, 8.
[16] H. Kersten, G. Thieme, M. Fröhlich, D. Bojic, D. H. Tung, M. Quaas, H. Wulff, R. Hippler, *Pure Appl. Chem.* **2005**, *77*, 415.
[17] F. Verheest, *Plasma Phys. Control. Fusion* **1999**, *41*, 3A.
[18] R. L. Merlino, Plasma Physics Applied, Transworld Research Network, Kerala, India **2006**, **2006**.
[19] S. C. Sharma, A. Gahlot, R. P. Sharma, *Phys. Plasmas* **2008**, *15*, 043701.
[20] M. S. Sodha, S. K. Mishra, S. Misra, *Phys. Plasmas* **2009**, *16*, 123701.
[21] M. S. Sodha, Kinetics of Complex Plasmas, Springer, India **2014**.
[22] G. Sambandan, V. K. Tripathi, J. Parashar, R. Bharuthram, *Phys. Plasmas* **1999**, *6*, 762.
[23] M. S. Sodha, S. K. Mishra, S. Misra, *Phys. Plasmas* **2011**, *18*, 023701.
[24] F. Verheest, P. Meuris, *Phys. Lett. A* **1996**, *210*, 198.
[25] J. Jia, C. Yuan, R. Gao, Y. Wang, Y. Liu, J. Gao, Z. Zhou, X. Sun, J. Wu, H. Li, S. Pu, *J. Phys. D: Appl. Phys.* **2015**, *48*, 465201.
[26] S. K. Mishra, S. Misra, *Phys. Plasmas* **2019**, *26*, 023702.
[27] R. Sharma, S. C. Sharma, *Contrib. Plasma Phys.* **2019**, *59*, 211.
[28] J. Jia, C. Yuan, S. Liu, F. Yue, R. Gao, Y. Wang, Z.-X. Zhou, J. Wu, H. Li, *Phys. Plasmas* **2016**, *23*, 043302.
[29] S. K. Mishra, S. Misra, M. S. Sodha, *Phys. Plasmas* **2011**, *18*, 043702.
[30] H.-b. Wang, Y.-y. Chen, H. Chung, *Optik-Int. J. Light Electron. Opt.* **2019**, *196*, 163148.
[31] M. S. Sodha, S. K. Mishra, S. Misra, S. Srivastava, *Phys. Plasmas* **2010**, *17*, 073705.
[32] M. Faisal, L. Bhasin, M. S. Sodha, *J. Geophys. Res.* **2009**, *114*, A01305.
[33] M. Sodha, A. Sharma, *J. Plasma Phys.* **2009**, *74*, 473.
[34] M. Faisal, M. P. Verma, M. S. Sodha, *Phys Plasmas* **2008**, *15*, 102301.
[35] S. A. Akhamanov, A. P. Sukhorukov, R. V. Khokhlov, *Sov. Phys. Usp.* **1968**, *10*, 619.
[36] M. S. Sodha, R. S. Mittal, S. Kumar, V. K. Tripathi, *Optoelectronics (London)* **1974**, *6*, 167.
[37] R. Fowler, Statistical Mechanics, Cambridge University Press, Cambridge **1936**.
[38] M. S. Sodha, A. Dixit, *Appl. Phys. Lett.* **2009**, *95*, 101502.
[39] J. D. Huba, NRL Physics Formularly, Naval Research Laboratory, Washington, D.C. **1998**.
[40] M. S. Sodha, A. K. Ghatak, V. K. Tripathi, Self-filamentation of Laser Beams in Dielectrics, Semiconductors and Plasmas, Tata-McGrawHills, Delhi **1974**.
[41] M. S. Sodha, A. K. Ghatak, V. K. Tripathi, *Prog.Opt.* **1976**, *13*, 169.

How to cite this article: Sharma R, Sharma SC. Theoretical analysis for transmission of Gaussian and sine time irradiance of electromagnetic beam in collisional dusty plasmas. *Contributions to Plasma Physics*. 2020;e201900175. https://doi.org/10.1002/ctpp.201900175

Rs. 30.00
ISSN- 0566-2257

UNIVERSITY NEWS

A Weekly Journal of Higher Education

Association of Indian Universities

Vol. 58 • No. 28 • July 13-19, 2020

Seema Singh
Theoretical Framework of Collaboration between MSME and Private Engineering Colleges to Co-exist during COVID-19

Preeti Dharmik, Rekha Sharma and C G Dethe
Impact of Orientation Programmes on Knowledge, Attitude and Skills of Senior University and College Teachers

Pranita Gopal
Good Practices in Virtual Presentations for Higher Education

M Venkaiah Naidu
Transformations through Innovative Technological Interventions

Convocation Address

Sajal Dasgupta
Innovative Initiatives Taken for Reforming Conduct of Education During COVID-19 Pandemic Lockdown Period

Communication

Celebrating 90 Years of University News

#Let'sBeatCoronaTogether

UNIVERSITY NEWS
Vol. 59 July 13-19
No. 28 2020
Price Rs. 30.00

A Weekly Journal of Higher Education Published by the Association of Indian Universities

In This Issue

ITEMS PAGE

Articles

Theoretical Framework of Collaboration between MSME and Private Engineering Colleges to Co-exist during COVID-19 3

Impact of Orientation Programmes on Knowledge, Attitude and Skills of Senior University and College Teachers 12

Good Practices in Virtual Presentations for Higher Education 20

Convocation

National Institute of Technology, Tadepalligudem, Andhra Pradesh 25

Communication

Innovative Initiatives Taken for Reforming Conduct of Education During COVID-19 Pandemic Lockdown Period 28

CAMPUS NEWS 30

AIU NEWS 32

Theses of the Month 29

(**Science & Technology**) 38

Advertisements 40

New Subscription Tariff

(Effective April 01, 2020)

	Inland		**Foreign**	
Institutions	Academics/ Students *(at residential address only)*		Airmail	Surface Mail
	Rs.	Rs.	US$	US$
1 year	1250.00	500.00	210.00	170.00
2 years	2200.00	900.00	400.00	300.00

Subscription is payable in advance by Bank Draft/MO only in favour of Association of Indian Universities, New Delhi.

Opinions expressed in the articles are those of the contributors and do not necessarily reflect the views and policies of the Association.

Editorial Committee Chairperson:
Dr (Ms) Pankaj Mittal

Editorial Committee:
Dr Baljit Singh Sekhon
Dr Amarendra Pani
Dr Youd Vir Singh

Editor:
Dr Sistla Rama Devi Pani

#Let'sBeatCoronaTogether

Theoretical Framework of Collaboration between MSME and Private Engineering Colleges to Co-exist during COVID-19

Seema Singh*

In December 2019, a Novel Coronavirus (SARS-CoV-2) emerged in Wuhan, China which became a global pandemic within months (Bouey, 2020). The virus has spread to more than 215 countries with 7.8 million infected and has caused 429,782 deaths. As a preventive measure, Indian Government has imposed a complete lockdown, barring emergency services. This has played havoc on the life and livelihood of a large number of people, more on those who have little reserves to spend. The medical crisis has profound economic and social effects. The Micro, Small and Medium enterprises (MSMEs) are a significant part of India's economic landscape as in many other countries of the globe. The MSMEs generate revenue, enhance employment and provide access to commodities for remote communities (World Bank, 2004; Wiklund, 2005; Satpathy and Rani, 2017). As per the Report of the Working Group on Micro, Small and Medium Enterprises (MSMEs) growth for 12th Five Year Plan (2012–17), the sector accounts for 45 per cent of the manufacturing output and 40 per cent of the total exports of India. Further, it accounted for 30.74 per cent of GDP in the year 2014–15. As per the registration between October 2015 to December' 2019, the micro sector (88.6 per cent) is the largest component of the MSME than the other two, small (10.9 per cent) and medium (0.42 per cent) sectors and provides employment to 61.5 per cent of the workforce against 32.2 per cent and 6.3 per cent in case of small and medium enterprises respectively (GoI, 2020). However, the micro sector has remained comparatively under-researched (Cristian, 2016). Add to this, Indian engineering education system has registered an exponential growth during the last twenty-five years due to enhanced global demand for English-language speaking engineers, who are equally competent, but at lower wages. Moreover, the un-aided private sector is a significant component of the whole spectrum of engineering education systems in India. What is the effect of lockdown on the micro sector as per the anecdotal evidence available? The hundreds of thousands of stranded migrant workers across the country, suggests that MSMEs have been the worst casualty of Covid-19 induced lockdown (Mishra, 2020). Higher education in general and engineering education in particular has also been adversely affected by the pandemic and they are still in search of an alternative sustainable model. In this background, the study tries to investigate whether a collaboration between these two sectors will be a win-win situation for both. The second section of the paper provides relevant literature review. Objective and methodology have been discussed in

*Professor, Department of Economics, Delhi Technological University, Rohini, Delhi- 110 042. E-mail: seemasinghdtu@gmail.com

the third section of the paper. Fourth section provides an analysis. A model has also been developed for cooperation between these two sectors Last section concludes the discussion and provides suggestions.

Literature Review

The literature review has been done in the four groups as i. The Micro Sector and Impact of COVID-19 on it; ii. Private Engineering Education in India and impact of COVID-19 on it; iii. University and Firms collaboration:

The Micro Sector and Impact of COVID-19

MSMEs are perceived as the backbone of the economy and 'an engine of growth' not only in India (Satpathy and Rani, 2017) but in other parts of globe (European Commission, 2016; Audretsch, 2009; Tesfayohannes, 1998). They also have a high export propensity (Claudiu Cicea, Ion Popa, Corina Marinescu & Simona Cătălina Ștefan, 2019), Huong T.X. Nguyen & Viet Le (2019)). However, it needs to be kept in mind that the small firms are not homogenous. They all are unique in nature and not like miniature corporate firms. So, any theory developed for corporate firms cannot be replicated to them (Beaver and Prince, 2004). In spite of being so important a segment of the economy, they have very poor technological endowment and their up-gradation is essential for the economic wellbeing of the nation (Venkatesh and Muthiah, 2012). Their output may be high in quantitative terms, but is very low in quality which can be up-graded by technological up-gradation of the sector. The author further view that advanced technology enables the small firms to create quality employment improving remuneration, duration, and skills (Bhavani T.A.2011; Singh et al. 2012). However, the study conducted by Bargal et.al. regarding performance of small scale industries (SSIs) during the period of nineties vis-à-vis the pre-reform years the annual average growth rate of different parameters have concluded that they have declined during post-liberalisation period (Bargal et al., 2009). According to Lawson and Samson (2001), a firm which is able to link the technological strategy with its business and innovation strategy, are capable of bringing innovation in the firms. Experts as Yang (2012), Martín-de Castro et al. (2013) and Gupta and Nanda (2015) have also highlighted the importance of technological adaptation for enhancing competitiveness. There is consensus among scholars that small firms should concentrate on developing market reach (Wheelen & Hunger, 1999). However, they are unable to raise finances to invest in cost-reducing technology for the firm (Beaver and Prince, 2004).

During COVID-19, countries have taken unprecedented and drastic measures to contain the spread of the virus by closing towns, cities, and national borders, disrupting not only the big corporations but also millions of SMEs around the world (SMEFF, 2020). Though it is too early to assess its full impact on MSME, but given that the COVID-19 is a pandemic has affected humanity, MSMEs would not be immune from it. The world economy is expected to face a very serious crisis besides on the social and political sphere. Over a period of fewer than two weeks, since it was officially declared a pandemic by WHO, COVID-19 has disrupted lives and economies around the globe. According to a briefing issued by the United Nations Department of Economic and Social Affairs, the COVID-19 pandemic will likely shrink the global GDP by almost one per cent in 2020. In such a brief period, many international organizations have analysed, assessed, the impact of the COVID-19 pandemic, and formulated scenarios for world economic growth in 2020. According to Citi Research (published April 7, 2020), the world economy will decline by -2.3 per cent in 2020 (compared to 2.6 per cent in 2019, more than -1.7 per cent in 2009). In particular, China's GDP growth was only about 2.4 per cent (compared to 6.1 per cent in 2019), United States -2.6 per cent(compared to 2.3 per cent in 2019), Japan -1.9 per cent (compared with an increase of 0.7 per cent in 2019), while the Eurozone has regressed -8.4 per cent (compared to 1.2 per cent in 2019) (Dung, Thi Bich and Tho Trần Bá, 2020). About 9-18 million formal jobs may be lost and 100 mn informal jobs vulnerable to income loss in Sub Saharan Africa (McKinsey, 2020). Even studies have been done on the impact in India show that about 30 to 40 percent of the micro sector firms will not be able to survive. However, the impact will be differential depending on the activity of the firm. (ISLE-ILO, 2020).

The pandemic cannot be compared with any other natural disaster due to its magnitude. But, in absence of any other event similar to COVID-19, a comparison to flood can be made. Aruzzir (2018) in her study in Malaysia for another natural calamity, flood, has found that the effect of any natural disaster on SMEs is very severe (Auzzir, 2018). Bouey (2020) has done the immediate impact analysis on the Chinese SME sector. China is the second-largest

economy with international trade with almost all countries of the world which ranges from mineral and food products to highly sophisticated electronic gadgets and automobile parts. So, certainly, it will affect the world's economy (Bouey Jennifer' 2020). With fears of a new recession and financial collapse, Nicola et. al. (2020) have suggested that these times call for resilient and strong leadership in business and government along with wider society. Immediate relief measures need to be implemented and adjusted for those that may fall through the cracks. A broad socio-economic development plan including sector-by-sector plans need to be created, and along with it an ecosystem that encourages entrepreneurship so that those with robust and sustainable business models can be allowed to flourish. COVID-19 will put many small businesses on life support (Nicola et. al. 2020). Karen G. Mills has suggested policymakers on aid options, offers guidance to owners on the brink of ruin (Kost, 2020), and organizations need to develop a trust-based culture with their employees (Beer, 2020). Samuel in a study of 300 randomly sampled SME-Owners in Ghana has discussed the importance of financial literacy for the SME growth (Samuel, 2018).

According to the extant literature, MSMEs are more vulnerable to natural hazards than large firms due to the fact that MSMEs: tend to operate in sub-optimal locations; are smaller and financially weaker; have a more limited, usually local market; tend to implement less DRR measures and be more excluded from recovery programs (Zhang et. al. 2004; Battisti, Martina, and David Deakins, 2012). According to the US Institute for Business and Home Safety, many businesses do not reach the post-disaster stage (Institute of Business and Home safety).

Private Engineering Colleges in India and Impact of COVID-19

History of Engineering Education in India can be traced back to 1794, when the School of Survey was established at Guindy, Tamil Nadu on the recommendation of Michael Topping to train Indians on construction of roads, bridges, docks, etc. With the involvement of British in the Second World War and the first Indian war of independence, mechanical engineering was added with civil engineering to train the lower grade technicians to manage equipment used for the army, navy and other technical establishments for maintenance of British Empire. (Shaha S., 2011). Private higher education institutions are established by philanthropic, private business establishments, religious and for-profit organizations. In 1970, India had a total of 139 engineering institutions, and only four of them were private. (Singh and Singh, 2018).

After Globalisation and use of Information and Communication Technology in the production process, producers have the opportunity to shift a part of the production process to far off places, where it can be done with required quality at low cost. Many large and multinational corporations came to India to use its English-speaking engineers besides many other resources (Singh, 2005). With the country counting on its demographic dividend, the increase in the number of students seeking engineering degrees put up pressure on the higher education sector to expand. Limitation of the government funding in engineering education which was not sufficient to meet the growing demand, led to the historic decision of privatisation. The Private Universities (Establishment and Regulation) Bill was introduced in 1995 in the Parliament. Passing the bill gave a boost to the opening of many private engineering colleges. By the end of 2000, number of engineering institutions rose to nearly 1,400 out of which about 200 only belonged to the government (Singh and Singh, 2014).

As per the Table-1, private engineering colleges were established across India but number was quite impressive in southern states. This is necessary for higher education institutions (especially private engineering institutions) to collaborate or develop partnership with any reputed organization or elite institutions both nationally and internationally rather than depending on students pass out rate and placement only to attract students amidst the fierce competition (Singh & Singh 2018).

University and Institute Collaboration

The Higher Educational Institutions are indulging in market-like behaviour for revenue generation. They are collaborating with business firms for technology transfer all over the world for raising funds. On the basis of the 2008-10 data from National Research Foundation of Korea, Ministry of Education and Technology and Korea Foundation for Promotion of Private Schools, Han and Heshmati (2013) have analysed the financial rewards from university-industry research cooperation. As per the findings, participation of engineering faculty, patent approvals, the volume of research funds, and firms in

Table 1. Regional Presence of Engineering Institutions

Region	State/ Union Territory	Government Aided	Private Unaided	University Managed
North	Chandigarh	6	1	2
	Delhi	18	18	1
	Haryana	46	288	5
	Himachal Pradesh	17	36	1
	Nammu & Kashmir	9	18	1
	Punjab	34	186	6
East	Arunachal Pradesh	2	1	NIL
	Assam	15	5	2
	Jharkhand	16	16	1
	Maniour	2	NIL	NIL
	Meghalaya	3	1	NIL
	Orissa	23	172	2
	Sikkim	2	NIL	1
	Tripura	2	NIL	NIL
	West Bengal	59	93	3
West	Goa	7	3	NIL
	Maharshtra	71	660	8
South	Andaman and Nicobar	1	NIL	NIL
	Karnataka	158	239	3
	Kerala	90	95	13
	Andhra Pradesh	132	763	5
	Pondicherry	8	13	NIL
	Tamil Nadu	78	852	4
	Daman & Diu	1	NIL	NIL
	Goa	7	3	NIL
Central	Chattisgarh	25	47	3
	Gujarat	49	128	3
	Madhya Pradesh	66	213	6

Source: AICTE website, Singh & Singh 2018.

incubators within universities turn out to be significant contributors to externally sourced university revenues. They have given an exhaustive list of study conducted on the same issue (Table-2) which shows positive result out of their collaboration.

It appears that industry-university cooperation foundation programs are likely to play a significant role in private university finances.

In this background, the paper discusses the possibility of collaboration of private engineering colleges and micro sectorto overcome the setback from the COVID-19 pandemic. Whether it will be a win-win situation. The discussion has been made on the basis of secondary sources and small studies done by individual researchers and organisations.

Analysis

Recently, the government of India has tried to remove the ambiguity and clearly define these sectors as discussed in the Table-3. Along with investment, turnover criteria has also been added. The rand has also been increased. However, keeping micro and medium enterprise in one group is akin to comparing cheese and chalk. While the medium sector is qualified and have resources to hire experts and technology to improve their productivity and efficiency, take lone from the bank, the micro sector is not the perspective or if they have, they are not able to develop sufficient resources to accomplish the task.

Coronavirus and following nationwide lockdowns have put the economic future in jeopardy. However, it is a question of survival for micro sector As per various studies about 30 to 40 percent of the micro sector firms will not be able to revive again (ISLE-ILO, 2020). At the same time, many of them are working in such areas that if they become part of a network or ammelgoration, their return on investment will increase which needs strategic planning and technological updardation along with finances. So, in the light of government financial package for revival and mudra loan, micro firms need an institutional support for assist in strategic planning and technological upgradation.

As per AICTE dashboard, in 2018-19, there were 2,678 private unaided engineering institutions in India with 6,09,090 enrolments and 3,14,840 faculty. The intake capacity was 13,20,573. Its means more than fifty percent seats remained vacant during 2018-19. Now, in the post COVID-19 academic session 2020-21, loss of parent's job or cut in salary or delayed payment of salarymay have cascading effect on the demand for seats in such engineering colleges. Universities across the world are expecting shrink in sources of income and as an alternative, many world class universities have started offering on-line

Table 2: Study on University - Firm Collaboration

Researcher and Sample	Findings
Joshua and Patricia (2003) on the basis of study of 108 universities in the USA	Age of Technology Transfer Office, the number of outstanding engineering faculty members and research fund received by external organisations have a positive relation with performance respectively.
Thursby and Thursby (2002) on the basis of 64 universities in USA	Performance is related to the number of experts, faculty members, andpersons working at the Technology Transfer office and their relationship with external firms
Jenson and Thursby (2003) Interview and Regression (2003	Reward system through technology transfer is directly related with performance
Sigel et al. (2003a) Interview and Regression for 98 people who are directly related to TTO in US	Reward systems for interested persons have effects on performance
Power (2003) Interview and Regression for 66 enterprises and 312 university researchers.	Research funds including both public and private have a positive correlation with patent products. But this is not related to revenue through the licensing. ·The more distinguished faculty member a university has, the more patents and licensing a university has.
Kim and Lee (2007) Regression 61 universities in Korea	Research competency such as the number of SCIE papers and the number of patent registrations were significant, but managerial competency such as the scale of technology transfer organizations and the number of specialists was not statistically significant
Kwon and Han (2009) Regression for 169 universities in Korea, Explorative	• Public universities have higher performances than those of private universities in terms of the number of technology transfers and the amount of technology licensing fees. • The characteristics of university such as age, size, number of departments, faculty members, students, and experts, the number of SCIE papers are not statistically related to performances. • Regional universities have higher performance than universities in urban areas in terms of technology transfer.
Seo et al. (2005) descriptive	University's own firms based on university's own technology and holding companies are related with revenue through industry-university collaboration.
Sapsalis et al. (2006) Regression for 89 universities in United States. .	Scientific competency, the number of papers, and patents have a positive effect on revenue through industry-university collaboration.
Kim (2005) Survey for 54 universities and 79 firms in Korea	• Level of education of experts working at technology transfer office has a positive effect on revenue through industry-university collaboration • The number of patents has an effect on revenue through industry-university collaboration
Friedman and Silberman (2003), Jaffe et al. (1993) regression with 12 firms	• The degree of proximity of high-tech firms near school has positive effects on technology transfer. • Geographical location of universities has positive relation with knowledge spill over.
Foltz et al. (2000) Survey for bio-agricultural firms in United States	Research funds received from federal government and university have positive effects on revenue, while research funds received from industries have no relation with performance.
O'Shea et al. (2005) Byun (2004) survey 107 universities in Korea	It shows that university scale, age of professional institutions, and incentive systems for technology developers have positive effects on revenue through industry-university collaboration.

Source: Han Junghee and Heshmati Almas (2013), p.19-20

degrees. A representative list has been discussed in the Table 4.

These efforts may further bleak the future for unaided private engineering colleges in India. However, they have human capital, in the form of qualified faculty, laboratory assistants and young energetic students. Not to forget, the role of higher education institutions (HEIs) towards the development of their surrounding regions is not a new phenomenon. This has been illustrated by the establishment (in the late 1800s) of land-grant colleges in North America and technical and civic universities in Germany and England, respectively (Editorial, 2017).As Borrel-Domial et.al. (2010) have discussed that generally Ph.D. is pursued for following teaching profession but now, about 50 per cent are pursuing for a wide range of careers. On the basis of the experiences reported by stakeholders, the paper suggests working practices and supporting policy measures that can address the (potentially conflicting) requirements for a high quality of education, sound research and adequate preparation for diverse career with participation of private sector firms (Borrel-Domial et.al., 2010). Maya Ziswiler of the UBS Optimus Foundation (UBSOF) has discussed innovative ways the education sector is using both blended finance and outcome-based financing. For example, UBSOF's investment in education social enterprises with Acumen in East Africa provides loans and equity investments coupled with a grant for pre- and post-investment support and an outcome-based incentive for both the investment manager and the investees if outcome targets are reached. Another example mentioned for supporting education social enterprises was impact loans, where the interest rate fluctuates based on the outcomes achieved. social and development impact of bonds (SIBs and DIBs), a form of outcome-based financing where investors provide up-front capital to service providers with the opportunity for a return if agreed-upon outcomes are achieved (Gustafsson- Wright Emily, 2020). In India, firms are collaborating with engineering institutions but they are generally large firms or government organisations. It is not common for

Table 3: MSMEs Classification in India

Existing Criteria	**Investment in Plant & Machinery or Equipment**		
Characteristic	Micro	Small	Medium
Manufacturing Entreprises	Investment< 25 Lac	Investment< 5 Crore	Investment.< 10 crore
Service Entreprises	Investment.<10 Lac	Investment.< 2 Crore	Investment< 5 crore
Revised Criteria	Composite Criteria - Investment and Annual Turnover		
Manufacturing & Services	Investment< 1 crore and Turnover< 5 crore	Investment.< 10 crore and Turnover< 50 crore	Investment< 20 crore and Turnover< 250 crore

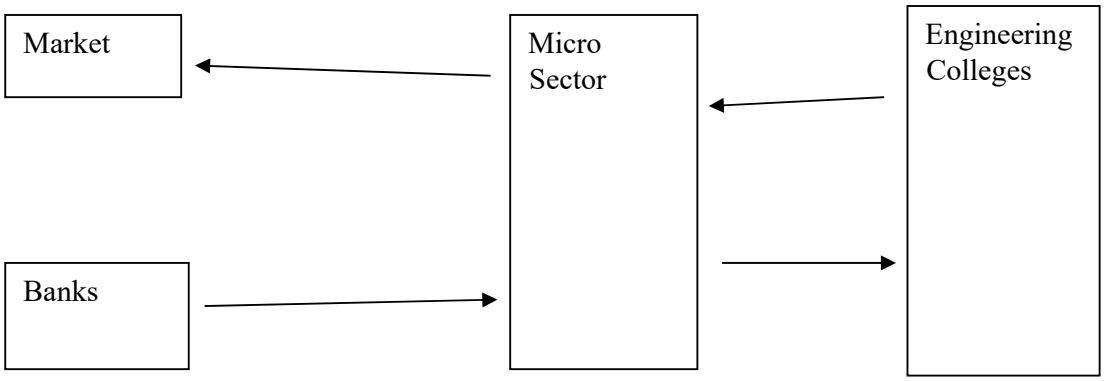
Source: Rao Harshitha, 2020

Table 4: List of On-line Courses

SN	**Course**	**Offered by**
1	MS in Computer Science	Arizona State University
2	Data Science for Engineers	IIT, Madras, July 20-Sept. 11, 2020
3.	Industry Ready Master Classes for Civil Engineering Students	contact@protrainy.com +91 7978817062, +91 8617833413 programme starting from June 7' 2020
4.	Basic Electrical Circuits	IIT Madras [July 20 to Oct 9' 2020
5.	M.Sc in Computer science	Liverpool John Moores University
6.	Master of Computer and Information Technology	University of Pennsylvania [16-40 Months, Fully Online]: Apply by Aug 1
7.	From 0 to N: Architecting on Alibaba Cloud	Alibaba Cloud Academy [Online, 5 Months] Gustafsson-Wright Emily (292

Source: Compiled by author

Fig 1: Model of Engineering Institutions and Micro Sector.



the medium sector, and almost absent in case of micro-sector. However, looking at their capacity to contribute towards the GDP and create employment and for enhanced use of ICT after COVID-19 in the production process, they need use of technology which can be provided by the Private engineering colleges and along with helping them in strategic planning. At present, there is no such platform. It is up to colleges, Industry bodies as CII or FICCI etc. which may work in that area. At Delhi Technological University, B.Tech. students visit slum to assess how knowledge of their branch of engineering can be used to improve life at slum since the academic year 2011-12. (Singh, 2013). The process has been integrated with the examination system as marks is given out of internal marks (Singh, 2016). Similar procedure can be adopted for associating with micro sector. The Unnat Bharat Abhiyan[2]of the Ministry of HRD is similar programme for rural development since 2014.Funds for Awareness, survey, visits etc @₹ 10000/- per village • Technical Intervention ₹ 1.0 lakh per technology for implementation in the village • Customization of Technology ₹ 50000/- per solution for villages.In the Fig-1, a Model for collaboration between Micro sector and Private engineering colleges (unaided) has been represented to achieve a win-win situation for both.

From the financial help under COVID-19 revival package of the government or Mudra loan, the Micro sector can approach engineering institutions for technological help or assistance in strategic planning. The enabled Micro Sector with the help from engineering colleges may go to market alone or may be as part of Cartel. They can become member of a production network for value addition also. This is the time to float loan to Micro Sector and knowledge input from engineering colleges to see the practical implication of the model. Professor ERichard Thaler and legal scholar Cass Sunstein have discussed about NUDGES in their New York Times best seller book entitled Nudge: Improving Decision about Health, Wealth and Happiness. It is crucial to get the necessary response from the beneficiaries, Micro sector firms and Private engineering colleges (unaided) to determine the success or failure of the model. NUDGES stand for incentives, Understand mappings, Defaults, Give feedback, Expect error, Structure complex choices. Through NUDGES, the model will improve further and become sustainable over a period of time.

Conclusion

The lockdown as a preventive measure during COVID-19 has disastrous impact on the micro sector which has employment potential and which contributes significantly to the GDP and exports. Though the government has provided a revival package, but *sans* competitive technological assistance or sustainable business model, survival will be difficult post-lockdown. On the other hand, Private engineering colleges will also have pressure of generating finances to survive. The paper concludes that more empirical research is needed to develop a more unified approach to MSMEs success factors, which informs thinking on this critical matter, to both policy makers and business owners (Lampadarios Evripidis; Kyriakidou Niki; Smith Gordon (2017) and engineering institutions. The paper has real implication for today's world situation which has never been experienced earlier. The model may be used as a win-win situation for both. A secondary source of financing for private engineering colleges and a hand-holder for micro sector.

Notes:

On 13th May 2020, Finance Minister Nirmala Sitharaman added the additional principle of turnover along with the investment. The Union Cabinet had approved the amendment to change the criteria to classify MSMEs from "investment in plant and machinery" to "annual turnover."

Unnat Bharat Abhiyan 2.0, an extension of the programme with improved feature has been stated in 2018.

References

1. Attisti, Martina, and David Deakins (2012). Perspectives from New Zealand Small Firms: Crisis Management and the Impact of the Canter bury Earthquakes, New Zealand Centre for Small & Medium Enterprise Research (Ed.), Business SM Easure. Palmerston North: Massey University.
2. Audretsch, D B (2009). The Entrepreneurial Society. *The Journal of Technology Transfer*, *3*, 245–254.
3. Ballesteros, Marife M, and Sonny N Domingo, (2015). Building Philippine SMEs Resilience to Natural Disasters. No. 2015-20. PIDS Discussion Paper Series, 2015-20, Econstor; https://www.econstor.eu/bitstream/10419/127030/1/pidsdps1520_rev.pdf dated 10.6.2020.
4. Bargal, H, Dashmishra, M., and Sharma, A. (2009), „Performance Analysis of Small Scale Industries - A Study of Pre-liberalization and Post-liberalization period", International Journal of Business and Management, Vol 1, No 2. [3].
5. Bavel Jay J Van, Baicker Katherine et.al.(2020), Using social and behavioural science to support COVID-19 Pandemic Response, Perspective, Nature Human Behaviour, https://doi.org/10.1038/s41562-020-0884-z, pp.
6. Beaver and Prince (2004). Management Strategy and Policy in the UK small business sector: a critical review, *Journal of Small Business and Enterprises Development*, 11(01).
7. Beer Michael (2020). Organizations will Develop Trust-based Cultures with Employees in Gerdman Dina, How The Corona Virus is already Rewriting the future of Business, Working Knowledge, Harvard Business School, March 16 https://hbswk.hbs.edu/item/how-the-coronavirus-is-already-rewriting-the-future-of-business, 14.5. 2020.
8. Bhavani, T.A. (2011). Dynamic Business Environments: What These Mean for Indian Small Enterprises in Micro and Small Enterprises in India: Era of Reforms: Keshab Das (Ed) pp. 27-45.
9. Borrell-Damian, L, Brown, T, Dearing, A. *et al. (2010)* Collaborative Doctoral Education: University-Industry Partnerships for Enhancing Knowledge Exchange. *High Educ Policy* 23 (2), pp**.** 493–514.
10. Bouey, Jennifer (2020). Assessment of COVID-19's Impact on Small and Medium-Sized Enterprises: Implications from China. *Santa Monica*, CA: RAND Corporation,. https://www.rand.org/pubs/testimonies/CT524.html.
11. Claudiu Cicea, *et. at.* (2019). Determinants of SMEs' performance: evidence from European countries, Economic Research-Ekonomska Istraživanja, 32:1, 1602-1620, DOI: 10.1080/1331677X.2019.1636699.
12. European Commission. (2016). Annual Report on European SMEs 2015/2016. Retrieved from http://ec.europa.eu/growth/smes/business-friendly-environment/performance-review-2016_ro
13. Gherhes Cristian, *et. al* (2016). Distinguishing micro-businesses from SMEs: A systematic review of growth constraints *Journal of Small Business and Enterprise Development,* 23(4), November.
14. Gupta, H and Nanda, T, (2015). A Quantitative Analysis of the Relationship Between Drivers of Innovativeness and Performance of MSMEs', *International Journal of Technology, Policy and Management*, Vol. 15, No. 2, pp.128–157.
15. Gustafsson- Wright Emily. (2020). Education In the time of COVID- 19: Reflection from the Charcha 2020 Forum Brookings, https://www.brookings.edu/blog/education-plus-development/2020/05/26/education-in-the-time-of-covid-19-reflections-from-the-charcha-2020-forum/?utm_campaign=Center%20for%20Universal%20Education&utm_medium=email&utm_content=88647927&utm_source=hs_email on 10.6.2020
16. Han Junghee and Heshmati Almas , (2013). Determinants of Financial Rewards from Industry-University Collaboration in South Korea, IZA Discussion Paper No. 7695 October, Bonn, Germany pp. 1-21.
17. Huong T.X. Nguyen & Viet Le (2019). Network Ties and Export Propensity of Vietnamese Small and Medium Enterprises, *Asia Pacific Business Review*, 25:1, 100-122.
18. Jensen. R. and Thursby, M (2003). Disclosure and Licensing of University Inventions: The Best We Can Do, *International Journal of Industrial Organization* 21, 1271- 1300.
19. Kohoutek, J, *et al.* (2017). The Role of Higher Education in the Socio-Economic Development of Peripheral Regions. *High Educ Policy,* 30 (4), 401–403 (2017).
20. Kost Danielle (2020). How Small Business can survive the Coronavirus Outbreak, Working Knowledge, Harvard Business School, March 20https://hbswk.hbs.edu/item/how-small-businesses-can-survive-the-coronavirus-outbreak
21. Lampadarios Evripidis, Kyriakidou Niki, and Smith Gordon (2017). Towards a new framework for SMEs success: a literature review, *International Journal of Business and Globalisation,* 2017 Vol.18 No.2, pp.194 – 232.

22. Lawson, B and Samson, D (2001). Developing Innovation Capability in Organisations: a Dynamic Capabilities Approach, *International Journal of Innovation Management*, Vol. 5, No. 3, pp.377–400.

23. Martín-de Castro, *et.al,* (2013). Linking Human, Technological, and Relational Assets to Technological Innovation: Exploring a New Approach', Knowledge Management Research & Practice, Vol. 11, No. 2, pp.123–132.

24. Metcalf, G., Jenkinson, K. and Johnstone, K A, (2010). Changing Climate for Business. 3rd ed. Oxford: UK Climate Impacts Programme, https://ukcip.ouce.ox.ac.uk/wp-content/PDFs/UKCIP_Business.pdf 12.6.2020

25. Mishre Udit (2020). Explained: Why are Medium, Small, Micro Enterprises worst hit by Covid-19 lockdown?, *The Indian Express*, May 7, https://indianexpress.com/article/explained/coronavirus-india-lockdown-msme-sector-crisis-government-relief-package-6395731/ on 14.5.2020

26. Nicola, M., *et.al*., (2020). The Socio-Economic Implications of the Coronavirus and COVID-19 Pandemic: A Review. *International Journal of Surgery (London, England)*, S1743-9191(20)30316-2. Advance online publication. https://doi.org/10.1016/j.ijsu.2020.04.018 on 11.6.2020

27. Rampal Jeevant, (2020). Corona Virus Crisis: A Behavioural Economist's guide to shaping India's Covid:19 Policy, https://scroll.in/article/958269/a-behavioural-economists-guide-to-shaping-indias-covid-19-policy on 26.5.2020.

28. Rao Harshitha (2020). Turnover to Define MSMEs: The New MSME Definition 2020, https://razorpay.com/learn/new-msme-definition-turnover-2020/ on 27.5.2020.

29. Saha, S.K. and Ghosh, S (2011). Engineering Education in India: Past, Present and Future. *Propagation*, 2(2), pp.114.

30. Samuel Kwaku Agyei, Christian Nsiah (Reviewing Editor) (2018) Culture, financial literacy, and SME performance in Ghana, *Cogent Economics & Finance*, 6:1.

31. Satpathy Saritaand and Rani Sailja (2017). A Study of Micro, Small and Medium Entrepreses: the Back Bone for Economic Development of the Indian Economy, International *Journal of Research and Scientific Innovation,* 04(06) June, pp 75-82.

32. Singh Antra and Singh Seema, (2018). Expansion of Private Engineering Institutions: Challenges Ahead, *Higher Education for the Future*, 5(1) pp. 20–39.

33. Singh Antra and Singh Seema, (2014). Private Engineering Education in India: Past, Present and Future, *International Journal of Management and Social Sciences Research, ISSN: 2319-442; 03(11),* November; pp. 39-46.

34. Singh, R, Verma, O P, and Anjum, B (2012). Small Scale Industry: An Engine of Growth, *Zenith International Journal of Business Economics & Management Research*, Vol.2 Issue 5.

35. Singh Seema, (2016). Integrating Social Responsibility of University and Corporate Sector for Inclusive Growth in India, *Higher Education for Future* (Sage Publication), 03(02) July 2016, pp. 183-196.

36. ----------, (2013). Social Responsibility of Engineering Institutions for Inclusive Growth, *Universities News, 2013,* 51(03), pp. 01-04-12.

37. ----------, (2005). MNCs Effect on Employment of Engineers, *The Indian Journal of Technical Education,* 28(2), April- June 2005, pp. 38-44.

38. SME Finance Forum (2020). Five Lessons From COVID-19 Crisis and SME, https://www.smefinanceforum.org/post/5-lessons-we-learned-about-covid-19-crisis-and-smes on 14.5. 2020.

39. Tesfayohannes, Mengsteab, (1998). The Promotion of Small and Medium-Scale Enterprises Financing as a Contribution to Sustainable Industrial Development- The Case of Eritrea. Aachen: Shaker Verlag, Germany.

40. Venkatesh, S and Muthiah, K, (2012). SMEs in India: Importance and Contribution, *Asian Journal of Management Research*, Vol. 2, No. 2.

41. Yang, J (2012), Innovation Capability and Corporate Growth: An Empirical Investigation in China', *Journal of Engineering and Technology Management*, Vol. 29, No. 1, pp.34–46.

42. Zairol Auzzira*, Richard Haighb , and Dilanthi Amaratungab, (2018). Impacts of Disaster to SMEs in Malaysia, *International Conference on Building Resilience using scientific Knowledge to informpolicy and practice in disaster risk reduction*, ICBR 2917 27-29 November; 2017.

43. Zhang, Yang, (2004). Modeling and Managing the Vulnerability of Community Businesses to Environmental Disasters, *45th Annual Conference of the Association of Collegiate Schools of Planning*, Portland, OR, October. Vol. 23. ❑

Impact of Orientation Programmes on Knowledge, Attitude and Skills of Senior University and College Teachers

Preeti Dharmik*, Rekha Sharma and C G Dethe*****

India's higher education system is the world's third largest in terms of students, next to China and the United States. In future, India will be one of the largest education hubs. India's Higher Education sector has witnessed a tremendous increase in the number of Universities/University level Institutions & Colleges since independence (Sheikh Younis Ahmad, 2017). Higher education is becoming a major driver of economic competitiveness in an increasingly knowledge-driven global economy. The imperative for countries to improve employment skills calls for quality teaching within educational institutions. In order to identify and promote good teaching practices, institutions may implement evaluation mechanisms. The environment of higher education institutions can enhance the quality of teaching through various means, for example, by issuing a national policy or recommendations by quality assurance agencies (Hénard, 2010).

The National Policy on Education (NPE) 1986 recognized the urgent need for creating effective systems to provide opportunities for professional and career development of teachers, necessary for quality education and research so that they are well equipped and motivated to accept new challenges emerging from growth of new knowledge, international competitiveness and changing requirements of learners, especially in the institutions of higher education. Therefore, it was proposed to enhance their motivation skills and knowledge through systematic orientation in specific subjects, techniques and methodologies, and thereby inculcate in them the right kind of values that would in turn encourage them to take initiatives for innovative and creative work (Guidelines for Regional Centre for Capacity Building and Human Resource Development Centre, 2015).

** Assistant Professor, UGC- Human Resource Development Centre, Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur- 440033. E-mail: preeti.dharmik5@gmail.com*
***Associate Professor, UGC- Human Resource Development Centre, Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur- 440033. E-mail: drrekha_sharma2000@yahoo.com*
****Director, UGC- Human Resource Development Centre, Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur-440033.E-mail: cgdethe@gmail.com*

In order to achieve the above objectives, the Academic Staff orientation scheme, later known as Academic Staff College (ASC) scheme was initiated by the University Grant Commission in 1987 based on the NPE recommendations. There are 66 ASCs in India spread over 26 states of our country. In 2015, these Academic Staff Colleges were renamed as Human Resource Development Centres. The Guidelines for UGC—Academic Staff College from 2002-07 and 2007-12 and UGC- Human Resource Development Centre, 2019 mentioned the four components to be included in orientation courses viz.,

Component A: Awareness of linkages between society, environment, development and education

Component B: Philosophy of education, Indian education system and pedagogy

Component C: Resource awareness and knowledge generation and

Component D: Management and personality development

After renaming of UGC-Academic Staff College to UGC-Human Resource Development Centres in 2015, the topics to be included in Orientation Programmes were redefined as follows:

- 10 per cent weight for topics in higher education such as issues of ethics, gender, marginalized communities, plagiarism etc.
- 10 per cent weight for issues related to environment.
- 10 per cent weight for issues concerning service matters of teachers.
- 20 per cent for broad cross discipline topics to motivate the trainees for development of interdisciplinary understanding and interest including basic legal awareness.
- 10 per cent weight for Research Methodology.
- 15 per cent weight for Communication Skills and Information Technology.
- 10 per cent weight for Microteaching

- The remaining 15 per cent of content of the First Level Programme will be focused on broad understanding of various subjects with recent global trends and developments.

Though the weightage for each domain were redefined, but the main thrust areas were similar to the UGCs Guidelines issued during 2002-07, 2007-12 and 2019. UGC-Human Resource Development Centre of Rashtrasant Tukadoji Maharaj Nagpur University was established in 1998 and till date it has organised 105 orientation programmes and catered to about 4363 teachers from university and colleges. Several researches have been carried out to study the impact of Orientation Programme on the professional development of teachers (Dhawan (2000), Rastogi Savita, (2001), Joshi Suresh and Pareek Sushila (2006), Pawar I. Anand and Mouli. S. Chandra, (2008), Survey and Bagul (2012), Satsangi *et al.* (2012), Sonawane and Survey (2015) and Khadilkar (2016)); however, there is a paucity of researches in terms of impact of different domains of the Orientation Programmes viz., teaching, ICT, research and soft skills and economic, social and environmental knowledge on teachers' knowledge, attitude and skills. Hence, the present study was carried out to assess the impact of Orientation Programmes on knowledge, attitude and skills of senior college and university teachers. The objectives of the present study were to assess the improvement in teachers in following domains after attending the Orientation Programme viz.,

1. Knowledge and practice of teaching skills and methodology;
2. Attitude of teachers towards students;
3. Research aptitude and skills;
4. ICT skills;
5. Life skills;
6. Knowledge of social, economical and environmental issues.

Materials and Methods

About 113 university and college teachers who attended Orientation Programme at UGC-Human Resource Development Centre of Rashtrasant Tukadoji Maharaj Nagpur University during the year 2001 to 2017 (20th to 98th Batches) were selected on the basis of judgemental sampling. The study was carried out during 2017. Survey method was used. A structured questionnaire was developed having questions related to personal characteristics of teachers and the core areas of Orientation Programmes. During the Orientation Programme, lectures, seminars, group discussion, panel discussion were organised on the topics stated in Guideline of UGC Academic Staff College (2002-07), (2007-12), and Guidelines for Regional Centre for Capacity Building & Human Resource Development Centre (2015) and Guidelines for UGC- Human Resource Development Centres (2019). Hence the questions included in the questionnaire were focused in the core areas of Orientation Programme. The teachers were asked to rate themselves on improvement in domains viz., knowledge and practice of teaching skills and methodology, attitude of teachers towards students, research aptitude and skills, ICT skills, life skills and knowledge of social, economical and environmental issues on five point Likert scale. The five point Likert scale had options viz., 'very low', 'low', 'neither low nor high', 'high' and 'very high'. The data was analysed for frequency, percent and chi square test. The confidence interval was set at 95 per cent.

Results and Discussion

Personal Characteristics of Teachers

The personal characteristics of the teachers studied were age, gender, educational qualification, faculty, location of the college and the teaching experience. The results are discussed here.About 67.9 per cent teachers were male and 32.2 per cent were female. About 42.7 per cent teachers were from the age group of 41 to 50 year and 39.1 per cent from 31 to 40 years. Only 7 per cent teachers were below 30 years. Majority (52.2 per cent) teachers were from Arts and Humanities whereas 16.8 per cent and 26.5 per cent were from Commerce and Science faculties respectively. Rest of 4.5 per cent teachers were from other disciplines. The majority teachers were from urban areas (54.8 per cent) and only 32.2 per cent from rural areas however 13 per cent teachers were from semi urban areas. There were 45 per cent and 43.5 per cent UG and PG teachers having teaching experience of up to 10 years whereas 20 per cent and 49.6 per cent UG and PG teachers having 10-20 years teaching experience respectively. Only 7.3 per cent teachers had teaching experience of more than 20 years.

Teaching Related Skills

During Orientation Programmes lectures and practical sessions were organised on effective teaching, teaching and learning styles, microteaching,

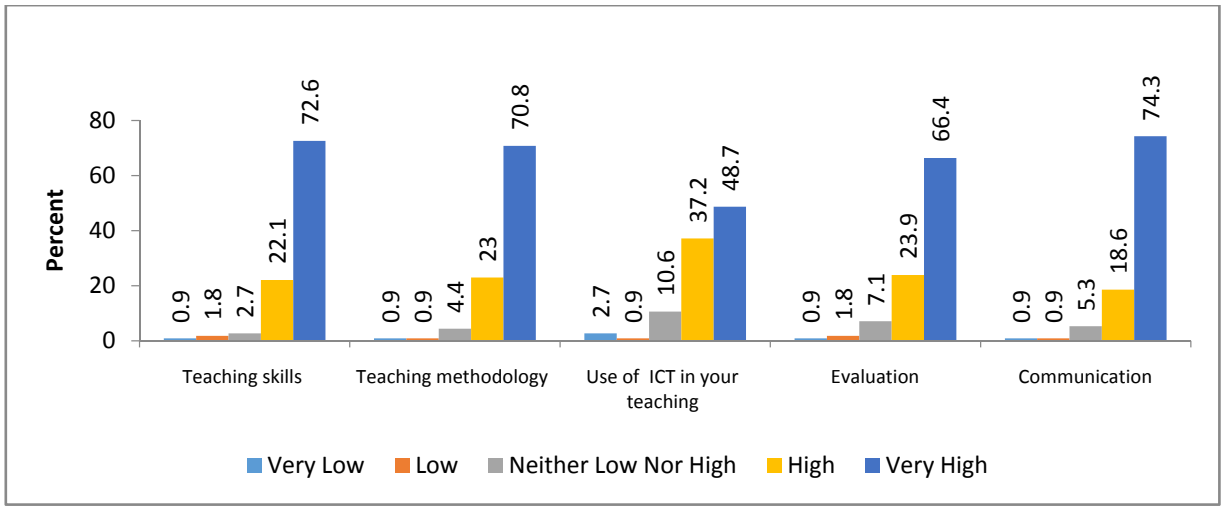
communication skills, use of ICT during teaching, learning and evaluation. The teachers were asked to rate themselves for the improvement in teaching skills, teaching methodology, use of ICT in teaching, evaluation and communication skills after attending Orientation Programme and results are presented in Figure 1.

The data presented in Figure 1 shows that the majority teachers rated themselves for 'high' and 'very high' for improvement in teaching related skills after attending Orientation Programmes. The highest number of teachers rated themselves 'very high' for improvement in communication (74.3 per cent) followed by teaching skills (72.6 per cent), teaching methodology (70.8 per cent) and evaluation (66.4 per cent). Slightly lesser percent of teachers (37.2 per cent and 48.7 per cent) rated themselves for high and very high for the use of ICT in teaching which may be because of those participants who attended Orientation Programme before 2005 and did not study computers during the course. The teachers, who rated themselves low and very low, ranged between 0.9 to 2.7 per cent. About 2.7 to 10.6 per cent teachers opted for neither low nor high for the improvement in their teaching skills.

The statistical analysis showed that teachers rating for their improvement significantly differs with respect to teaching skills (χ^2 = 212.796, p=0.000), teaching methodology (χ^2 =201.292, p=0.000), use of ICT in teaching (χ^2 = 105.717, p=0.000), evaluation (χ^2 = 171.204, p=0.000) and communication skills (χ^2 = 220.407, p=0.000). The study showed higher impact of Orientation Programme as compared to the reported studies. Satsangi *et al.* (2012) studied the impact of orientation and refresher courses on knowledge up gradation, motivation, leadership traits, technological information, communication skills, personal attributes, health awareness, stress management and national and social values on the participants. Authors reported that ASCs functioning of teaching in higher education in orientation and refresher programme were effective and its help participants in improving their teaching abilities. Similarly, Sonawane and Survey (2015) studied the impact of Orientation Programmes among the teachers. The study reported that orientation helped 10 per cent of the faculty to increase their abilities as a teacher, 9.33 per cent to improve their lecture delivery skills, 9 per cent to achieve improvement in practical and vocational skills, 8 per cent to improve their contents of teaching, only 7 per cent to improve their skills regarding student evaluation and assessment practice and 6.67 per cent of the student and classroom management skills.

Khadilkar (2016) carried out a study on role of faculty development programme and reported that out of total sampled faculty members from Maharashtra, 44 per cent of the faculty members strongly agree about improving communication skills, 50 per cent in improving teaching learning process, 44 per cent in improvement of lecture delivery, 50 per cent in improvement in contents of teaching, 36.67 per cent of the faculty members them in using innovative teaching pedagogy and 44 per cent in improving use

Figure 1: Percent Distribution of Teachers According to Improvement in Teaching Related Skills after Attending Orientation Programmes



of ICT in teaching, after attending the orientation program.

Understanding Student's Issues

Student counselling and guidance is an important core area which is covered during the Orientation Programme. In this section, the teachers were asked to rate themselves for understanding students' attitude towards learning, towards teachers, their psychological problems, peer group pressure, financial and employment problems as well as employment opportunities for the students and results are presented in Figure 2.

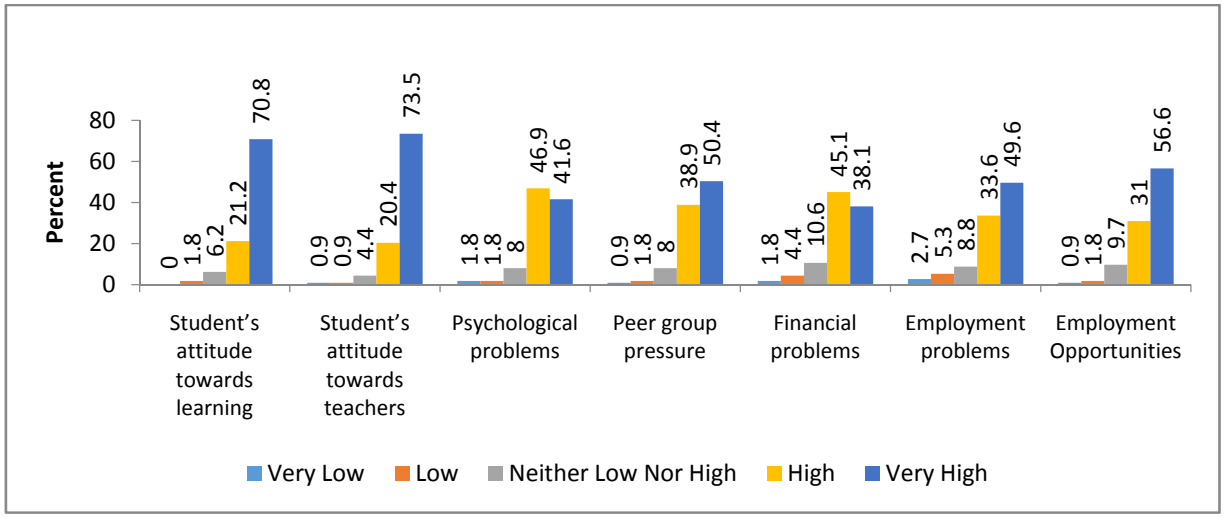
The data presented in Figure 2 indicates that the majority teachers rated 'very high' for themselves for their improvement in understanding student's attitude towards learning (70.8 per cent) and teachers (73.5 per cent), employment opportunities for students (56.6 per cent) and employment problems (49.6 per cent) after attending Orientation Programmes. The teachers understanding for peer group pressure were rated 'very high' by only 50.4 per cent teachers. In case of teachers' understanding of students psychological and financial problems the majority teachers rated themselves for 'high' (46.9 per cent and 45.1 per cent respectively) followed by 'very high' (41.6 per cent and 38.1 per cent). About zero to 2.7 per cent teachers rated for 'very low' whereas 0.9 to 5.3 per cent rated for 'low' scores. There were 4.4 to 8.8 per cent teachers who were indifferent.

The statistical analysis showed that teachers rating of themselves for improvement significantly differs with respect to understanding students' attitude towards learning (χ^2 = 135.814, p=0.000), towards teachers (χ^2 = 216.425, p=0.000), their psychological problems (χ^2 = 112.973, p=0.000), peer group pressure (χ^2 = 120.230, p=0.000), financial (χ^2 = 91.558, p=0.000) and employment problems (χ^2 = 96.071, p=0.000) as well as employment Opportunities for the students (χ^2 = 128.018, p=0.000). Several studies have been carried out on the impact of Orientation Programme on the professional development of teachers (Dhawan (2000), Rastogi Savita (2001), Joshi Suresh and Pareek Sushila (2003), Pawar I. Anand and Mouli S. Chandra, (2008), Survey and Bagul (2015), Satsangi *et al.* (2012) and Sonawane and Survey (2015)), however, none of the studies reported impact of Orientation Programmes on teachers with respect to understanding problems of their students.

Research Aptitude and Skills

Research in higher education help in discovery dissemination and interpretation of new knowledge, ideas, facts and motivate the teachers. All teachers should normally engage in research because research is essential to their professional identity (Satsangi et al.,2012). Research methodology and statistical analysis is an important core area which is covered during the Orientation Programme. The teachers are acquainted with research process, writing review of literature, bibliography, analysing

Figure 2: Percent Distribution of Teachers According to Improvement in Teachers' Understanding on Issues of Students after Attending Orientation Programmes



data and interpreting results, writing research proposals publishing research papers at national and international levels besides plagiarism, citation, statistical analysis and statistical tools, during the Orientation Programmes. In this domain, the teachers were asked to rate themselves for their improvement in research aptitude and skills after attending the Orientation Programme and the results are presented in Figure 3.

The data presented in Figure 3 shows that the majority teachers (73.5 per cent) benefitted 'very high' for searching review of literature whereas 67.3 per cent and 66.4 per cent teachers rated 'very high' for writing research proposals and research papers at national levels respectively. About 59.3 per cent, 54 per cent and 49.6 per cent teachers rated themselves very high for writing bibliography, writing research papers at international levels and analysing and interpreting data respectively.

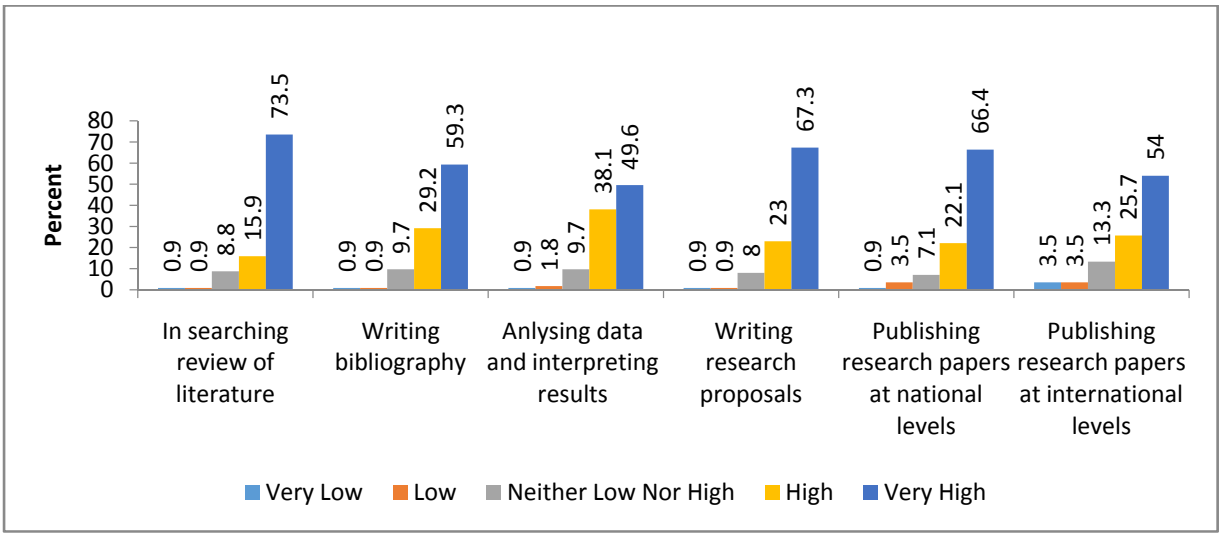
The statistical analysis showed that teachers rating of themselves for improvement significantly differs with respect to knowledge and skills of teachers viz., writing review of literature (χ^2 = 210.673, p=0.000), bibliography (χ^2 = 139.257, p=0.000), analysing data and interpreting results (χ^2 = 113.150, p=0.000), writing research proposals (χ^2 = 176.159, p=0.000), publishing research papers at national ((χ^2 = 167.133, p=0.000) and international levels (χ^2 = 100.230, p=0.000). Satsangi *et al.* (2012) studied the impact of orientation and refresher courses on teachers with respect to research aptitude of teachers viz., knowledge up-gradation, art of thinking, writing skills, technological information and research skills. It was concluded that, ASCs contribution towards inputs in research in higher education in orientation and refresher programme were effective and its help participants in improving their research skills.

Use of ICT Skills in Teaching

Information Communication Technology (ICT) is one of the important buzzwords of today's IT world. It has changed the society into information society and our way of life (Sinha, 2008). During Orientation Programmes, the teachers are acquainted with Microsoft Office. The teachers were given hands on in preparing power point presentation, using M.S. Excel for data analysis, using M.S. word for writing papers, using internet for searching reviews and communicating through emails as well as introduction to web 2.0 tools viz., Blogs/ You tube/ Face book. In this section, the teachers were asked to rate themselves for their improvement in use of ICT skills after attending the Orientation Programme and results are presented in Figure 4.

The data presented in Figure 4 shows that about 63.7 to 66.4 per cent teachers rated 'Very high' for improvement in preparing power point presentation, using M.S. word for writing papers, using internet for searching reviews and communicating through emails. Only 40.7 per cent and 47.8 per cent teachers rated "very high' for improvement in using MS

Figure 3: Percent Distribution of Teachers According to Improvement in Research Knowledge and Skills after Attending Orientation Programmes



Excel for data analysis and knowing web 2.0 tools viz., Blogs/ You tube/Face book respectively. The statistical analysis showed that teachers rating for improvement significantly differs with respect to preparing power point presentation (χ^2 = 172.973, p=0.000), using M.S. Excel for data analysis (χ^2 = 89.611, p=0.000), using MS word for writing papers (χ^2 = 104.876, p=0.000), using internet for searching reviews (χ^2 = 156.336, p=0.000) and communicating through emails (χ^2 = 87.389, p=0.000), and knowing web 2.0 tools viz., Blogs/ You tube/Face book (χ^2 = 59.080, p=0.000) after attending Orientation Programme.

Sinha (2008) studied Information Communication Technology (ICT) and internet awareness amongst the college and university teachers attending refresher course in ICT. The results showed that 90.24 per cent teachers were using IT in their classroom activities whereas only 9.75 per cent are not using IT for this purpose. 63.41 per cent teachers were comfortable in using MS-Office for their class work whereas 36.58 per cent were not able to use this. MS-Office was used frequently by 42.30 per cent teachers which are followed by 38.46 per cent for moderately use whereas 19.23 per cent were using MS-Office often for their class work. Internet literacy was very much poor among the respondents. Out of 41 participants only 36.59 per cent were Internet literate whereas maximum respondents (63.41 per cent) were Internet illiterate. While assessing the rating of Internet /e-mail services, e-mail ranks first which is followed by WWW (2nd), e-journals access (3rd), INFLIBNET Databases Search (4th) and Chatting (5th). For e-mail service, Yahoo (33.33 per cent) emerged as most preferred e-mail service which is followed by Indiatimes mail service (26.66 per cent), BSNL and Rediffmail (13.33 per cent) and 6.66 per cent for NICNET and Google mail service. 53.33 per cent respondents prefer www.goggle.com search engine which is followed by www.yahoo.com (26.66 per cent) whereas www.msn.com, www.rediffmail.com and www.indiatimesmail.com are preferred by only 6.66 per cent each.

Soft Skills

The Component D of the guidelines of ASC/HRDC states inclusion of topics related to management and personality development. During these Orientation Programmes, teachers were given training in soft skills like time management, team work, self analysis, self confidence, interpersonal skills, change management, personality development, communication skills, attitude, motivation level and stress management. Teachers were asked to rate the improvement in these soft skills after attending the Orientation Programme and the results are presented in Figure 5.

The data presented in Figure 5 shows that about 72.6 to 77.9 per cent teachers rated 'Very high' for improvement in soft skills viz., team work, communication skills, motivational level, personality development and self confidence. About 64 to 69 per cent teachers rated 'very high' for stress management,

Figure 4: Improvement in ICT Skills of Teachers after Attending Orientation Programmes

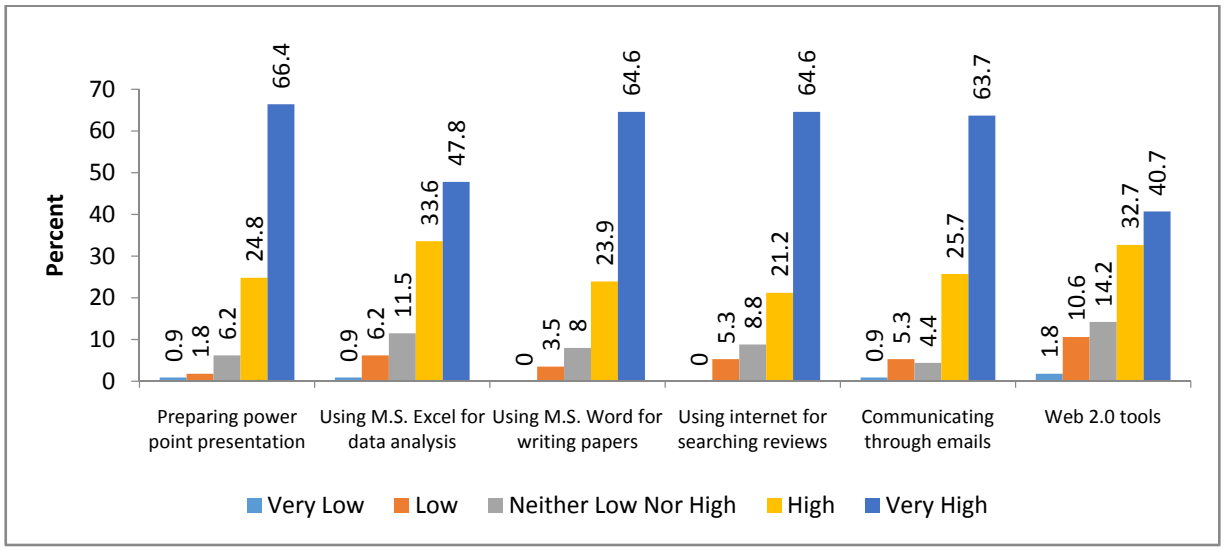
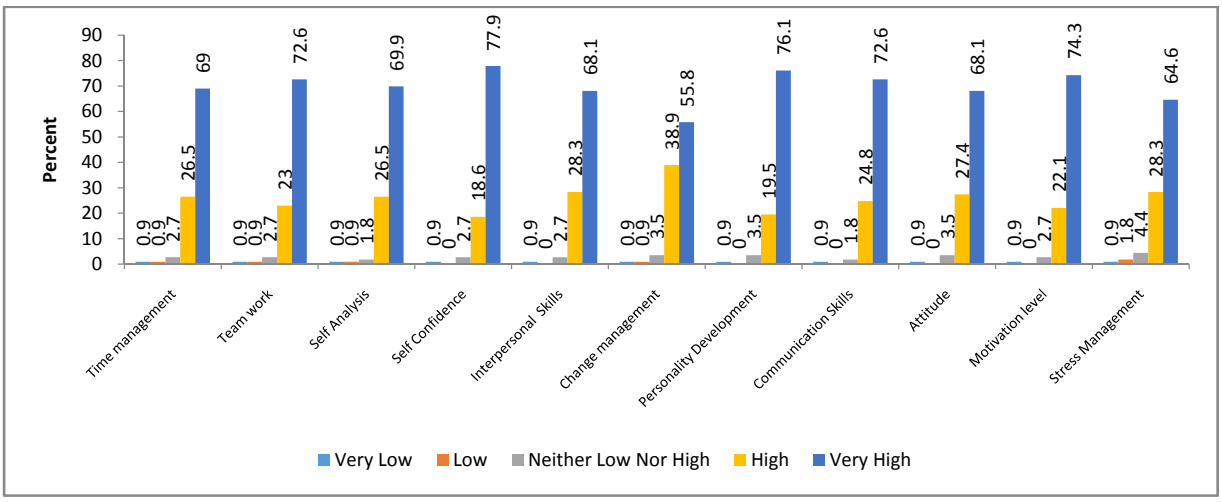


Figure 5: Percent Distribution of Teachers According to Improvement in Soft Skills after Attending Orientation Programmes



attitude, interpersonal skills, time management and self analysis. Only 55.8 per cent teachers rated for 'very high' for change management. About 0.9 to 1.8 per cent teachers rated for low and very low for improvement in all the soft skills after attending Orientation Programme. There were 1.8 to 4.4 per cent teachers who rated for 'neither high nor low' for improvement in all the skills after attending the course.

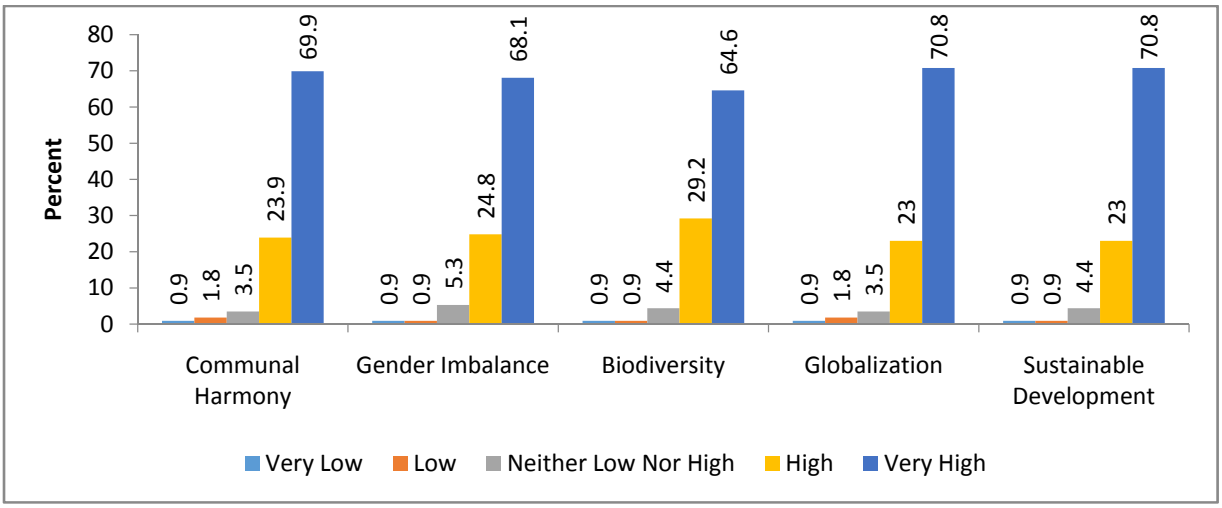
The statistical analysis showed that teachers rating for improvement significantly differs with respect to soft skills like time management (χ^2 = 196.513, p=0.000), team work (χ^2 = 214.920, p=0.000), self analysis (χ^2 = 203.239, p=0.000), self confidence (χ^2 = 177.088, p=0.000), interpersonal skills (χ^2 = 133.478, p=0.000), change management(χ^2 = 149.080, p=0.000), personality development (χ^2 = 166.540, p=0.000), communication skills (χ^2 = 152.947, p=0.000), attitude(χ^2 = 131.496, p=0.000), motivation level (χ^2 = 159.248, p=0.000), and stress management (χ^2 = 169.434, p=0.000) after attending Orientation Programme.

The extensive review of literature showed the paucity of data on improvements in soft skills of the teachers after attending Orientation Programme.

Knowledge on Social, Economical and Legal Aspects

During Orientation Programmes teachers are

Figure 6: Scores for Improvement in Teachers' Knowledge on Social, Economical and Environmental Issues after Attending Orientation Programmes



informed about social, economical, environmental and legal issues. The teachers were asked to rate themselves for improvement in their knowledge about communal harmony, gender imbalance, sustainable development, biodiversity and globalization and the results are presented in Figure 6.

The knowledge gained for social, economical and environmental issues during Orientation Programmes were rated 'very high' by 64.6 to 70.8 per cent teachers. About 70.8 per cent teachers each rated 'very high' for gain in knowledge in sustainable development and globalization whereas for communal harmony, gender imbalance and biodiversity 69.9 per cent, 68.1 per cent and 64.4 per cent teachers rated 'very high'(Figure 6). The statistical analysis showed that teachers rating for improvement significantly differs with respect to communal harmony (χ^2 = 196.336, p=0.000), gender imbalance (χ^2 = 185.717, p=0.000), sustainable development (χ^2 = 172.177, p=0.000), biodiversity (χ^2 = 201.027, p=0.000), and globalization (χ^2 = 201.292, p=0.000).

Satsangi *et al.* (2012) also studied the impact of Orientation Programmes on personal attributes, health awareness, environmental awareness, stress management and national and social values and concluded a positive impact on teachers. Sonawane and Survey (2015) reported that 5.67 per cent of the faculty members were positive about the statement that the orientation helped them in improving the interdisciplinary subject knowledge. Khadilkar (2016) also reported that out of total sampled faculty members from Maharashtra, 11.33 per cent of the faculty members strongly disagree, 8.67 per cent of the faculty members disagree, 4 per cent of the faculty members are neutral, 26 per cent of the faculty members agree and 50 per cent of the faculty members strongly agree about the orientation program helped them in improving social awareness.

Conclusion

It can be concluded from the present study that Orientation Programmes helped the teachers to improve their knowledge and practice of teaching skills and methodology, attitude of towards students, research aptitude and skills, ICT skills in imparting education, life skills and knowledge of social, economical and environmental issues. The improvement in all domains was found to be significantly (p=0.000) "very high" in likert scale in all the domains. Imparting knowledge, and skills through Orientation Programme may go a long way in improving the quality education through well trained teachers.

References

1. Dhawan Rani, (2000). Impact of Academic Staff College's programmes on Teachers and Students, *University News*, 38 (16) April 17, 2000.
2. Guidelines for Academic Staff Colleges, University Grants Commission (2002-2007) X Plan.
3. Guidelines for Academic Staff Colleges (2007-12) XI Plan.
4. Guidelines for Regional Centre for Capacity Building (RCCB) & Human Resource Development Centre (HRDC) (Implementable from 01 February, 2015) .
5. Guidelines Human Resource Development Centre (HRDC), (2019).
6. Hénard, Fabrice (2010). Learning our lesson. *Review of Quality Teaching in Higher Education.* https://www.researchgate.net/publication/44841258_Learning_our_lesson_Review_of_quality_teaching_in_higher_education
7. Joshi Suresh and Pareek Sushila, (2006). Impact of Orientation Courses on teachers in higher education. *Indian Journal of Adult Education*, New Delhi, Jan-June 2006.
8. Khadilkar Sujay Madhukar, (2016). Role of Faculty Development Program, *Scholarly Research Journal for Interdisciplinary Studies*. 3(22), 1655-1671
9. Pawar I Anand and Mouli S Chandra, (2008). Impact of training on university and college teachers: An Empirical Study. *University News*, 46 (49), 08-14, 2008.
10. Rastogi Savita (2001). Teachers Development in Higher Education in Academic Staff Colleges in India, *www.Google.co.in*. September 2009.
11. Sachin Sadashiv Surve and Dhananjay Bagul, (2015). The Orientation Programmes - Challenges Ahead. *Scholarly Research Journal for Interdisciplinary Studies*. 3(19), *375-39.*
12. Satsangi Falguni, Bhatnagar Sanjeev and Kaushik S P, (2012). An Evaluative Study of Functioning of Academic Staff College in India and Its Impact, Ph.D. thesis, Department of Management, Faculty of Social Sciences, Dayalbagh Educational Institute, Dayalbagh, Agra, https://shodhganga.inflibnet.ac.in/handle/10603/14183
13. Sheikh Younis Ahmad, (2017). Higher Education in India: Challenges and Opportunities. *Journal of Education and Practice*. 8(1), 39 - 42.
14. Sinha, Manoj (2008). Information Communication Technology (ICT) and Internet Awareness amongst the College and University Teachers. 6th International CALIBER-2008, University of Allahabad, Allahabad, February 28-29 & March1, 2008 © INFLIBNET Centre, Ahmedabad .https://www.researchgate.net/publication/269334809
15. **S**onawane Sanjeev A, and Surve Sachin Sadashiv, (2015). Professional Development of Academic Staff in Higher Education: Impact of Orientation Programmes, *Scholarly Research Journal for Interdisciplinary Studies*, 3(20), 1149-1157. ❑

Good Practices in Virtual Presentations for Higher Education

Pranita Gopal*

The COVID pandemic has forced educators tap into the potential of technology. On the face of it, technology appears to be quite useful in breaking the barriers of location and time but it steals away the infectious enthusiasm that one experiences in a live classroom. Educators today are relying heavily on virtual modes of presentation to help students and colleagues around move past these strange times. This paper attempts to understand virtual presentations and how educators can benefit from delivering engaging virtual presentations for their students.

Purpose of a Presentation

All presentations fall into three categories:

- *Inspirational:* The best example to illustrate an inspirational presentation is to think of TED Talks (www.ted.com). The inspirational presentations focus on certain problems and thereafter leave the audience with ideas and questions to think and ponder about. Using the audience's emotions and intellect, inspirational presentations put forth viewpoints that have the ability to inspire.
- *Informative:* Informative presentations on the other hand aim at sharing information – they are expected to communicate instructions, or deliver a point or share a research or simply educate the audience regarding areas of interest to either the organization (business or institution) or the audience. Educative presentations, training presentations could fall under this category.
- *Influential:* These presentations could also be called persuasive presentations where the presenter expects to the audience to accept the proposal or idea at hand or accept the solution presented to a particular problem. Sales pitch or business proposals could be an example of an influential presentation.

In education, the presentations draw heavily from the genres of inspirational to the informative presentation.

In higher education we have another lens to view a presentation. Presentations could be considered as opportunities:

- *to contribute*: where the presenter has the ability to contribute towards a discussion or a problem-solving process or the learning journey of a fellow learner.
- *to connect*: where the presenter has the ability to make connections with others present. It helps the presenter network with like-minded people and grow in their respective fields.
- *to change*: the perception of the audience towards issues or problems.

Face to Face Presentations Vs Virtual Presentations

Gendelman (2010) explains the difference between face to face presentations and virtual presentations under the following four heads:

- *Location & Time*: In face to face presentation, the location and time are fixed and the audience and the presenter are expected to be at a certain physical location at the same time, while in the case of a virtual presentation there are opportunities for the presentation to be happening synchronously (where the audience and the presenter are present virtually at the same virtual place & at the same time (irrespective of the time zones) or asynchronously (where the audience and the presenter can break away from the barriers of time and place).
- *Size of the audience*: Face to face presentations have a physical limitation with regard to the size of the audience and the logistics involved thereafter. In virtual presentations, the size of the audience is not gauged by the physical constraints of space but that of the internet bandwidth and technological platform.
- *Content delivery tools*: Often in face to face presentation, presenters are limited by the tools they can use to deliver the content (projector, LCD, video projector) as often it is difficult to create a complete multimedia experience in all presentation venues. With virtual presentations the choice of content delivery tools increases with options of screen sharing and using the internet simultaneously within the presentation. Using break-out rooms within the presentation framework allows for short group discussions without disturbing the entire group.

**Instructional Coach & Educational Consultant, O 1090 Devender Vihar, Sector 56, Gurgaon, Haryana. Email: pranitagopal@gmail.com*

- *Modes of Communication*: In face to face presentation, the mode of communication is mostly audio, that can create accessibility issues. Using closed captions in a virtual presentation can help break down the accessibility barriers.

Challenges in a Virtual Presentation in Higher Education

In spite of having advantages like, inclusion of large number of participants, lower economic costs for organizers, virtual presentations come with their own challenges. Gendelman (2010) list down the following challenges:

- *Absence of visual/ facial cues*: As a face to face presenter, we all rely on facial cues to give us reinforcements while we are presenting. In the world of virtual presentation, there is generally a lack of visual cues with regard to the audiences' mood and receptivity and these effects the enthusiasm of the presenter.
- *Struggle in building rapport with the audience:* Virtual presenters often struggle in building rapport with their audience due to internet bandwidth issues visible via the poor-quality audio or video transmission.
- *Internal distractions:* As virtual presentations are conducted using the laptop or mobile device, the presenters are always facing challenges of participants getting distracted by the emails or chats or other websites.
- *External distractions:* The audience of virtual presentations could be either logging in from their homes or office spaces where simultaneously other activities could created is tractions for the presenter and the audience. Also, another example of an external distraction could stem from the unmuting of the audio function in the device leading to distractions from the barking of pets to squabbles amongst family members.
- *Struggle in encouraging students:* In face to face presentations, the presenter can use visual cues and facial expressions to create a rapport with the participants. In a virtual world, often the audience has the ability to login to a presentation and switch off the video feed, leaving very little room for the presenter to engage with such participants.

Minimizing the Challenges while Presenting Virtually

These challenges could affect the communication channel, but with a bit of planning and conscious efforts, one can overcome the challenges, like for example:

- *Keeping the presentation short*: Unlike face to face presentations that could go on for an hour or more, virtual presentations need to be short with ample room for interaction. Although there is no fixed formula to ensure the distribution of presentation time vis a vis the question-and- answer time, a general rule of thumb followed by good presenters is to use only 60 per cent of the allotted time for presenting and the rest 40 per cent of the time for questions.
- *Using a brisk pace*: Unlike face to face presentations, in the virtual presentation, maintaining a brisk pace ensures the presentation reflects the enthusiasm of the presenter. A brisk pace doesn't mean to speak fast, instead it means that the presenter should be so well versed with the content that the presentation flows like an oft repeated story.
- *Keeping the interaction alive during presentation*: Interactivity is a key element to create memorable virtual presentations. Interaction keeps the audience actively engaged away from the various distractions.
- *Using more visuals than texts*: Using more visuals than texts helps keep the attention of the students. They ensure the presentation doesn't become dull. Free copy right images and images licensed under the creative commons can be found in sites like Unsplash, Pixabay, Wikicommons.
- *Add FIRES to the presentation:* Virtual presentations that have the necessary FIRES (F-Fresh, I – Informative, R- Relevant, E – Enthusiastic, S-Story) have the ability to engage the participants. In the case of higher education, the presenter can mix and match audio and visual clips to add freshness to the topics.

Planning & Delivering a Virtual Presentation

Every presentation has a certain purpose and that purpose is built on understanding the elements of a good presentation. Research has shown there are four elements to effective presentation:

- *Understanding the audience*: Concepts across disciplines and grade levels often overlap. Understanding the audience – both students and teachers- includes understanding the age, the role of the audience, socio-economic background, linguistic diversity, technological capability, accessibility issues etc. Understanding the audience provides the presenter a perspective to deal with the content and leads to the next important element of the presentation.

- *Preparing the content*: If we wish that our presentation is interesting and our audience is attentive during our presentation, the content needs to mirror audience needs and interests, along with the coherence of content. Students have access to all material that is available on the internet, what they lack is the ability to string the content into a tapestry that is meaningful and age appropriate.
- *Delivering confidently*: Confidently delivered presentation have "hooks" that engage the participants and create room for inter activity between the students and the teacher. One of the hallmarks of a confidently delivered presentation is linking what is being discussed to the larger context of the subject and demonstrating how small pieces of concepts have shaped the domain knowledge.
- *Controlling the environment*: Creating ground rules for the presentation, giving clear instructions on acceptable behaviors in an online environment, letting the participants know when their questions would be answered are some ways in which the presenter can control the presenting environment.

Durate (2012) describes three phases of planning and delivering a presentation. This holds true for even virtual presentations.

Phase I: Conceiving the presentation: Conceiving a presentation includes finding answers to the following questions:

- Who is the audience?
- What do I want the audience to know after hearing the presentation?
- What do I want the audience to do after hearing the presentation?
- What do I want the audience to feel after hearing the presentation?
- How does the presentation answer the question of "*WIFM – What's in it for me*?" for the audience?
- How does the presentation organize the content? Does it follow a chronological order or causal-effect or does it build on research or does it present a logical argument?
- Does the presentation flow like a story?
- What are the media modes that have been chosen for communicating the message?
- Are there alternatives for images so that accessibility issues in the audience is catered for?

Phase II: Visualizing the presentation: It is in this stage the presentation takes the form of a narrative with the help of slides or any presentation modes. While visualizing the presentation it is useful to plan answer the following questions:

- Are the visuals used free from cultural, gender or socio-economic barriers?
- Are the visuals such that they easily convey the point that the presenter is trying to make?
- Does the image help to make the presentation memorable?
- Does the image supplement the content and the topic that is being presented?
- Are the visuals too bulky in terms of their size?

It is important to remember that unlike face to face presentations, in virtual presentations, the presenter has to plan for more slides so as to keep the audience engaged.

Phase III: Delivering the presentation: There are two aspects that Durate (2012) suggests we consider at this stage – one pertains to delivery and the second pertains to the impact of the presentation. It is important to consider the following points while delivering the presentation:

- Is there enthusiasm in the presentation?
- Is the voice modulated for driving the maximum impact?
- Depending on the content and audience is the presenter's voice communicating assertiveness or caution or is it critical or is it motivational?
- Is there a way to measure the impact of the presentation?

Virtual Presentation-Good Practices for Higher Education

There is an analogy that one often reads about when a presenter's face to face and virtual presentation skills are compared. The analogy comes from the world of aviation – where pilots are not only trained but expected to have dual proficiency – first is the proficiency to fly with sight and what is visible in front of them (akin to face to face presentations where the audience is visible) and the second is the proficiency to fly with just the instruments in the cockpit – that is without any visibility of the world outside (akin to virtual presentations – where the presenter is facing just a camera and the computer screen). And therefore, like how pilots gain competency and dual proficiency following certain good practices can help a presenter gain proficiency in presenting virtually as well as in face to face environments.

- *Have a wingman*: A wingman is a supporter; a person who will assist the presenter in many ways – like to present the introduction, to monitor the chat, to monitor the questions, to steer the flow of questions to the presenter, to be have a copy of the presentation. The wingman can also keep providing useful summaries at the end of each section, giving the presenter a break from continuous speaking. If the planned presentation is of a longer duration, then the wingman can help break the monotony of a single presenter.
- Giving students the opportunity to be a wingman during the presentation, helps boost the student's confidence and creates a sense of responsibility within the student towards her learning.
- *Understand the audience*: A good practice to deliver effective virtual presentations revolves around understanding the audience composition – which institutes they belong to, what is their educational background, what is their work experience – all this information can be obtained when the audience registers for the presentation. This information also helps the presenter plan and share specific examples pertaining to certain specific areas of interests of the audience – automatically increasing the interaction amongst the participants.
- *Create a presentation title & description that will intrigue/ appeal the audience*: Presenters need to plan their topic to center around areas of interest of the audience or current need. Like for example, in today's pandemic times, any presentation that helps final year students increase their chances of getting employed will automatically appeal to a large majority of the students. In the same manner, presentations that deal with the mental health issues that students undergo due to lockdown will help students cope better as human beings.
- *Understand the presenting environment*: Virtual presenting environments have the following common features: a *presentation area* (where the presentation is available for display to the audience) , a *picture area* (where the presenter's video feed is available), a *chat area (*where the audience can chat either with the presenter or amongst themselves), a *break-out room (*where the larger groups can be broken into smaller groups for small group interactions), a *polling area (*where the presenter can create polls to understand the audience), a *screensharing area* (where the presenter can share the computer screen – this area could be overlapping with the presentation area), a *whiteboard* (which can be used during the presentation), the *audio/ video controller (*which is used to control the audio levels and the video clarity of the presentation.) With so many areas under one screen, it is important to understand the presentation environment as all apps and products have the same functions in different areas of the screen. \
- *Test the platform once before going live*: As presentation environments could vary across operating systems and amongst themselves, it is a good practice to test the presentation on any new platform before the final presentation. Also, when the presentation is tested before the final presentation, the presenter is able to gauge how the presentation fonts and images will be visible when the final presentation is screened across devices. When a presenter tests the presentation before delivering, it displays professionalism and behaviors that we wish the next generation imbibes.
- *Remember -Lag time*: Due to the internet bandwidth issues a virtual presenter needs to be aware of the lag time that could occur while presenting. The lag time is often witnessed by the audience as a slow-motion picture.
- *Avoid scroll quickly*: As many times, a lag time occurs during virtual presentations, it is important to remember that while viewing websites or programs via screen sharing, one must not scroll through the page quickly.
- *Avoid Animations*: In face to face presentations, animations if used correctly can help hold the attention of the audience. In a virtual presentation, they generally do not provide the same assistance as they could add to the lag during the presentation. Animations during virtual presentations could create unwanted distractions while moving between slides.
- *Present information in short logical chunks*: Presenting information in short logical chunks will help the presenter maintain the momentum of the presentation. It will assist the presenter move from one topic to another with confidence and agility that are hallmarks of a good presentation.
- *Using visuals*: The visuals in the presentations act as anchors that help the presenter and the audience. In a virtual world, the images help carry the talk forward. Visuals are great tools for creating transitions within the presentation topic and help the presenter support learners with diverse learning styles.

- *Promote interaction*: Polls are very useful tools to promote interaction. Polling tools can also transform into tools for checking the perception of the audience, understand the conceptual clarity of a topic. These questions can then be used as hooks to grab the attention of the audience.

  The wing man can also play an important role in adding to the interaction of the presentation by asking planned questions that help propel the discussion/ presentation ahead.

  Integrating live results or direct poling sites within a presentation can also help add to the inter activity within the presentation.

  The polling option can also be used as a tool to see the perception of the audience before and after the presentation.
- *Answer questions*: A good practice is to clearly set the stage as to when the questions would be answered. Many presenters prefer the questions are answered at the end of the presentation. Also, another good practice is to repeat the question and the name of the person who asked the question to help others understand the question in perspective. At times to add to the interactivity within the presentation, a good idea could also to be ask leading / investigative questions to the audience and then break them into smaller groups to think about the answer.
- *Keep a glass of water handy*: Like all presentations, when the speaker speaks for a longer duration, having a sip of water helps maintain the clarity of speech.
- *Do some vocal exercises*: In Julian Treasure's TED Talk (How to speak so that people want to listen) there are some exercises that help the speaker warm up her voice. These simple exercises add to the clarity of speech that helps the speaker keep the attention of the listener.
- *Reboot the computer*: Often we do not realize the number of background activities that take place in the computer and many at times the computer could update itself and start rebooting right when the presentation is on. Therefore, it is generally suggested that the computer/laptop be rebooted once before a presentation.
- *Upload supporting material*: It is also recommended not to share the presentation handouts before the presentation, instead the material shared once the presentation is over. Many presenters also share the presentation handouts/ supporting material once the presentation is over but the question and answer session is still on, giving the audience and students an opportunity to ask their questions before the event is wrapped up.
- *Gather feedback*: In the face to face situations, one is able to gather feedback based on the interaction after the presentation is over. In virtual presentations, the presenter has the provision to solicit specific feedback via forms and questionnaires. The feedback always assists the presenter in either collaborating with other like-minded colleagues or helps improve on their own presentation.

To Sum Up

In the UGC Guidelines on Examinations and Academic Calendar for the Universities in View of COVID-19 Pandemic and Subsequent Lockdown, issued in April 2020 a case for blended learning model has been put forward. The guideline also emphasizes that the responsibility of the teacher doesn't end by just sharing the e-content, instead it lays emphasis for interaction with the students.

Presenting in the virtual world and working with students and colleagues via a virtual medium necessitates that interactivity and engagement are deeply rooted within all presentations and discussions. With adequate planning and following the good practices from world over virtual presentations can be a great scaffold for in the learning journey of individuals.

References

1. Campbell, G M, (2002). Bulletproof Presentations. Career Press. New Jersey.
2. Duarte, N (2012). HBR Guide to Persuasive Presentations. Harvard Business Review Press.
3. Duarte, N (2019). Data Story: Explain Data and Inspire Action Through Story. Ideapress Publishing.
4. Gendelman, J, (2010).Virtual Presentations That Work, McGraw-Hill Education, New York.
5. Reynold, Garr (2012). Presentation Zen: Simple Ideas on Presentation Design & Delivery. 2nd Edition. Peachpit.
6. Reynold, Garr (2011). Naked Presenter. Peachpit. California.
7. Treasure, Julian (2013). TED Talk - https://www.ted.com/talks/julian_treasure_how_to_speak_so_that_people_want_to_listen.
8. University Grants Commission, (2020). UGC Guidelines on Examinations and Academic Calendar for the Universities in View of COVID-19 Pandemic and Subsequent Lockdown. From: https://www.ugc.ac.in/pdfnews/5369929_Letter-regarding-UGC-Guidelines-on-Examinations-and-Academic-Calendar.pdf. ❑

Transformations through Innovative Technological Interventions

M Venkaiah Naidu, Hon'ble Vice President of India delivered the Convocation Address at the First Convocation of National Institute of Technology, Tadepalligudem, Andhra Pradesh on December 24, 2019. He said, "The challenge I place before all of you as young technocrats today, as you embark on the next phase of your life, is to find technological solutions to combat the pressing problems. Incremental improvements and frugal innovations can be carried out constantly to bring about advancements. By solving these problems, you will create a virtuous cycle which will help pull millions out of poverty and improve the quality of their lives. Please remember that that the ultimate aim of science and technology is to make lives of the people comfortable and happy." Excerpts

I am delighted to be with all of you today at the first convocation ceremony of NIT Andhra Pradesh, a young and promising institution nestled in the pristine land of Tadepalligudem. A visit to Andhra Pradesh is like a homecoming to me. I was born here and was nurtured and nourished by this great land.

National Institutes of Technology (NITs) that have evolved from Regional Engineering Colleges (RECs) have been at the forefront of technical education in the nation. NIT AP had the privilege of being mentored by NIT Warangal, the first REC in the country. I am happy to know that NIT Andhra Pradesh is the 31st NIT in the nation. I am told that it started functioning with the support of its mentor institute, NIT Warangal, in Sri Vasavi Engineering College Campus, Tadepalligudem. I fondly remember the foundation laying ceremony of NIT Andhra Pradesh which was held on 20 August 2015 when I was the Union Minister for Urban Development and Parliamentary Affairs. It is truly noteworthy that the construction work of the campus which started in October 2018, was completed in thirteen months. The transformation of an erstwhile runway to an awe-inspiring campus of an institution of national importance is indeed remarkable. It is clear from the report presented by the Director that the youngest NIT of the country has made great strides in all areas.

My Dear Young Students, let me take this opportunity to congratulate each and every one of you for this tremendous accomplishment. You have the proud privilege of being the first batch of students to graduate from NIT, Andhra Pradesh. NIT Andhra Pradesh was born with you and will grow with you and prosper with you. As you leave the portals of this Institution today, please remember that you are amongst the brightest young minds of our country. You live in a very exciting era of scientific and technological disruptions.

Technological breakthroughs in the form of Automation, Artificial Intelligence, Internet of Things, Big Data and analytics have truly transformed the way we live and work. The existing frontiers of science are being constantly challenged through interaction between various disciplines ranging from arts to humanities to engineering to Biotechnology. Young engineers like yourself have the never-before opportunity to further revolutionize these technologies and find new ways to use them for the benefit of humanity. Let me stress that the ultimate aim of all these technological advances must be the betterment of the life of the common man. They must lead to the discovery of solutions to the most pressing problems of our time.

There is no doubt in my mind that innovation is the watchword for the 21st Century. Institutions such as the IITs and the NITs must become hubs of innovation. They must recognize potential and nurture it to bring out the best in each student. The curricula and teaching methods in these institutions must never be set in stone. They must constantly evolve to suit the times. Institutes that render technical education must never hesitate to teach and experiment with the latest of technologies. They must allow the free flow of ideas and must never stifle innovative thinking through rigid protocols. Let me focus on a few areas that I think can be transformed through innovative technological interventions.

The first is agriculture. There is a vast scope to introduce new and innovative technologies to spur agriculture growth and improve the lot of the farmers. This is the time for young engineers to come up with out-of-box solutions to the problems faced by agriculturists from climate change to increasing crop productivity. I urge young technocrats to understand the real problems faced by Indian agriculture, including fragmented land holdings and insufficient market access.

I want you to think:

- Can we understand soil better?
- Can we carry out more efficient irrigation by overcoming shortages and avoiding surpluses?
- Indian agriculture is greatly dependent on weather. Keeping the changing climate and precipitation change in mind, can we create better weather prediction systems and make agriculture become more resilient?
- Can we facilitate better price discovery and market access to farmers?

It is of paramount importance that we boost our food production to overcome hunger and feed a billion people. I have always maintained that imported food security is never a solution. The second is the arena of clean and renewable energy. Our climate patterns and weather systems are changing. Global Warming is a reality and it is happening now. We cannot be in denial anymore.

India is one of the most carbon-efficient economies in the world today. The government of India under the leadership of our visionary Prime Minister is all set to surpass the target we had set for ourselves of 175 GW of renewable energy by 2022. Our Solar capacity has increased by eight times between FY14-18 and we have spearheaded the International Solar Alliance (ISA), a testimony to India's commitment to the development of 'zero emission sources'. I want engineers and technocrats like you to champion the cause of clean energy.

I strongly believe that a balance between environment and development can be achieved through good technology. It is essential that we strike that balance. Our development has to be sustainable. Let conservation be the cornerstone of every single one of your innovations.

The third is urban development. India's urban population is expected to rise to 517 million by 2020 and will cross the 700 million mark by 2050. Our urban spaces are our engines of growth. Unfortunately, our towns and cities have exceeded their carrying capacity and are choking. Our most prosperous cities are also home to large slums. Our cities must grow inclusively and sustainably. We must find solutions for sustainable housing, drinking water, mass mobility and provide economic opportunities to bridge the huge income gaps in cities. We must address concerns of pollution and congestion and create sustainable waste management solutions and waste-to-wealth initiatives. We must make our urban spaces safe for women and children and accessible to the differently-abled.

We must check distress migration by providing quality healthcare, education, other amenities, and economic opportunities in rural areas as well. Urban development is possible only if we achieve rural development, a cause that is very close to my heart.

Mahatma Gandhi also believed that self-sufficient villages are necessary if we are to build a prosperous nation. Our rural areas too need technological interventions that are tailor-made to suit their unique requirements. I also urge you to find solutions to challenges of water management. Many parts of India are facing an acute water crisis. This crisis will only worsen in the time to come.

According to the Composite Water Management Index (CWMI) report released by the Niti Aayog in 2018, 21 major cities might reach zero groundwater levels by 2020. The Union government recently formed a new Jal Shakti (water) ministry, which aims at tackling water issues with a holistic and integrated perspective. The ministry has announced an ambitious plan to provide piped water connections to every household in India by 2024. The government alone will not be able to solve the entire problem. We must augment the efforts of the government.

We must use technology for better rainwater harvesting and to improve water use efficiency. We must urgently transit from this 'supply-and-supply-more water' provision to measures which lead towards improving water use efficiency. The challenge I place before all of you as young technocrats today, as you embark on the next phase of your life, is to find technological solutions to combat these pressing problems. I am hopeful that you will come up with solutions to these challenges through hard work and creative thinking. You do not have to wait for big-ticket innovations. Incremental improvements and frugal innovations can be carried out constantly to bring about advancements. By solving these problems, you will create a virtuous cycle which will help pull millions out of poverty and improve the quality of their lives. Please remember that that the ultimate aim of science and technology is to make lives of the people comfortable and happy.

My Dear Youngsters, In spite of temporary global slowdown, the Indian economy has the potential to become a $ 5 economy in the coming years as the macroeconomic fundamentals are strong and resilient. Government is also taking various reforms from time to time to improve GDP growth, cut down red tape and attract more investments.

A bright future awaits all of you. All the stakeholders in the country, particularly youngsters like you, must strive to transform the nation and the lives of the people in tune with the three-word mantra of the Prime Minister, Shri Narendra Bhai Modi to *'Reform. Perform and Transform'.*

Even as our technical education institutions such as NITs equip young technocrats with such skills, also take care to nurture life skills and moral values in them. Always remember the words of Mahatma Gandhi. He said that education without character is a sin. Every graduate passing out of the portals of this institution must not only be academically proficient, they must also be ethical, compassionate and honest individuals.

Before I conclude, I must raise one more issue. I read with mounting concern reports about young students resorting to extreme measures because of the inability to handle stress and cope with challenges. Our campuses must be the safest spaces in our nation, for in our campuses, our future is cast. I urge our higher educational institutions to take extra care to teach our youngsters to manage stress. I understand that the curriculum is heavy. But students must be taught to take time out. We must make sure that our students are physically fit and mentally happy.

My Dear Young Friends, never hesitate to reach out to your friends, teachers or loved ones when you are in need of help or support. Be optimistic and embrace the challenges that await you in your life and career with a balanced mind. The practice of Yoga and meditation will greatly help you to deal with stressful situations with poise and equanimity. I am confident that you will make the impossible possible with your grit and determination.

Let me congratulate all of you once again. Go forth and create a bright, beautiful future for you, our nation and the world. ❑

ATTENTION : SUBSCRIBERS UNIVERSITY NEWS

The NEW RATES of Subscriptions effective April 01, 2020 shall be as per following:

	Institutions Rs.	Teachers/Students/Individuals* Rs.
1 year	**1,250.00**	**500.00**
2 years	**2,200.00**	**900.00**

***AT RESIDETNAIL ADDRESS ONLY**

The payable amount is required to be remitted **in advance** by any of the following modes of payments:

a) AIU WEB Portal (b) Cash Deposit (c) Demand Draft/At Par Cheque and (d) NEFT/RTGS/Net Banking/G-Pay/ BHIM APP, etc.

1	**Bank Account No.**	**0158101000975 (Saving)**
2	Beneficiary Name and Address	**ASSOCIATION OF INDIAN UNIVERSITIES**
		16, Comrade Indrajit Gupta Marg, New Delhi – 110 002
3	Bank & Branch Name	CANARA BANK, DDU MARG
4	Bank's Address	"URDU GHAR", 212, Deen Dayal Upadhyaya Marg, New Delhi – 110 002
5	Branch Code	0158
6	IFSC Code	CNRB 0000158
7	Contact No. & E-mail ID	(011) 23230059 Extn. 208/213 (M) 09818621761

THE NEFT/RTGS/ONLINE PAYMENT TRANSACTION/UTR NUMBER MUST BE SENT BY MAIL IMMEDIATELY WITH COMPLETE MAILING ADDRESS & PIN CODE FOR LINKING AND ITS SETTLEMENT AT OUR END.

For further information/enquiries, ***send Mail at : subsun@aiu.ac.in / publicationsales@aiu.ac.in***
Website : ***https://www.aiu.ac.in***

COMMUNICATION

Innovative Initiatives Taken for Reforming Conduct of Education During COVID-19 Pandemic Lockdown Period

Sajal Dasgupta*

We are going through a very difficult situation now due to this Corona virus invasion, COVID 19 pandemic. Regular traditional teaching learning process (face to face) has been completely halted. We do not know when things will normalize. Fortunately, we had taken lots of exemplary initiatives in last two years to go digital in line with Digital India Initiatives launched by Government of India. In the month of Feb end 2020 when we realized the seriousness of the situation and anticipated a nationwide preventive drive is going to be implemented shortly to encounter the outbreak , our statutory bodies decided to confront the threat, crisis and panic created due to COVID pandemic with a positive attitude of transformational leadership to avail its opportunities and become self reliant through a paradigm shift from offline to online mode of education in such a manner so that there is no loss of available academic hours as per our academic calendar being followed strictly before the pandemic period. The very objective of these initiatives taken was not to reinvent education but to continue it uninterruptedly with a different methodology.

As per the directive of the Government we had notified to suspend all face to face classes from 16th March 2020 and closed down the University on and from 23rd March 2020. Since then all our employees including faculty are locked down in home as per the decision of the govt. to protect against Covid-19 pandemic. We firmly believed that if we delay the teaching learning process and examination and declaring results, enormous damage will be done to the families who are waiting for their child to financially supplement family income. About 90 percent of our final year students have got job offers. Families of these students are looking forward eagerly for time when their ward will start supplementing family income. Employer company will not wait further.

**Sajal Dasgupta, Vice Chancellor, University of Engineering and Management, Kolkata, West Bengal- 700160, E-mail: dasguptasajal@yahoo.com.*

In order to maintain our academic calendar unaffected and to protect the career of our students, we decided to continue our teaching learning process through digital platform. Around 2000 lectures done by our teachers were videotaped. Our teachers and students collectively practiced mock examinations on digital platforms. University authority was committed to ensure no students loose time due to corona virus invasions.

Accordingly, since 16th March 2020 all our faculty were sincerely engaged in classroom teaching online and session / lab classes in virtual mode following our daily academic class routine. The syllabi for even semester in 2019 -20 academic session completed satisfactorily. The daily attendance of faculty, progress of syllabi and % attendance of students were monitored daily through online mode. An academic audit was conducted online by the IQAC of our university to oversee the quality of online education and timely completion of courses which was found to be satisfactory in nature. We were confident to come out of the crisis period with complete victory.

During this lockdown period all over India due to COVID-19 pandemic, our group of institutions have taken lot of innovative initiatives to continue academic as well as students/ faculty development activities through digital platform. Some of the initiatives are mentioned hereunder

- Digital Attendance of Faculty- Online morning and afternoon meeting daily with all faculty(which is one of the good practice being followed in all our group of institutions since inception) through youtube live, Goto Webinar- Everyday all faculty meet via Youtube Live, Goto Webinar in the morning and afternoon, comprising of all 500+ members of the group and everyday departmental meetings and faculty assignments reported via Google meet are monitored and considered as a part of compliance for virtual attendance of all faculty and officers.

- Online classes through Google Hangout meet, YouTube live , every day as per routine were held. The digital tools used are - Youtube Live, Google Classroom, IEMCRP (our proprietary educational ERP software), Simulators, like, MATLAB, NS-2, GridSim, Cloud Sim etc., Online Coding platforms like, Hacker Rank, Code Chef, Hacker Earth etc.
- Online Labs by virtual mode using MATLab, Hacker Rank - online coding platform
- Online lecture by corporate experts- With the continuous persuasion and strong initiatives of various departments corporate connect programme are being conducted through online sessions in the evening for our students, thereby not hampering their everyday WFH schedule, and much to our delight, many of them have responded spontaneously and a good number of sessions have been done which are immensely beneficial for students making them aware and ready for the upcoming battle in their career.
- University have conducted online mid-term exam for students as a part of continuous assessment and prepared for end semester exam. These exams were held where students sitting in their home with pen and paper answered the. descriptive questions sent to them digitally . This was not an open book system of examination. Questions taken from question banks were framed in such a manner so that no two question papers were identical. Answer scripts were scanned by the students with mobile phone and uploaded digitally for evaluation. Teachers took help of zoom and later google meet for proctoring (a group of students under one mentor) so that students cannot take help of others during examination. This hybrid on line system with 4500 students during our 2nd term online exam was introduced and found to be very effective . Almost 95 percent students attended the examination without any problem. Students will later submit the hard copies of answer script for keeping record.
- With this hybrid on line system of examinations tested and proven to be very effective , our university has recently completed end semester examinations using the same model for final year students successfully as per the approval of our statutory bodies and in accordance with UGC/ AICTE guidelines . Almost 100 percent students have appeared for the examination. Rest of the students will get another opportunity to appear for examination after the situation normalizes.
- University has facilitated development of faculty members during this testing time by organizing online FDPs, which are taken by eminent academicians of India.
- Online Admission- New batch admission process has been made completely online by making entrance test, IEMJEE online, followed by online counselling-cum-interview session.
- Online Placement- Many recruiters have resorted to online recruitment process and our current final year students have coped up with those opportunities well and grabbed the job offers. ❑

CAMPUS NEWS

National Webinar on Education in Post-pandemic Era

A ten-day National Webinar on 'Education in Post-pandemic Era : Challenges and Opportunities' was organized by the School of Education, Ravenshaw University, Cuttack during June 4-13, 2020. A total of 2452 participants registered from almost all the states and Union territories of the country and from more than 100 universities. The nature of participants was PG/ UG/B.Ed. students, research scholars, school teachers, NGO officials, Government officials, faculty members, *etc.*

In the inaugural session, Dr. Sudarshan Mishra, Head, Department of Education and Organizing Secretary of the Webinar read the theme of the seminar and introduced the guest. Prof. Ishan K Patro, Vice Chancellor, Ravenshaw University, in his inaugural address deliberated upon the problems and opportunities during post-pandemic context. He encouraged participants to rely on self-study which will help in developing higher order thinking skills. He also encouraged the faculty members to motivate the students for self-study. He said that during this pandemic context, the role of the teacher has been changed from mere instructor to a facilitator who will play supportive role for motivating students for self-study. He also emphasized upon continuous assessment of learners which will also help the learners for self-study as well as self-assessment. He cited some practical examples which the universities will going to face during post-pandemic time such as, sanitization of hostel, university campus, departments, maintaining social distancing norms, resuming teaching and examination, *etc.*

Every day one resource person is invited to deliberate upon a topic related to the theme of the webinar. Dr. Sudarshan Mishra delivered a talk on 'Open Educational Resources, Its Need, History and Development, Advantages and Challenges'. Dr. Manas Ranjan Panigrahi, Sr. Programme Officer, Commonwealth Educational Media Centre for Asia (CEMCA), New Delhi delivered a talk on 'SWAYAM MOOCs: An Opportunities for Self-Learning' who focused on teaching-learning through SWAYAM, development of course proposal for SWAYAM, *etc.* Prof. G C Nanda, Senior ICSSR Fellow, Department of Education, Ravenshaw University delivered a talk on 'Autonomy and Accountability in Higher Education' who emphasized upon levels of autonomy, how to promote autonomy and ensure accountability. Prof. Dhaneswar Harichandan, Former Professor cum Director, Institute of Distance and Open Learning, University of Mumbai, Mumbai delivered a talk on 'ICT and MOOCs through SWAYAM'. He gave a conceptual orientation about ICT and step by step procedure for development of MOOCs on SWAYAM. Dr. Amita Puri, Director, Optimus Centre for Wellbeing, Citizen Hospital and De addiction Centre, Gurgaon delivered a talk on 'Developing Resilience and Positive Mental Health during COVID Times'. She discussed about stress and practicing skills in developing unique styles of reacting to stress. Prof. Bharti Dogra, School of Education, Indira Gandhi National Open University, New Delhi delivered a talk on 'Blended Learning and UGC Guidelines, Blending Models, Credit Equivalency, Online Activities in Blended Learning', and how to adopt blended learning. Prof. C P Nanda, Department of History, Ravenshaw University delivered a talk on 'State, People, and Technology: A Regime of New Normality in Post-COVID World Order'. He said that the crisis that neoliberalism faces as a result of COVID-19 is much more deep rooted. He said it is time to reset the certitudes of neoliberalism, including that being pedaled in education. While technology is an important component of life today, the larger issues of life and life world must be through by educationist today. This requires planetary rather than an anthropogenic perspective, a reset of view from body and matter to perhaps soul, vitality and joy. Dr. T K Pany, Associate Professor, Department of Commerce, Ravenshaw University delivered a talk on 'Challenges of Services Sectors with Reference to Higher Education'. Prof. Sanjay Satapathy, Dean, School of Commerce and Business Studies, Ravenshaw University delivered a talk on 'Nourishing Personality'. He highlighted on effective time management, study skills and future classroom.

In the valedictory session, Dr. Sudarshan Mishra read the webinar Report. Prof. Nityananda Pradhan, Principal, Regional Institute of Education, Bhopal delivered the valedictory speech on 'Pedagogy 2.0 in the Era of Education 4.0'. In his talk, he highlighted about history of Revolution in Education, Framework of Pedagogy 2.0, key pedagogical practices, flipped classroom, critical pedagogy, and capacity building of

teachers. In Valedictory Address, Prof. C B Sharma, School of Education, IGNOU, New Delhi said that digital medium can only change the Post-COVID-19 education system. He said that we could not use the technology optimally in education in the first twenty years of twenty-first century. This is the right time to use digital technology in education. In future, teachers will be the most celebrated personality in the media and they will be the highest paid professionals. The world will listen to them. Prof CP Nanda, Department of History, Ravenshaw University, in his presidential address recognized the challenges and opportunities that come with the future education system and hoped for a rapid change in the model of education reform and the distribution of educational processes in the coming days.

Conference on Creativity, Innovation and New Business Models

A two-day Conference on 'Creativity, Innovation and New Business Models for Reviving the Economy' is being organised by TAPMI School of Business, Faculty of Management and Commerce, Manipal University Jaipur in association with Manipal Instituite of Management, Manipal Academy of Higher Education, Manipal and Atal Incubation Centre, Manipal University, Jaipur during August 27-28, 2020 at Jaipur.

Economies all over the world are witnessing prolonged slowdown beset by structural, global, and cyclical factors resulting in of job losses; trade wars and the subsequent reduction in trade flows across the international borders, record-high unemployment rates, liquidity crisis, crashing demand for consumer goods, and pessimistic investor sentiments. A large chunk of businesses is coming under severe stress and strain to sustain or stay afloat in the market. It even requires sustainable business innovations and reforms to withstand the economic storm. Economic revival in developing countries would undoubtedly require pro-growth initiatives, evolving business models and development initiatives to boost the economy. The role of creativity and innovation in recognizing hitherto unexplored opportunities for the introduction of new techniques, new products, and new services needs no emphasis on rejuvenating entrepreneurs are critical components of society which take calculated risks, manage uncertainty and engage in revamping economic activities in explored terrains. Sustainable growth combined with a progressive distribution of income is achievable through inclusive and innovative business models. The themes of the event are:

- Finance and Accounting.
- Economics and Health Care.
- Marketing.
- Human Resource Management.
- Operation Management.

For further details contact, Organising Secretary, TAPMI School of Business, Manipal University Jaipur-303007, Rajasthan, Phone: 0141-3999100 Ext.: 317, 868, 378, E-mail: samita.sharma@jaipur.manipal.edu. For updates, log on to: www. jaipur.manipal.edu

Online Workshop on Basics of C and C++ Programming

A five-day Online Workshop on 'Basics of C and C++ Programming' is being organized by the Department of Mathematics, Rajiv Gandhi University Arunachal Pradesh in collaboration with Spoken Tutorial Project, IIT Bombay during July 15-19, 2020. The students of universities/colleges, research scholars, and academicians may participate in the event.

C is middle level programming language which combines the features of both high level and low level languages. It was developed by Dennis Ritchie during the early 1970s while working at AT&T Bell Labs in USA. The need to redesign the UNIX operating system to enable it to be used on multiple computers led to the development of C. C is a highly portable structured general-purpose programming language which is used for scripting system applications which form a major part of Windows, UNIX and Linux operating system. It also efficiently works on enterprise applications, games, graphics, and applications requiring calculations. C++ is a highly portable language and is often the language of choice for multi-device, multi-platform app development. In fact, C++ is an extended object-oriented programming language which includes classes, inheritance, polymorphism, data abstraction and encapsulation. C++ also allows exception handling, and function overloading which are not possible in C. Being a powerful, efficient and fast language, it finds a wide range of applications - from GUI applications to 3D graphics for games to real-time mathematical simulations. On account of the wide significance of C and C++ programming languages, it has become an integral part of course syllabus in most of the higher educational institutions around the world. Now a day, it helps a great deal in research oriented works in the field of mathematics, Physics, Computer Science, *etc.*

For further details, contact Convener, Dr. Nipen Saikia, Assistant Professor, Department of Mathematics,

Rajiv Gandhi University, Arunachal Pradesh-791112, Mobile: +91 9435507484, E-mail: nipen.saikia@rgu.ac.in.

International Web Lecture Series on Qualitative Research Methodology

A Ten-day Online Faculty Development Programme on 'Crafting Qualitative Research: How can We Study the World as it is Lived?' is being organised by the University of Science and Technology, Meghalaya from July 15-24, 2020. The Programme is particularly for faculties, research scholars especially interested to know about qualitative research. Independent researchers and senior master degree students may also participate in the event.

For further details, contact Convenor, Amit Choudhury, Professor and Dean, School of Business Sciences, University of Science and Technology, Meghalaya-793101, Mobile: +918761007133/ + 919854453170, E-mail: sbuswebls@gmail.com. For updates, log on to: www.ustm.ac.in.

AIU NEWS

West Zone Vice Chancellors' Meet–2019-20
on
Internationalization of Higher Education and Global Rankings

The West Zone Vice Chancellors Meet (2019-20) of the Association of Indian Universities (AIU), was hosted by IIS University, Jaipur on December 11th-12th, 2019. Shri Kalraj Mishra, Hon'ble Governor of Rajasthan graced the Inaugural Session as Chief Guest. Prof. Bhushan Patwardhan, Vice Chairman, UGC was the Guest of Honour and Prof. M M Salunkhe, President AIU, presided over the Meet. Dr. Pankaj Mittal, Secretary General, AIU Convened the Meet. Prof. Ashok Gupta, Vice Chancellor of the IIS University was the Co-convener of the Meet. Dr Alok Mishra, Joint Secretary, AIU was the Nodal Officer of the Meet and Mr Vijendra Kumar took care of logistics and coordination. Dr S Rama Devi Pani, Editor University News was the Nodal Officer for Academics. Dr. Raakhi Gupta, Registrar IIS University was the Nodal Officer of the Meet at host University. The Meet was attended by 80 Vice Chancellors from Gujarat, Maharashtra and Rajasthan apart from international experts from World Bank, Shastri Indo-Canadian Institute, Times Higher Education, British Council, South Asian University & IIM Ahmadabad. Main theme of the Meet was *'Internationalization Of Higher Education and Global Rankings'***.** Under the main theme, following three technical sessions were conducted in the Meet: ***'Promoting Indian Higher Education Abroad and Building International Collaborations'***; *'Improving International Rankings of Indian Universities through Internationalization'; 'Attracting International Students and Faculty to India--Issues and Challenges'*. In addition to Inaugural Session, Valedictory Session and three Technical Sessions, the Meet had AIU Business Session and newly induced Session on Interaction with Apex Bodies. The University had arranged Cultural Evening.

The Chief Guest; Hon'ble Governor of Rajasthan, Shri Kalraj Mishra initiated his in his inaugural address reminded the audience of their constitutional duties and appealed for taking pledge to abide by them. In his peech, he highlighted the role of academia in sensitizing students towards constitutional duties. Teachers should inculcate in students strong value system and must ensure that students not only understand their fundamental duties but also perform them religiously. He emphasised that concerted efforts are required by the Indian universities to be competitive in this era of globalization. Higher education should be more employment oriented and steps should be taken to prevent brain drain from India, he said. Chief Guest released the Special Issue of the University News edited by Dr S Rama Devi Pani was released to commemorate the occasion.

Prof. Bhushan Patwardhan stressed on the pressing need for internationalization of Indian HEIs to survive in this globalized world. He deliberated on various initiatives taken by UGC and AIU to give impetus to internationalization of higher education. He recommended an indispensable shift in Indian Higher Education System; to be student-centric rather than teacher-centric. Prof. M M Salunkhe, in his presidential address focused upon the need for internationalization of education. He envisioned the transformation of Indian higher education through

"Study in India" programme of MHRD.Dr. Pankaj Mittal, Secretary General, AIU, New Delhi outlined AIU's accomplishments in providing value added services to the member universities, apart from bringing visibility to AIU by creating Twitter handle, digitization of the AIU Library, creating portals for admissions, jobs for the students, and also facilitating collaboration with the overseas universities to encourage global partnerships. Dr. Ashok Gupta expressed confidence that the deliberations during the meet will go a long way in redefining operational strategies in Indian higher education within the national boundaries and beyond.

Session on Interface with Apex bodies

The Interface with officers from Apex bodies was chaired by Dr. Bhushan Patwardhan, Vice Chairman, UGC and Dr. Pankaj Mittal, Secretary General, AIU was the Co-Chair. Prof. Amiya Kumar Rath, Adviser represented NAAC. Discussing about the roles AIU, UGC and other apex bodies of education, Dr. Patwardhan suggested that consistent communication between AIU and other apex bodies is very significant in maintain the relevance and standard of higher education. He also shared information about the proposed National Academic Credit Bank.

Prof. Rath during the discourse, shared success stories of NAAC focusing on two major aspects i.e. changes during more than two decades of NAAC's existence and Revised Accreditation Framework. He deliberated on the IIQA application process, SSR submission readiness guidelines, distribution of metrics, key indicators across each criterion, reforms such as data based qualitative indicator evaluation; data validation by external professional agencies, instituting key indicators on alumni engagement and student satisfaction survey. Dr. Mittal threw light on various schemes of UGC like ARPIT, IMPRESS, MOOC's and emphasized that community engagement programmes should be undertaken by the universities to fulfill their social responsibilities. The important take aways from the session are:

1. The universities should undertake curriculum revision based on learning outcomes as revised under the UGC guidelines. Focusing on relevant, updated and learning outcome-based curriculum will prepare the HEIs as well as students for futuristic roles and competing in an international system.
2. Expert committee should be appointed by the universities to assess the quality of research papers published by the faculty. Research profile of a faculty members should not be adjudged on the basis of number of research papers rather it should be based on the quality of research work.
3. NBA accredits only few selected programs; NAAC accredits all the programs but there is dichotomy in the mandates of accreditation agencies which creates confusion. There should be a single accrediting agency.
4. NAAC manuals should provide examples with reference benchmarks.
5. UGC search webpage should be more informative with detailed information about each scheme.
6. IIMs & IITs should handhold the general universities, including self-financing universities which are not having good grades in NAAC to enable them to learn and adopt their best practices. AIU may facilitate interactions and signing of MoUs between the aforesaid institutions.
7. Many rankings such as ICAR, NIRF provide well laid down and transparent benchmarking. Accordingly, NAAC benchmarks be made available to help in self-assessment of university, evaluate their SWOT and make directional efforts for improvement in their grading.
8. More international accreditations should be introduced in India.
9. The fees of some self financing universities is too high, causing financial hardship to parents. To ease this out, institutions may provide avenues to 'Earn-while-learn'. 'Earn-while-learn' should be introduced in all universities. The data of students' earning during degree programs should be one of the criteria for NAAC assessment. It should be essential for getting A+ grading.

Session on Promoting Indian Higher Education Abroad and Building International Collaborations

The first Technical Session; on the theme 'Promoting Indian Higher Education Abroad and Building International Collaborations' was chaired by Prof. A. M. Pathan, former President, AIU and former Vice Chancellor, Maulana Azad

National Urdu University and Central University of Karnataka. The speakers of the session were Prof. Francisco Marmolejo, World Bank's Lead Tertiary Education Specialist, Prof. Manjula Rao, Director, Higher Education and Society, British Council and Dr. Ajit Motwani, Head, Development Office, IIM Ahmedabad, former Director EdCIL, and former India Director, Institute of International Education, IIIE, New York. Prof. Pathan introduced the topic, defining the concept of Internationalization of higher education and outlined the place that India holds in pan-national education.

Prof. Francisco emphasized that India's education system is extremely dynamic and increasingly getting complex. He discussed key trends prevalent in global higher education i.e. uneven expansion, relevance and internationalization of education. Speaking on the need for comprehensive Internationalization, he also conversed on the possible solutions, the most important of which was internationalization of the curriculum. He opined that India's position in terms of capacity, innovation and large contributions might help in making room for international collaborations. His presentation concluded by expressing need for universities to be future ready in terms of constantly changing focus and paradigms in accordance with the challenges being faced by entire tertiary education system in the years to come.

Prof. Manjula Rao discussed the practicalities of internationalization of higher education in India. She stressed on the need for universities to explore each other, forge partnerships and sustain these corporations for mutual benefits. Dr. Rao also emphasized the need for understanding the perception of international audience about the Indian education, brand building and accordingly universities being proactive. Dr. Ajit Motwani opined that collaborations cannot be over emphasized. He stated that the roadmap for internationalization must be different for each university / Institution. The recommendations of the session are:

1. Regional Committees of universities located in an area should be formed to facilitate collaborative regional studies, universities' strengths, contribution, USP etc. and should explore; how their internationalization can be promoted with respect to mobility, course content, cultural exchange, projects and research.
2. Universities must create international facilities in the campuses to ensure comfortable and harmonious stay of Indian and foreign students together.
3. Universities should essentially invite experts from foreign universities for workshops, seminars, examinations and various other activities to promote Indian education.
4. Government should ensure proper distribution of foreign students among universities.
5. Government should invite and arrange some free trips to foreign students and faculty to India to promote Indian higher education.
6. Dual degree at postgraduation level in collaboration with foreign universities other than parent university should be facilitated. AIU should ensure equivalence of dual degrees.
7. Regulatory authorities should act as facilitators instead of controllers or inspectors.
8. Internationalization requires a lot of support from the Government and embassies for visa & FRRO with police department. Therefore, Government should make the process easy, fast and smooth.
9. AIU should popularize the initiatives of the Government of India towards promoting Indian higher education abroad and international collaboration through workshops, seminars and University News.
10. Indian Government should create permanent country-wise platform involving Indian diaspora to promote Indian higher education abroad.
11. Government should launch promotional campaigns like 'Incredible India' to promote best Indian universities/institutes of higher education abroad.
12. Barriers/challenges need to be identified that hinder the inward mobility of foreign students to Indian universities.
13. The skills and perspectives of international students on campus need to be understood and shared with other students.
14. Alumni to be leveraged significantly so that Institute builds its brand in other countries; alumni being the spokesperson.

Session on 'Improving International Rankings of Indian Universities through Internationalization'

The Technical Session on 'Improving International Rankings of Indian Universities

through Internationalization' was chaired by Prof. Sandeep Sancheti, former President of AIU and Vice Chancellor, SRM Institute of Science and Technology, Tamil Nadu. The panelists of the session were Mr. Ritin Malhotra, Regional Director (South Asia) Times Higher Education and Prof. Kamlesh Mishra, Vice Chancellor, Rishihood University, Bahalgarh. In his opening remarks, Prof. Sancheti enumerated various advantages and disadvantages of rankings.

Mr. Ritin Malhotra discussed the framework of Times Higher Education (THE) ranking and parameters for internationalization of colleges. He stated that the alumni are the goodwill ambassadors of academic excellence of an institution and the reputation of an institution is built by its alumni. Mr. Kamlesh Mishra shared his experience in evolution of ranking systems and the challenges faced by the system. Following recommendations emerged from the session:

1. Industry & Education sector are lacking in research activities. There is need to encourage the research initiatives in both the sectors which complement each other. Emphasis should be laid on research collaboration, academic reputation and citation impact
2. Yoga courses should be introduced in universities as a science with both theoretical & practical component and not just as a casual practice for international acceptance.
3. While trying to compete with other players in the world, we need to have an established system with defined guidelines, vision and mission. The standards, goals and priorities need to be set and be followed in principle accordingly.
4. Ranking and ratings do not necessarily indicate quality in real terms. More emphasis should be laid on quality of the professionals that are being produced in the universities. Balance should be done by universities on the priorities and strengths not only for ranking but to elevate the overall standards of institutions.
5. Universities should look at the changing education dynamics in international scenario and align their syllabus in line with emerging challenges with a definite industry interface.
6. In order to raise the standards of the universities, there is an urgent need to groom and mentor the skills of faculty members for their overall development through faculty exchange programmes and tie-ups with national and international universities. Student exchange programs within the country and with different nations also need to be organized.
7. There is a need to hire thinkers and innovators with enhanced skills at par with global standards. National programmes like 'Make in India' need to expand their horizons further, to engage indigenous professionals, provide employment and curb brain drain. Close collaboration with international media is also the need of the hour.
8. The mapping of expertise and expectations is very important. Students need to be equipped with proper resources of knowledge and skill. Universities should be proactive in communication, specifically involving research activities and research outputs.
9. The prime objective of the universities should be addressing local problems and align themselves with the development of surrounding areas and resident communities. The reputation of a university is largely impacted by their community outreach programme, wellbeing of neighboring area and its inhabitants. Internationally, universities should strengthen their public relations and media presence.
10. There should be well defined criteria on attracting international students to India. Serious and scaled efforts need to be made to help Indian Higher Education brand building and accelerate inward students' mobility.
11. HEIs must concentrate on quality research and low impact factor in research needs to be improved. HEIs need to focus on entrepreneurship and being job creators instead of job seekers.

Session on Attracting International Students and Faculty to India: Issues and Challenges.

The third technical session on 'Attracting International Students and Faculty to India: Issues and Challenges' was chaired by Prof. A V S Ramesh Chandra, President, South Asian University, New Delhi. The speakers of the session were Dr R.L. Raina, VC, JKLU, Jaipur and Dr. Prachi Kaul, Director, Shastri Indo- Canadian Institute, New Delhi.

The session began with insightful observations from Prof. Chandra who remarked at the changing perceptions about Indians in foreign countries. He stated that Indians professionals represent values such as hard work, trust, teamwork and loyalty. He opined that comfortable and affordable campus life and universal acceptability of grades are some of the important factors that play a crucial role in attracting foreign students. He remarked that Indian institutions are striving hard to improve the quality of education. Prof. Raina reckoned multiple factors; playing vital role in improving the quality of education; for it to compete internationally. He opined that enriching classroom experience with focus on learning, pedagogy, curriculum, quality of faculty, evaluation process, and innovative courses can significantly improve the quality of Indian Higher Education.

Dr Prachi Kaul, Director, Shastri Indo Canadian Institute forwarded suggestions that can help the universities in attracting foreign students and faculties. She stressed on the need to outreach at a global level by marketing abroad. She reiterated that flexibility in grading, curricula and training can attract foreign students. She also recommended that HEIs need to work on employability factor and bring in required change in academic processes and systems. Major recommendations of the session are:

1. Internationalization of higher education can be augmented by creating Non-Resident Indians (NRI) and foreign students' quota on supernumerary basis in HEIs.
2. Academic credit exchange should be systematized and seamlessly aligned so that uniform standards of credit transfer can be developed among countries. This will facilitate International students' inward and outward mobility.
3. Research domains should be prioritized for international collaboration. The priority areas must be identified keeping in view the international trends of academic research.
4. Internationalization of higher education mandates developing world class infrastructure, amenable academic environment enabling exchange of students, teachers, academic programmes, high quality research projects and conducive research ambience.
5. Encouraging regulations are essential to facilitate cross border movement of academic stakeholders in higher education. The draft education policy must clearly spell out regulations covering all aspects of Internationalization.
6. There is an immediate need to inculcate cultural sensitivity among academia to seamlessly amalgamate international stakeholders. Adequate safety and security provisions should be ensured on the campus. Home Stay, involving international students in various cultural activities and familiarizing them with indigenous culture are some of the desirable steps to encourage inward mobility. The counseling bureaus of the universities must organising workshops and orientation programmes time to time for cultural orientation of foreign students.
7. Institutional arrangements are mandatory to address various concerns of out-bound and in-bound students. The policy must include standardized, tangible modalities; compulsory for all universities admitting international students. An effective and speedy grievance redressal mechanism must be in place to address the concerns of international students.
8. Internationalization of higher education can be achieved by striking a balance between indigenous and global knowledge systems. New academic programs must have international orientation without losing sight of the indigenous values. Existing programmes may be redesigned to accommodate both indigenous and international components.
9. Policies should be formulated to support institutional collaborations for comprehensive advancement of higher education as a universal public good.
10. Indianisation of higher education should be first step toward internationalization. Indianisation can only be feasible by removing regional barriers and state reservations in the country.
11. At present around, five lac students go out of India to study in foreign HEIs. As compared to this; only 54000 foreign students are admitted in Indian HEIs. The inflow of foreign students needs to be maximized by building world class infrastructure, simplifying visa rules, easy entry and exit provisions, good hostel

facilities, curricular revision and international orientation, etc.

12. There is a need to attract foreign faculty and incentivizing them not only financially but also stimulating their intellect by interacting with cross-cultural, bright students and colleagues.
13. It is high time for Indian higher education institutions to achieve desired credibility and visibility among national and global circles. Collective efforts need to be made by the Government, apex regulatory bodies and universities.
14. Enriching experiences by students and faculty inside as well as outside the classrooms are critical factors to attract foreign faculty and students to HEIs.
15. Apart from the educational infrastructure and faculty upgradation programs, India should consider sensitizing Indian diplomatic mission abroad to make concerted efforts to attract foreign students; looking for educational opportunity in India.
16. MHRD, apex regulatory bodies and other ministries should work in tandem for ease of policies, regulations, providing required support, infrastructure as well as human resource support to foreign students coming to study in Indian universities.
17. AIU should provide platform to the universities to share best practices and experiences; worth emulating from leading global universities. University News may have column of best practices in which every week best practices adopted by one best university can be brought out.
18. Universities should rope in and actively involve Indian embassies in highlighting strengths of Indian HE system.
19. It is evident that the number of students admitted in Indian HEIs and supported by Indian Government scholarships is not impressive. In fact, more than half the students come from Afghanistan. Entire foreign students' policy needs to be revised taking into consideration prevailing geopolitical scenario, availability of higher education system in the respective countries and India's relationship with different countries.
20. Nowadays, education has also become commercial entity. Hence, in order to compete in international market, role of policy makers is as important as that of universities. At university level, infrastructure, faculty, hostel facilities and research structure should be enriched to attract, adapt and sustain foreign students in Indian conditions. Cost of education should also be competitive.
21. Academic hubs should be developed in India to attract foreign students.

Valedictory Session

Dr. Subhash Garg, Minister of Technical Education, Rajasthan graced the valedictory session as the Chief Guest. The session was chaired by Prof. Sandeep Sancheti, former President AIU. Ms. Shuchi Sharma, Secretary Higher Education, Government of Rajasthan, Prof. Vivek Saoji, Vice Chancellor, KLE University, Belgam & Governing Council Member, AIU were the Guests of Honour. Dr Pankaj Mittal, Secretary General, AIU convened the Session and Dr. Ashok Gupta, Vice Chancellor, IIS University felicitated the Guests and Officers of AIU. Dr S Rama Devi Pani presented the Report of the Meet. The Meet ended with vote of thanks by Dr Alok Mishra, Joint Secretary, AIU.

. ❑

THESES OF THE MONTH

SCIENCE & TECHNOLOGY

A List of doctoral theses accepted by Indian Universities (Notifications received in AIU during the month of Feb-March, 2020)

AGRICULTURAL & VETERINARY SCIENCES

Agricultural Economics

1. Kausadikar, Harsha Hemant. **Economics of production and marketing of sweet orange in Marathwada Region of Maharashtra**. (Dr. S R Nagargoje), Department of Agricultural Economic, Vasantrao Naik Marathwada Agricultural University, Parbhani.

2. Srikanth, Bandi. **Econometric analysis of production and marketing of maize in Telangana State**. (Dr. K V Deshmukh), Department of Agricultural Economic, Vasantrao Naik Marathwada Agricultural University, Parbhani.

Biotechnology

1. Malik, Palvi. **Genome wide selection for rapid introgression of productivity traits from *Oryza rufipogon* into O sativa**. Department of Biotechnology, Punjab Agricultural University, Ludhiana.

2. Verma, Annu. **Gene networks involved in cold stress response in chickpea (Cicer arietinum L) anthers**. (Dr. H K Chaudhary), Department of Agricultural Biotechnology, CSK Himachal Pradesh Krishi Vishvavidyalaya, Palampur.

Entomology

1. Bokan, Sanjog Chandrakant. **Seasonal incidence, crop loss assessment and bioefficacy of insecticides against major insect pests of soybean (*Glycine max* (L) Merrill)**. (Dr. P R Zanwar), Department of Agricultural Entomology, Vasantrao Naik Marathwada Agricultural University, Parbhani.

Genetics & Plant Breeding

1. Mohd Shamshad. **Genetic and physiological basis of nitrogen uptake and utilization in wheat (*Triticum aestivum* L)**. Department of Plant Breeding & Genetics, Punjab Agricultural University, Ludhiana.

Horticulture

1. Damodhar, Vijay Pundlik. **Effect of nutrients and plant growth regulators on yield, quality and post-harvest life of banana Cv grand naine**. (Dr. G M Waghmare), Department of Horticulture, Vasantrao Naik Marathwada Agricultural University, Parbhani.

2. Ingole, Revati Hemantrao. **Studies on nutrient management in response to growth, yield, quality, shelf life and value addition in papaya (*Carica papaya* L)**. (Dr. G M Waghmare), Department of Horticulture, Vasantrao Naik Marathwada Agricultural University, Parbhani.

3. Zoremtluangi, Jenny. **Nutrient management in Khasi mandarin (Citrus reticulata blanco) under subtropical agroclimatic condition in Mizoram**. (Dr. Debashis Mandal), Department of Horticulture, Mizoram University, Aizawl.

Veterinary Science

1. Kondampati, Kasturi. **Effect of different anaesthetic protocols on the plasma levels and disposition kinetics of ampicillin-cloxacillin combination in horses**. Department of Veterinary Pharmacology and Toxicology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana.

2. Mir, Insha. **Aflatoxin contamination in cattle feed value chain and its impact on milk quality in Punjab**. Department of Veterinary and Animal Husbandry Extension Education, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana.

3. Sharma, Deepti. **Ultrasonographic evaluation of canine liver and kidney with special reference to shear wave elastography in health and disease**. Department of Veterinary Surgery and Radiology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana.

4. Srinu, B. **Epidemiological study on the anthropogenic disseminated environmental pollutants in the aquatic ecosystem and their association with antimicrobial resistance emergence**. Department of Veterinary Public Health & Epidemiology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana.

BIOLOGICAL SCIENCES

Bioinformatics

1. Tripathi, Swetakumari Pauharisharan. **Identification and characterization of potential drug targets for glioblastoma: An *in-silico* approach**. (Dr. Rakesh M Rawal), Department of Bioinformatics, Gujarat University, Ahmedabad.

Biotechnology

1. Gangadasu, Seetharavamma. **Development of monoclonal antibody based ELISA and Lateral Flow Strip (LFS) for detection of cry 1 Ac Bt protein in genetically modified cotton plants**. (Dr. P Sudhakar), Department of Biotechnology, Acharya Nagarjuna University, Nagarjuna Nagar.

Botany

1. Patel, Svetaben Bharatbhai. **Genetic diversity assessment of *Acmella paniculata* (Wall Ex Dc) R K Jansen collected from different regions**. (Dr. H A Solanki), Department of Botany, Gujarat University, Ahmedabad.

Life Science

1. Raval, Nidhi Pankaj. **Identification and validation of Y-sperm specific markers for semen sexing in dairy animals**. (Dr. Linz-Buoy George), Department of Life Science, Gujarat University, Ahmedabad.

2. Sharma, Bhagwati Sharan. **Comparison of Locus Control Regions (LCRs) and designing of a system to understand regulatory potential in complex gene expression**. (Dr. R J Verma), Department of Human Genetics, Gujarat University, Ahmedabad.

Zoology

1. Umrigar, Namrata Rajesh. **Biodiversity and ethological study of water birds in and around Surat City**. (Dr. Arunkumar H Dholakia), Department of Zoology, Veer Narmad South Gujarat University, Surat.

EARTH SYSTEM SCIENCES

Environmental Science

1. Parmar, Payal Vinod. **Biodiversity analysis and habitat suitability studies of certain important wetlands of Gujarat using aquatic macroinvertebrates as the niche indicators**. (Dr. Rita N Kumar), Sophisticated Instrument Center for Applied Research and Testing, Sardar Patel University, Anand.

2. Shakti. **Preparation characteriazation and pharmacological evaluation of polyphenolic nanoformulations from camellia sinensis**. (Dr. Neeraj Dilbaghi and Dr. Dinesh Dhingra), Department of Bio and Nano Technology, Guru Jambheshwar University of Science & Technology, Hisar.

ENGINEERING SCIENCES

Civil Engineering

1. Amitha, N R. **Some studies on strength behaviour of concrete due to improper casting sequence**. (Dr C Sashidhar), Department of Civil Engineering, Jawaharlal Nehru Technological University Anantapur, Ananthapuramu.

Computer Science & Engineering

1. Bhavsar, Shreyaben Bhupendrabhai. **Multilingual multimedia based crops disease management system using mobile technology**. (Dr. Dipti B Shah), P G Department of Computer Science & Technology, Sardar Patel University, Anand.

2. Naralasetti, Neelima. **Cost and profit analysis of web application deployment on cloud**. (Dr. Bobba Basaveswara Rao), Department of Computer Science & Engineering, Acharya Nagarjuna University, Nagarjuna Nagar.

3. Preeti. **Metaheuristic techniques for feature selection in face recognition**. (Prof. Dinesh Kumar), Department of Computer Science & Engineering, Guru Jambheshwar University of Science & Technology, Hisar.

Electrical & Electronics Engineering

1. Dhamecha, Vishal Dineshkumar. **Studies on compound semiconductor ($ZnGa_2Se_4$) thin film electronic devices**. (Dr. Basumati H Patel), P G Department of Electronics, Sardar Patel University, Anand.

Electronics & Communication Engineering

1. Kanna, T Venkatesh. **Analysis of wavelet extensions for epileptic seizure detection using support vector machine**. Department of Electronics & Communication Engineering, Hindustan Institute of Technology and Science, Chennai.

Mechanical Engineering

1. Mishra, Santosh Kumar. **Numerical studies on non-Newtonian fluid flow through pipe using CFD**. (Dr. Arun Arora and Dr. H S Chandra), Department of Mechanical Engineering, Chhattisgarh Swami Vivekanand Technical University, Bhilai.

MATHEMATICAL SCIENCES

Mathematics

1. Rabish Kumar. **Spectral theory and Eigen functions expansions for a pair of second order matrix differential equations**. (Dr. Amar Kumar), Department of Mathematics, T M Bhagalpur University, Bhagalpur.

2. Sunita Rani. **Deformation of a stratified poroelastic media due to surface loads and buried sources**. (Prof. Sunita Rani), Department of Mathematics, Guru Jambheshwar University of Science & Technology, Hisar.

MEDICAL SCIENCES

Biotechnology

1. Salappa, Arun Ramling. **Solubility enhancement of poorly water soluble simvastatin and fenofibrate**. (Dr. E Jayachandran), Department of Biotechnology, Acharya Nagarjuna University, Nagarjuna Nagar.

2. Zodinpuii, Doris. **Characterization of clinically significant mutations associated with breast cancer in Mizo population**. (Prof. N Senthil Kumar), Department of Biotechnology, Mizoram University, Aizawl.

Nursing

1. Kori, Asha Deepak. **Community Initiatives of corporate social responsibility in engineering organizations in Mumbai: A study of women beneficiaries**. (Prof. Vandana Chakrabarti), Department of Nursing, S.N.D.T. Women's University, Mumbai.

2. Waghe, Aakanksha Ashok. **Effect of planned teaching on KAP of mothers and nutritional status of infant/children**. (Dr. Shakuntala Prabu), Faculty of Science and Technology, S.N.D.T. Women's University, Mumbai.

Pathology

1. Panicker, Soumya Mohanan. **A study on liquid-based cytology-new possibilities in the diagnosis of cervical lesions in India**. (Dr. Menka H Shah), Department of Pathology, Sardar Patel University, Anand.

Pharmaceutical Science

1. Hadia, Jayeshkumar Jasubhai. **Development and evaluation of colloidal drug delivery system containing**

poorly water soluble drug. (Dr. Ashvin V Dudhrejiya), Department of Pharmacy, Saurashtra University, Rajkot.

2. Pandey, Ajit Kumar. **Synthesis and biological evaluation of some thiadiazole derivatives and validation by UV spectrophotometric method**. (Dr. Pranita Kashyap and Dr. Chanchal Deep Kaur), Faculty of Pharmacy, Biomedical Engineering and Biotechnology, Chhattisgarh Swami Vivekanand Technical University, Bhilai.

3. Samir Odhavajibhai, Rabadiya. **Effect of selective histone deacetylase inhibitors on cardiac and renal complications associated with typeI I diabetes mellitus in rats**. (Dr. Ashvin V Dudhrejiya), Department of Pharmacy, Saurashtra University, Rajkot.

4. Sawarkar, Hemant Arunrao. **Phytochemical and pharmacological investigation of traditionally used herbs for treatment of Buccal ulcer**. (Dr. Pranita Kashyap and Dr. Chanchal Deep Kaur), Faculty of Pharmacy, Biomedical Engineering and Biotechnology, Chhattisgarh Swami Vivekanand Technical University, Bhilai.

5. Sheokand, Sneha. **Development of nanocrystalline solid dispersion of hesperetin and its oral formulation, with enhanced oral bioavailability for prevention of breast cancer**. (Dr. Arvind K Bansal), Department of Pharmaceutics, National Institute of Pharmaceutical Education and Research, Mohali.

Physiology

1. Vaidya, Lopa Madhusudan. **A comparative study of audiogram in ear sound devices users and non-users in age group of 18-40 years**. (Dr. Nirupama J Shah), Department of Physiology, Gujarat University, Ahmedabad.

PHYSICAL SCIENCES

Chemistry

1. Jasminder Singh. **Study on CO_2 capture using nanostructure carbon adsorbents**. (Dr. Soumen Basu and Dr. Haripada Bhunia), School of Chemistry and Bio-Chemistry, Thapar Institute of Engineering and Technology, Patiala.

2. Kamani, Ronakkumar Dhanjibhai. **Isatin and pyrazole clubbed heterocycles: Synthesis, characterization and biological studies**. (Dr. K H Patel and Dr. D K Raval), P G Department of Chemistry, Sardar Patel University, Anand.

3. Kanthecha, Darshana Natvarlal. **Synthesis, characterization and biological investigations of some novel gold (III) and ruthenium (III) complexes**. (Dr. M N Patel), P G Department of Chemistry, Sardar Patel University, Anand.

4. Prakash, Ajay. **Transition metal complexes of schiff bases and their biological studies**. (Prof. Rajesh Malhotra), Department of Chemistry, Guru Jambheshwar University of Science & Technology, Hisar.

5. Savaliya, Ravikumar Bhagvanjibhai. **Synthesis, characterization and application of graft copolymers and its blends with thermoplastics polymers**. (Dr. M M Raj), P G Department of Chemistry, Sardar Patel University, Anand.

Physics

1. John, Jinya. **Study of cloud microphysical properties over the Indian Region using satellite data**. (Dr. V Sathiyamoorthy), Department of Physics, Gujarat University, Ahmedabad.

2. Patel, Hiral Timin. **A comprehensive study of physical properties of some simple and non-simple metals by using an effective and local pseudopotentials**. (Dr. P R Vyas), Department of Physics, Gujarat University, Ahmedabad.

3. Singh, Rachana. **AB-initio modeling and performance study of light emitting electro chemical cells**. (Dr. Mohan L Verma), Department of Applied Physics, Chhattisgarh Swami Vivekanand Technical University, Bhilai.

4. Tiwary, Prem Pyari. **Laser and plasma interaction at high power laser flux**. (Prof. Vibha Rani and Prof. R P Sharma), Department of Physics and Computer Science, Dayalbagh Educational Institute, Agra. ❑

Bishop Chulaparambil Memorial College (BCM College)

Kottayam, Kerala -686001 | Tel. 0481-2562171

www.bcmcollege.ac.in

Affiliated to Mahatma Gandhi University, Kottayam

Applications are invited from qualified candidates for appointments to the permanent posts of **Assistant Professor** in the following disciplines, subject to the approval from Mahatma Gandhi University, Kottayam/Govt. of Kerala.

Subject	Category & No. of Posts
Zoology	Community - 1
Botany	Community - 1
	Open Merit - 1

Qualification, Age and Approval etc as per Kerala Government/UGC/University norms. Apply to the Manager within one month from the date of publication of this advertisement in the prescribed form which can be had from the college office on payment of Rs. 1000/- (Rs. 1100/- by post).

Sd/-, Manager

ASSOCIATION OF INDIAN UNIVERSITIES

ADVERTISEMENT TARIFF

UNIVERSITY NEWS JOURNAL

GST AT PRESENT RATE OF 5% IS PAYABLE FOR PUBLICATION OF ALL TYPES OF ADVERTISEMENTS IN UNIVERSITY NEWS W.E.F. APRIL 01, 2020 IN ADDITION TO THE PAYABLE CHARGE FOR EACH CATEGORY AS GIVEN BELOW)

A. FOR EDUCATIONAL INSTITUTIONS, GOVT. ORGANIZATIONS, PUBLISHERS, BOOK SELLERS & DISTRIBUTORS DISPLAY

(Amount in Rupees)

Categories of Advertisement	1 Insertion	4 Insertions	8 Insertions	12 Insertions
Full Page	15000	45000	85000	120000
Half Page	8000	28000	50000	68000
Quarter Page	5000	16000	28000	40000
Cover (Inside)	16000	55000	100000	144000
Cover (Back)	20000	65000	120000	165000

B. TARIFF FOR SPECIAL NATURE OF MATTERS/ITEMS (DOUBLE THE RATES)

TARIFF FOR SUPPLIERS OF COMPUTERS, COMPUTER STATIONERY & PERIPHERALS, SCIENTIFIC & SURGICAL INSTRUMENTS, SPORTS GOODS AND OTHERS (*NOT COVERED IN ANY FORM OF THE TARIFF)* WILL BE AT DOUBLE THE RATES AND TARIFF CAN BE HAD ON REQUEST.

C. CONCESSIONAL TARIFF (For Publishers/Book Distributors- Exclusively for Books)

Per Square Cm (Display)	1 Insertion	4 Insertions	8 Insertions	12 Insertions
	30.00	28.00	26.00	24.00

MECHANICAL DATA OF JOURNAL

Size of Page 21 cms x 27 cms

PRINT AREA

Full Page 23 cms (Height) x 16.5 cms (Width)
Half Page 12 cms (Height) x 16.5 cms (Width)
Quarter Page 11 cms (Height) x 8 cms (Width)

The Art Work/CRC IN PDF as per above Print Area (in BLACK & WHITE ONLY) or the OPEN FILE IN MS WORD may be sent accordingly. The text is required to be sent positively on E-Mail IDs as shown below. *MATTER FOR ADVERTISEMENT MUST REACH SEVEN (07) DAYS IN ADVANCE FROM THE DATE PUBLICATION OF EACH ISSUE OF UNIVERSITY NEWS*, WHICH IS EVERY MONDAY.

ADVERTISEMENT AGENCIES (INS ACCREDITED) ARE ALLOWED 15% DISCOUNT.

Full advance payment could be sent by Demand Draft/At Par Cheque favouring "ASSOCIATION OF INDIAN UNIVERSITIES", payable at New Delhi. Also, the details for payment via NEFT/RTGS are available in AIU Website. Also, the required data could be provided on request.

***For further information write to* :-**

Publication & Sales Division
Association of Indian Universities
AIU House, 16, Comrade Indrajit Gupta Marg, New Delhi - 110 002
EPABX : 011-23230059 (Extn. 208/213), FAX : 011-23232131
E-mail IDs : advtun@aiu.ac.in /publicationsales@aiu.ac.in
Website : http://www.aiu.ac.in

Towards Social & Engaging Peer Learning: Predicting Backchanneling and Disengagement in Children

Mononito Goswami*, Minkush Manuja*, Maitree Leekha*

arXiv:2007.11346v1 [cs.HC] 22 Jul 2020

***Abstract*—Social robots and interactive computer applications have the potential to foster early language development in young children by acting as peer learning companions. However, studies have found that children only trust robots which behave in a natural and interpersonal manner. To help robots come across as engaging and attentive peer learning companions, we develop models to predict whether the listener will lose attention (Listener Disengagement Prediction, LDP) and the extent to which a robot should generate backchanneling responses (Backchanneling Extent Prediction, BEP) in the next few seconds. We pose LDP and BEP as time series classification problems and conduct several experiments to assess the impact of different time series characteristics and feature sets on the predictive performance of our model. Using statistics & machine learning, we also examine which socio-demographic factors influence the amount of time children spend backchanneling and listening to their peers. To lend interpretability to our models, we also analyzed critical features responsible for their predictive performance. Our experiments revealed the utility of multimodal features such as pupil dilation, blink rate, head movements, facial action units which have never been used before. We also found that the dynamics of time series features are rich predictors of listener disengagement and backchanneling.**

***Index Terms*—Educational Data Mining, Time Series Classification, Backchanneling, Engagement, Permutation Feature Importance, Interpretability, Machine Learning, Deep Learning**

I. Introduction

MANY intervention studies have shown that engaging in narratives can help children prepare to succeed in formal education in school [Curenton, 2010]. Conversational activities like storytelling play a pivotal role in children's early language development. Storytelling, for instance, nurtures children's narrative skills by forcing them to combine multiple ideas in the form of sentences into a coherent whole. Moreover, early conversational exchanges that children have with their parents and peers significantly influences their narrative skills. Peterson and Jesso [Peterson et al., 1999] revealed that children whose mothers encouraged longer narratives through backchannel responses and open-ended questions, had a better vocabulary and overall narrative skills in the long term [Peterson et al., 1999]. Narrative conversations with peers can also cultivate children's communication skills much in the same way. Furthermore, peer learning (and tutoring) is not only beneficial in theory [Topping and Ehly, 1998], and many studies have reiterated the positive outcomes of peer learning for children belonging to low-income households and urban areas [Tymms et al., 2011].

Given the benefits of peer learning, can we design peer learning companions to support early language learning in children? Studies in the past have shown that interactive computer applications and social robots indeed have the potential to educate young children [McReynolds et al., 2020, Mostow et al., 2003, Aist and Mostow, 1997, Aist et al., 2001, Movellan et al., 2009]. Children not only trust social robots and believe that they have emotions [Kahn Jr et al., 2012], but also readily learn new information from them [Movellan et al., 2009]. However, recent studies found that children not only trust '*social*' robots more [Breazeal et al., 2016], attend to them and find them more attentive, but are also distracted by non-socially contingent robots [Park et al., 2017a].

For robots or tablet applications to be socially-contingent, they must interact in a 'natural' and interpersonal manner. Having a natural conversation is a product of a successful, continuous and back-and-forth process of speaker cues and listener backchannels. Backchanneling responses are subtle verbal or non-verbal signals that indicate the speaker that the listener is following the conversation [Yngve, 1970]. In this paper, we focus on making robots more natural and engaging. While our long term goal is to create social and engaging peer learning companions to foster *children's* early language development, our methods are reasonably generic and can be readily adapted for adult-adult interactions.

In this work, we introduce two models to help robots come across as active listeners and engaging storytellers. An active listener must consistently communicate with the speaker through backchanneling responses [Park et al., 2017a]. On the other hand, a good storyteller must be able to predict when a listener will lose attention and prevent disengagement by making an engaging gesture [Sidner et al., 2005]. To this end, we develop two models: a (*i*) listener disengagement prediction (LDP) to anticipate listener disengagement, and a (*ii*) backchanneling extent prediction (BEP) model to predict the extent to which a robot should generate backchanneling responses in the next few (3) seconds. We also carry out multiple statistical and machine learning experiments to shed light on the influence of socio-demographic and economic factors on the amount of time children spend listening and backchanneling. Finally, in order to explain the decisions made

*All authors contributed equally and wish that they be considered as Joint First Authors.

This work was carried out when all authors were with the Department of Computer Science & Engineering, Delhi Technological University, New Delhi INDIA

by our models so that they are more useful for educational technologists and educators, we extensively examine critical features responsible for their predictive potential.

This paper extends our prior work [Jindal et al., 2020] in several directions. First, we examine the predictive utility of a wide range of time series characterization techniques as inputs to our Random Forest models. Second, we analyze socio-demographic features responsible for differences in the amount of time children spend listening to their partners. Third, we generalize the Permutation Feature Importance algorithm to examine features relevant to our ResNet models. Fourth, we view our rich multimodal feature set in the context of features used by prior work through multiple feature ablation experiments. Fifth, we devise an algorithm to quantify the importance of time series features for our models.

The rest of the paper is organized as follows. The next section reviews existing work on backchanneling, disengagement and learning technology for children. Section III describes the dataset used, feature engineering and the LDP and BEP models. Section IV discuss our experiments and their results. Finally, Section V concludes our work with a discussion of our contributions and findings.

II. Relation to Prior Work

A. Peer tutoring and Robots as peer learning companions

Peer tutoring usually refers to two children working in a tutor-tutee dyad, where each individual has well-defined roles and protocols of interactions [Tymms et al., 2011]. Topping and Ehly [Topping and Ehly, 1998] had theoretically modelled the cognitive advantages of peer tutoring and learning. In the course of peer tutoring, a tutor manages activities such that they remain in the 'Zone of Proximal Development' through adequate support and scaffolding. This is a consequence of co-construction, wherein a peer tutor acts as a co-learner and minimizes their partner's frustration resulting from excess challenge. Since peer tutors do not hold a position of authority, tutees may also develop a trusting relationship with them and voluntarily disclose their ignorance or misconception. Many studies have reported encouraging outcomes of peer tutoring young students and those from low-income households and urban areas [Tymms et al., 2011].

Interactive tablets and social robots have shown great promise in educating young children worldwide and have the potential to become excellent peer tutors. While a majority of learning technologies are developed for children of WEIRD (Western, Educated, Industrialized, Rich and Democratic) countries, a recent large-scale intervention study showed that carefully designed educational software can provide quality basic education to all children regardless of their cultures and socio-economic status [McReynolds et al., 2020]. In the context of early language education, learning agents such as Project LISTEN's Reading tutor have been shown to improve children's reading comprehension significantly [Aist and Mostow, 1997]. An intervention study compared the Reading Tutor with regular classroom activities and found that children using the tutor outgained their classmates in comprehension [Mostow et al., 2003]. Another surprising study [Aist et al., 2001] found that students tutored by certified teachers did no better than children using a shared Reading tutor in tasks like word identification, word & passage comprehension and fluency. Among other factors, learning technologies dramatically improve learning gains by steering children towards their 'Zone of Proximal Development'.

However, for a learning agent to be a successful peer learning companion, it must be able to leverage social cues and our means of communication (facial expressions, gestures, verbal expressions etc.) to interact with us more 'naturally'. A recent study [Park et al., 2017b] had shown that children not only direct their attention towards social robots and find them more attentive, but are also distracted by non-contingent robots. To this end, in this paper, we develop models to enable peer learning companions to be social and engaging.

B. Social and Engaging communication

To communicate effectively, both speakers and listeners must go beyond merely issuing utterances and hearing & understanding, respectively. They must not only actively coordinate on the content of the communication, but also its process. Both the speaker and listener must establish a mutual understanding of the subject of discussion (content coordination). Moreover, the speaker must speak only when the listener is paying attention and understanding what the speaker says. Likewise, the listener must clearly indicate to the speaker that s/he is doing just that (process coordination) [Clark et al., 1991]. Thus, effective communication is a continuous, dynamic and collective process of the first order.

Listeners communicate through subtle backchannel responses such as gaze locking, nodding, gestures and short verbal expressions (*'yeah'*, *'ohh'*, *'hmm'*). These responses serve several cognitive functions such as indicating the state of engagement, understanding, clarification of ideas and sentence completion. For learning agents to be perceived as better listeners, they must backchannel adequately and at appropriate times. Few studies have been able to predict backchanneling opportunities successfully using the speaker's vocal prosodic features such as the pitch and energy [Morency et al., 2010, Noguchi and Den, 1998, Ward and Tsukahara, 2000]. In their seminal work, Ward and Tsukahara [Ward and Tsukahara, 2000] proposed a rule-based model which associated backchanneling responses with a region of low pitch lasting a few milliseconds. A decade later, Morency *et al.* [Morency et al., 2010] demonstrated how sequential probabilistic models such as Hidden Markov Models can automatically learn how to predict backchanneling opportunities from a database of adult-adult interactions. They also demonstrated that using multimodal features derived from the speaker (prosody, spoken words and eye gaze) resulted in a significant improvement over Ward and Tsukahara's unimodal model. This study not only opened up the exciting possibility of scaling up the development of backchanneling opportunity prediction models by automatically analysing human interactions, but also established the predictive utility of multimodal features. Poppe *et al.* [Poppe et al., 2010] also used the speaker's speech and gaze and multiple rule-based strategies to determine the timing of

backchanneling. A major drawback of all these models is that they were trained on adult interactions and voices, and thus unlikely to generalize to child-child interactions. Recently, Park *et al.* [Park et al., 2017a] developed a backchannel prediction model based on the non-verbal behaviour they observed in children. They identified both speaker cues (gaze, pitch, energy etc.) which listeners respond to and backchanneling responses (leaning towards, smiling etc.) which indicate engagement, and developed a backchanneling opportunity prediction model based on their findings.

Furthermore, to enable learning companions to be engaging speakers, they must be able to predict when their listeners are likely to be disengaged. Moreover, experiments conducted by Sidner *et al.* [Sidner et al., 2005] confirmed that human listeners were more likely to attend to robots with engagement gestures. They had used connection events like directed & mutual facial gaze, conversational adjacency pairs and backchannels to predict listener disengagement. In another study, Michalowski *et al.* [Michalowski et al., 2006] devised a system that used spatial information from a laser tracker along with head movements to predict engagement. However, the deployment of such a system is not practical, especially in underserved regions of the world. Instead, our work only makes use of a camera which is an easily available and cheap device. In [Bohus and Horvitz, 2009], Bohus *et al.* designed policy-based computational models for managing engagement in a multiparty environment. With multiple agents interacting with the system at a time, they made use of the head pose of the engaged agents, along with several other features to infer the focus of attention and the relationship between the agents to manage engagement. Our LDP model, however, is limited to dyadic interactions at present. Later, Park *et al.* [Park et al., 2017a] studied the impact of annotated visual and prosodic features in analysing listener disengagement. Lastly, Lee *et al.* [Lee et al., 2017] were amongst the first to show that both listener and speaker behaviours must be taken into account to best infer the attentive state of the listener. Drawing inspiration from their work, we use features from both the speaker and listener to predict the attentive state of the latter.

Our work is different from the existing literature in the following ways. First, we use a rich set of automatically extracted features (with the exception of a few hand-annotated features) across visual and vocal modalities. We incorporate features such as Facial Action Units (FAUs) [Sariyanidi et al., 2014], gross head and eye movements, pupil dilation and blink rate, which to the best of our knowledge, have never been used to predict the extent of backchanneling and listener disengagement. These features have, however, been used in other fields of study. For instance, FAUs [Sariyanidi et al., 2014] that detect phenotypical facial expressions, have been widely used in the affective computing literature. Similarly, studies such as [Goswami et al., 2020, D'Mello et al., 2012, Bhatia et al., 2019] have used gross body movements for a variety of tasks such as identifying emotions, cognitive states of children and predicting the severity of depression. The blink rate of the eye has been used to infer alertness [Haq and Hasan, 2016]. Moreover, many behavioural studies have shown that the pupil size is an important social signal which can influence social impressions [Kret, 2018].

Second, most of the existing approaches use a few time-series characterizations, such as the mean and standard deviation of pitch or gaze to predict backchanneling opportunities and disengagement. However, in this study we not only experiment with a wide range of time series characterizations such as approximate entropy, quantiles etc. (Section IV), but also utilize dynamic time series features to predict backchanneling and listener disengagement. In our earlier work [Jindal et al., 2020], we had hypothesized that dynamic time series features are a rich source of information which can help us better predict listener disengagement and backchanneling. In this paper, we devise an algorithm to quantify the importance of these time series features over their aggregates.

III. Methodology

A. Dataset

Datasets of social interactions play an essential role in gaining a deeper understanding of verbal and non-verbal human communication, and consequently in informing the design of social technologies. Multiple datasets exist that capture social interactions between adults [Malisz et al., 2016, de Kok and Heylen, 2011], and adults & children [Rehg et al., 2013, Nojavanasghari et al., 2016]. However, the social interactions of children have only been studied through the lens of adult–child interactions [Singh et al., 2018]. It is reasonable to believe that children interact with their peers differently from adults. Thus, while prior literature offers rich insights into adult–adult and adult–child social interactions, there is only a limited understanding of peer-to-peer interactions in children.

The P2PSTORY dataset is among the first attempts at capturing and analyzing peer-to-peer storytelling interactions in young children of five to six years of age. Like our prior work [Jindal et al., 2020], we use the P2PSTORY dataset to develop engaging and social agents to support early language learning in children. Below we present a brief description of the dataset and the data collection methodology. We urge the interested reader to refer to [Singh et al., 2018] for further details.

In order to collect data from children in a natural environment, the authors recruited eighteen children of a single kindergarten classroom from a public elementary school in Boston. The participants were 5.22 years old on an average and belonged to diverse socio-cultural and economic backgrounds. Most of the children were reported as having "typical development" with a few exceptions. The participants had an average Ages and Stages Questionnaire (ASQ) score of 30. ASQ is a standardized measure to evaluate a child's social and emotional development.

In the storytelling task, each child engaged in three rounds of storytelling with different partners and text-less storybooks. Each storybook had a sequence of coloured pictures with evolving scenes and characters that children could use to develop their own story. Each round differed in terms of the amount of instruction, number of scenes in the story and the type of partner. In a session, a dyad of children took turns to narrate a story to their partner, each turn resulting in a

storytelling episode. The dataset comprised of 29 sessions and 58 storytelling episodes. While each dyad session averaged 15 minutes and comprised of four stages ("Instruction", "Story Construction", "Story Sharing", and "Recall & Questionnaire"), the actual storytelling episode only lasted a little less than one and a half minutes on an average ($\mu = 77$ seconds).

Three cameras captured each storytelling episode, one facing the storyteller, the other facing the listener and the last one capturing a "bird's-eye view" of the dyad. The audio was also recorded using a high-quality microphone and synchronized in time with all the cameras. Multiple annotators coded each episode for a variety of verbal and non-verbal behaviours using ELAN [Brugman and Russel, 2004] (an annotation software). The following behaviours were annotated: gaze, posture, nod, eyebrow movement, mouth, utterances, voicing, on/off task, and attentive state. These behaviours were chosen either because they were found in prior work or were frequently observed in the storytelling sessions. Furthermore, the dataset also captured the participants' socio-demographic parameters such as the age, gender, household income, mother's qualification and results from the ASQ questionnaire etc. to analyze their influence on the acquisition of speaker cues and listener responses.

B. Feature Engineering

We extracted several useful features across the visual and vocal modalities from the high-quality audio and video recordings provided in the P2PSTORY dataset. We used OpenFace [Baltrušaitis et al., 2016] at a frequency of 30 Hz to render a set of highly detailed facial features shown in Table I. In addition to 18 Facial Action Units (FAUs), we extracted the velocity and acceleration of eye and head movements to better capture children's reactions in terms of their gross body movements. Moreover, we also calculated children's blink rate and pupil dilation using eye landmarks. The interested reader may refer our previous work [Jindal et al., 2020] for further details on these features.

Several studies [Jiang et al., 2011] have established that in addition to visual cues, audio features are also useful in increasing emotion recognition accuracy. Therefore, we used OpenSmile with a sampling frequency of 30 Hz, to extract some prosodic features listed in Table I. Specifically, we included the pitch (F0), first and second-order derivatives of Root Mean square energy of the fundamental frequency, and Mel-Frequency Cepstral Coefficients (MFCC), along with their derivatives. All derivatives were computed using Numpy's* gradient function which uses central differences in the interior and first differences at the boundaries with the default distance set to 1.

C. Listener Disengagement Prediction

The ability to detect when the listener is losing attention can be beneficial for a speaker. It can not only help in analyzing when and why the latter felt disengaged, but also allow the speaker to regain the listener's attention. The speaker may do so by commencing an engaging gesture [Sidner et al., 2005] or changing the subject of conversation. With this aim, we model the task of detecting listener disengagement as a time series classification problem.

Features	Description
FAUs	Indicate the presence of 18 Facial Action Units
gaze_vel	Velocity of eye gaze
head_vel_T	Translational velocity of head
head_vel_R	Rotational velocity of head
head_acc_T	Translational acceleration of head
head_acc_R	Rotational acceleration of head
blink_rate	First order differential of Eye Aspect Ration
pupil_dilation	Size of pupil (averaged for both the eyes)
F0	The fundamental frequency computed from the Cepstrum
mfcc′, mfcc″	First and Second order derivatives of Mel-Frequency Cepstral Coefficients 1-12
pcm_RMSenergy′, pcm_RMSenergy″	First and Second order derivatives of Root-mean-square signal frame energy

TABLE I
VISUAL AND VOCAL PROSODIC FEATURES USED IN THE STUDY. WHILE VISUAL FEATURES WERE EXTRACTED FOR BOTH THE SPEAKER AND LISTENER, PROSODIC FEATURES WERE ONLY EXTRACTED FOR THE SPEAKER. FOR A FEATURE X, (X)′ AND (X)″ ARE ITS FIRST AND SECOND ORDER DERIVATIVES.

Inspired by the recent study by Park *et al.* [Park et al., 2017b], we split all the storytelling episodes into disjoint 3 second windows. A 3 second window with 90 time steps or time-ordered observations†, is not only short enough to be considered stationary (and therefore easier to analyze) but also captures enough information to enable us to model its temporal dynamics. We used vocal prosodic features from the speaker, and visual & behavioural features from both the participants to model listener disengagement. We model LDP as a classification problem, by training our models to predict the listener's engagement state at the last (90^{th}) time step using the features of the first 89 time steps of the window.

Before formulating the LDP problem any further, we first had to resolve a mismatch between our feature sets, some of which were sampled at a different frequency. As discussed in Section III-B, the features extracted using OpenSmile and OpenFace were sampled at a frequency of 30 Hz. In contrast, annotated behavioural features from the P2PSTORY dataset were only available at a lower sampling frequency of 5 Hz. We resolved this mismatch by aggregating each annotated feature across all its observations sampled during the time window. Specifically, we introduced dummy variables for each value of a behavioural feature. Each dummy variable was associated with the ratio of time steps where a particular dummy variable was True to the total time steps in a window (89). For example, at any time step, a listener could either have *raised*, *furrow* or *neutral* eyebrows. Therefore, we added three dummy variables replacing the *eyebrows* feature: *eyebrows_raised*, *eyebrows_furrow* and *eyebrows_neutral*. Therefore, we had two categories of features for each 3 second window: visual & vocal multivariate time series features, and a set of aggregated annotated features.

*https://numpy.org/

†Features were sampled at a frequency of 30 Hz (Section III-B).

We could also have down-sampled the OpenFace and OpenSmile features, but that would have meant a lot of lost data. On the other hand, we could also have over-sampled behavioural features. However, we found that these features hardly showed any variation in the chosen 3 second interval, and hence, the dynamics of behavioural features would have had very little information. Therefore, instead of over-sampling, we transformed these behavioural annotations into window-level features by aggregation.

Finally, we provide a mathematical formulation for our LDP model. Let $\mathtt{W}_i$ be the $\mathtt{i}^{\mathtt{th}}$ 3 second window. Also let $\mathtt{T}_j$ be the visual & vocal multivariate time series features for the $\mathtt{j}^{\mathtt{th}}$ time step of $\mathtt{W}_i$. Therefore, the dynamic features for the first 89 steps in the window $\mathtt{W}_i$ can be represented as $[\mathtt{T}_1, \mathtt{T}_2, ..., \mathtt{T}_{89}]_i$. Furthermore, let $\mathcal{A}_i$ be the aggregated annotated features for $\mathtt{W}_i$. Then, for the speaker to predict the engagement state of the listener ($\mathcal{E}_i$) at the $90^{\mathtt{th}}$ time step of window $\mathtt{W}_i$, we need to model the following mapping $\mathcal{F}_{\mathcal{LDP}}$:

$$\mathcal{F}_{\mathcal{LDP}}([\mathtt{T}_1, \mathtt{T}_2, ..., \mathtt{T}_{89}]_i, \mathcal{A}_i) \mapsto \mathcal{E}_i \tag{1}$$

For our experiments, we focused on two widely used algorithms for time series classification: Random Forests and ResNet. We used different sets of time series characteristics (`mean`, `basic`, `tsfresh` etc.) concatenated with the aggregated annotated features as input to our Random Forest models. ResNet is a state-of-the-art deep learning model for time series classification [Fawaz et al., 2019]. We used ResNet to learn the dynamics of the vocal and visual time series features. We also made use of the aggregated annotated features, by combining them with the latent representation of the temporal features learned by ResNet, and then passing them all to the final softmax layer for prediction.

D. Backchanneling Extent Prediction

Backchannel responses act as feedback to speakers, which enables them to analyze the extent to which the listeners understand them [Dennis and Kinney, 1998]. These responses are pivotal to communicating effectively, and therefore an active artificial listener must be able to backchannel appropriately. While predicting the exact timing of backchanneling signals is a challenging task, many studies in the past have proposed models to predict backchanneling opportunities. In fact, recently Park *et al.* [Park et al., 2017b] proposed a backchanneling opportunity prediction model based on child-child interactions. In this work, however, we investigate speaker cues which elicit *high* and *low* levels of backchanneling responses from listeners, instead of predicting their timing.

Like LDP, we model the task of predicting backchanneling extent as a time series classification problem using 3 second windows. However, we only make use of speaker cues (visual and prosodic features) to predict the extent of backchanneling. Since most of the annotated prosodic features (pauses between phrases, speaker's energy, etc.) available in the P2PSTORY dataset had limited variation during episodes, we did not include these features in our analysis, unlike prior work [Park et al., 2017b]. Based on the speaker cues for a particular window $\mathtt{W}_i$, the BEP model learns to predict the extent of listener's backchanneling response in the next window *i.e.,* $\mathtt{W}_{i+1}$. Taking inspiration from Dennis *et al.* [Dennis and Kinney, 1998], we considered both verbal (listener utterances such as *"hmm"*, *"ooh"*, etc.) and non-verbal (smile, nod, onset of partner gaze, lean towards, raising brows) gestures as backchannels for our BEP model. To label a window as having '*high*' or '*low*' backchanneling, we first find the proportion p of time steps wherein the listener generates one or more of the aforementioned backchannel responses. Next, we introduce a backchanneling threshold parameter τ to convert p to binary labels, *high backchanneling* or *low backchanneling* as follows:

$$label_{BC} = \begin{cases} \textit{high backchanneling} & p > \tau \\ \textit{low backchanneling} & p \leq \tau \end{cases} \tag{2}$$

The threshold parameter τ regulates the minimum number of time steps with backchanneling responses (($\tau \times 90$ time steps)) required for the window to qualify as being indicative of *high backchanneling*. For our experiments, we use two values of τ: 0.25 and 0.50. A higher τ includes responses that are relatively more prolonged, and on the other hand, a lower τ includes both subtle and long-lasting backchannels.

Now, we mathematically formulate the BEP problem. Let $\mathtt{W}_i$ be the $\mathtt{i}^{\mathtt{th}}$ 3 second window, and $[\mathtt{T}_1, \mathtt{T}_2, ..., \mathtt{T}_{90}]_i$ be the visual & vocal multi variate time series features for $\mathtt{W}_i$. Further, let $\mathcal{B}_{i+1} \in \{\textit{high backchanneling}, \textit{low backchanneling}\}$ be the target variable indicating the extent of listener backchannels in the following window ($\mathtt{W}_{i+1}$). Then, to predict the extent of backchanneling, our must learn the following mapping $\mathcal{F}_{\mathcal{BEP}}$:

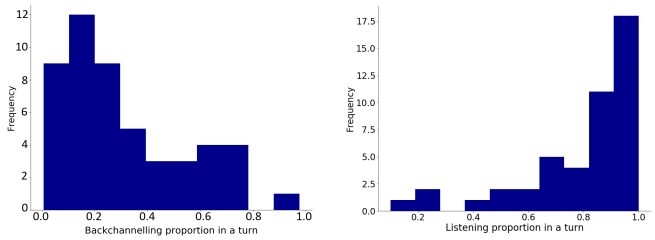
$$\mathcal{F}_{\mathcal{BEP}}([\mathtt{T}_1, \mathtt{T}_2, ..., \mathtt{T}_{90}]_i) \mapsto \mathcal{B}_{i+1} \tag{3}$$

Like LDP, we use both Random Forests and ResNet to predict the extent of backchannel responses.

E. Influence of Socio-demographic factors

Children gradually learn the art of weaving meaningful stories through "culturally-laden" interactions with their parents, peers and teachers. Studies across diverse cultural, socio-demographic and economic contexts have indicated that these factors may influence the development of narrative skills in young children [Curenton, 2010]. Like narrative skills, the acquisition of listener responses may also exhibit individual differences [Dittmann, 1972]. In our earlier work [Jindal et al., 2020], we had examined which socio-demographic and economic features influence the extent of backchanneling. The extent of backchanneling was expressed in terms of *backchanneling proportion* per storytelling episode *i.e.* the fraction of time (time steps) the listener exhibited any form of backchanneling in an episode. Figure 1 *(i)* illustrates the distribution of backchanneling proportion across all the storytelling episodes. For the sake of brevity, we shall not delve any further on the influence of socio-demographic factors on the extent of backchanneling and refer the interested reader to [Jindal et al., 2020] for more details.

In this paper, we build on our prior work [Jindal et al., 2020] and analyze whether socio-demographic and economic factors influence how much time children spend listening to



(i) (ii)

Fig. 1. Distribution of (i) *backchanneling* and (ii) *listener proportion* across all storytelling episodes.

their peers. We quantify the extent of listening in terms of the *listening proportion* or the fraction of time the listener actively listens to the storyteller. Figure 1 *(ii)* shows the distribution of listening proportion across all storytelling episodes.

In order to determine the socio-demographic features influencing the amount of time children spend listening, we conducted statistical tests to discover socio-demographic and economic features causing significant differences in the distribution of listening proportion across all the story-sharing episodes. In particular, for each socio-demographic or economic factor noted in Table II, we carried out a two-sample Kolmogorov-Smirnov (K-S) test under the null hypothesis that the distribution of listening proportion conditioned on its (the socio-demographic factor) values is similar. A two-sample K-S test examines whether the empirical cumulative distribution (ECDF) of a random variable significantly differs across the two samples. For example, consider the gender of the storyteller. The storyteller can either be a `male` or a `female`. Thus, the K-S test either rejects or accepts the null hypothesis that the distribution of listening proportion when the storyteller is male is not significantly different from its distribution when the storyteller is a female. As shown in Table II, the K-S test on the gender of the storyteller rejects the null hypothesis.

A notable exclusion from Table II is the age of the children. This is because all the children were approximately of the same age *i.e.* between five to six years old [Singh et al., 2018]. It must be noted that the K-S test only measures the difference between the empirical distributions of the two samples. Thus, socio-demographic factors having more than two values had to be 'treated'. We binned the values of such factors into '`High`' and '`Low`' buckets or chose the two most frequently occurring values for the K-S test. For example, the participants' ASQ scores were classified as '`High Score`' or '`Low Score`' based on multiple thresholds (Thresh $=$ 20 or 30), such that children having ASQ scores less than the threshold belonged to the '`Low Score`' category and vice versa. Likewise, the participants' total household income was also classified into '`High Income`' (Over $\$100,000$) and '`Low Income`' ($\$30,000$ to $\$75,000$) groups. Furthermore, participants' mother's highest education also had multiple values: `Graduate or professional training`, `College Graduate`, `High school graduate/ GED`, `Some college or vocational school` and `Some high school`, but since most mothers were college graduates, we used the two most frequently-occurring qualifications (`Graduate or professional training` or `College Graduate`) as our two samples. We also carried out randomization tests ($n = 1000$) by randomly permuting feature values and computing the K-S statistic for each randomly permuted feature set. We discuss our findings at a greater depth in Section IV-A.

F. Interpretability: Feature and Time Series Importance

Given the nature of our study, interpretability of our machine learning models is as important as their predictive accuracy. In this section, we provide a brief overview of some techniques which give us insights into features that most influence our models' predictions. While there are many ways to rank essential features, in this study, we primarily use Mean Decrease in Impurity (MDI) [Breiman, 2001] and Mean Decrease in Accuracy (MDA) [Breiman, 2002] also known as permutation feature importance.

MDI evaluates the importance of a feature by measuring the total decrease in impurity as a consequence of splitting on it, weighted by the proportion of samples reaching the node and averaged over all the trees. The importance of a feature is directly proportional to its MDI score. However, a significant drawback of MDI is that it is biased towards high-cardinality features *i.e.* categorical features having many values. To supplement our analysis and prevent misleading conclusions, we also calculated feature relevance using MDA or permutation feature importance. The elegant yet straightforward assumption behind the permutation feature importance algorithm is that randomly shuffling the observations of an irrelevant feature in the testing dataset should only have a negligible impact on the performance of a well-trained model. On the other hand, shuffling the observations of an important feature should confuse the model and result in a significant drop in performance (`accuracy`). Specifically, the MDA for a feature f is the total decrease in a model's `accuracy` as a result of randomly shuffling its observations, averaged over s independent runs of the algorithm. In our work, in addition to accuracy, we also estimated the importance of features using the model's `AUC` and `log-loss`[‡]. It must be noted that MDA is computed with the help of a trained model on the testing data.

The vanilla MDA algorithm [Breiman, 2002] expects testing data to be a two-dimensional matrix having dimensions $\mathcal{O} \times \mathcal{F}$, where $|\mathcal{O}|$ is the number of observations whereas $|\mathcal{F}|$ is the number of features. However, ResNet is trained and tested on a three-dimensional matrix, where each observation is a time series instead of a scalar. To discover important features for our ResNet models, we propose Algorithm 1, which adjusts the shuffling procedure of the vanilla MDA algorithm for three-dimensional datasets and Convolutional Neural Networks. Algorithm 1 takes a trained model $\mathbb{M}$, a three-dimensional feature matrix $\mathbb{X}$, observed values of the target $\mathbb{Y}$ and a scoring metric $\mathbb{S}$ as input, and returns the *importance score* of each feature. The feature matrix $\mathbb{X}$ is of the order $\mathcal{W} \times \mathcal{T} \times \mathcal{F}$ where $|\mathcal{W}|$ is

[‡]Accuracy, AUC and log-loss are referred to as scoring metrics. The log-loss metric was only used for our ResNet models.

Algorithm 1: Permutation Feature Importance

Input: Model ($\mathbb{M}$), Feature Matrix ($\mathbb{X}$), Targets ($\mathbb{Y}$), Scoring metric ($\mathbb{S}$)
Output: *Importance scores* = $\{(f, i)\ \forall f \in$ feature set, i is the corresponding feature importance$\}$

1 **begin**
2 $\quad \mathcal{S}_b \leftarrow$ *Scorer*($\mathbb{M}$, $\mathbb{X}$, $\mathbb{Y}$, $\mathbb{S}$) // `Baseline score`
3 $\quad$ Initialize *Importance scores* $\leftarrow \phi$
4 $\quad$ Initialize $\mathcal{S}_p \leftarrow 0$
5 $\quad$ **for** $f \leftarrow 1$ **to** $|\mathcal{F}|$ **do**
6 $\quad\quad \mathbb{X}_p \leftarrow$ *Shuffle*($\mathbb{X}$, f) // `shuffle feature` f
7 $\quad\quad \mathcal{S}_p \leftarrow$ *Scorer*($\mathbb{M}$, $\mathbb{X}_p$, $\mathbb{Y}$, $\mathbb{S}$)
8 $\quad\quad$ *score difference* $\leftarrow \mathcal{S}_b - \mathcal{S}_p$
9 $\quad\quad$ *Importance scores* $\leftarrow$ *Importance scores* $\cup$ (*score difference*, f)
10 $\quad$ **end**
11 $\quad$ **return** *Importance scores*
12 **end**

Algorithm 2: Time-Series Importance

Input: Model ($\mathbb{M}$), Feature Matrix ($\mathbb{X}$), Targets ($\mathbb{Y}$), Scoring metric ($\mathbb{S}$), Feature list (*feature_list*)
Output: Importance score (*Importance score*)

begin
$\quad \mathcal{S}_b \leftarrow$ *Scorer*($\mathbb{M}, \mathbb{X}, \mathbb{Y}, \mathbb{S}$)
$\quad$ Initialize $\mathcal{S}_p \leftarrow 0$
$\quad$ Initialize $\mathbb{X}_p \leftarrow \mathbb{X}$
$\quad$ **for** $f \in$ *feature_list* **do**
$\quad\quad \mathbb{X}_p \leftarrow$ *Shuffle*($\mathbb{X}_p, f$)
$\quad$ **end**
$\quad \mathbb{M}' \leftarrow$ *Train*($\mathbb{X}_p, \mathbb{Y}$)
$\quad \mathcal{S}_p \leftarrow$ *Scorer*($\mathbb{M}', \mathbb{X}_p, \mathbb{Y}, \mathbb{S}$)
$\quad$ *Importance score* $\leftarrow \mathcal{S}_b - \mathcal{S}_p$
$\quad$ **return** *Importance score*
end

the number of windows (observations), $|\mathcal{T}|$ is number of time steps in a window and $|\mathcal{F}|$ is the total number of features. The scoring metric $\mathbb{S}$ can be any performance metric such as `AUC`, `accuracy` etc. Algorithm 1 first evaluates the performance of the trained model $\mathbb{M}$ using the scoring metric $\mathbb{S}$ (baseline score, $\mathcal{S}_b$) (Line 2). Next, it permutes the feature matrix $\mathbb{X}$ (Line 6) using the `shuffle` procedure and re-evaluates the model on the same scoring metric resulting in the permuted score $\mathcal{S}_p$ (Line 7). The *importance score* of a feature is then the difference between the baseline and permuted score (Lines 8 & 9).

The `shuffle` procedure randomly permutes the time series observations of a given feature based on a randomly generated permutation map. A permutation map $\mathbf{M}$ is simply a $|\mathcal{W}|$-dimensional vector which defines a mapping $\mathbf{M} : \{(1,\ 2,\ ...\ |\mathcal{W}|)\} \rightarrow \mathbb{P}$, where $\mathbb{P}$ denotes the permutation set of $(1,\ 2,\ ...\ |\mathcal{W}|)$, such that if $\mathbf{M}[i] = j,\ (i \neq j)$ then the algorithm installs the time series observation from the i^{th} window of $\mathbb{X}$ to j^{th} window of $\mathbb{X}_p$ for a given feature f. We must note that the time series observations of only one feature are shuffled at a time while the temporal-order of each time series remains the same in both $\mathbb{X}$ and $\mathbb{X}_p$.

We also used Partial Dependency Plots (PDPs) to investigate how some important features influence our target variables. PDPs illustrate the functional relationship between a subset of features and the target variable when the influence of all the remaining features are averaged-out. They are useful in determining whether the relationship between a feature and the target is linear, monotonic or complex [Friedman, 2001].

Finally, to demonstrate whether the dynamics of time series features bear any importance to the predicting capability of our ResNet models, we devised an Algorithm 2 on similar lines as Algorithm 1. The fundamental assumption behind Algorithm 2 is that if the dynamics of time series features significantly influence the predictive performance of ResNet, then destroying the temporal information (by randomly shuffling $\mathbb{X}$ along the time dimension, $\mathcal{T}$) must result in a sharp drop in performance. The elegance of Algorithm 2 however, lies in the invariant that most characteristics of the shuffled time series are unaffected by the permutation. This is because time series characteristics (Section IV) like `mean` and `standard deviation` are independent of the temporal ordering of values in a time series. This invariant entails that ResNet models 'trained' on the permuted data learn from information no more than any Random Forest model.

Algorithm 2 differs from Algorithm 1 in the following ways. Whereas in Algorithm 1, a trained model is evaluated on $\mathbb{X}_p$ in the testing phase, in Algorithm 2, we train a new ResNet model and evaluate it on every permuted feature matrix. This way, we force the ResNet model to depend on information other than the dynamics of time series features. Secondly, we only permute the top $\eta = 10$ most important features found using Algorithm 1 to allow the model to have some temporal information. Finally, the `shuffle` procedures of Algorithms 2 and 1 are slightly different. In Algorithm 2, the top η features are permuted along the time dimension ($\mathcal{T}$) to destroy temporal information unlike Algorithm 1 where features are permuted along the $\mathcal{W}$ dimension and the temporal information of each time series observation remains intact.

IV. Results

We begin our analysis by investigating the influence of socio-demographic and economic features on active listening and backchanneling in Section IV-A. This is followed by an extensive examination of the listener disengagement (LDP) and backchanneling extent (BEP) prediction models in Sections IV-B and IV-C, respectively. Finally, we demonstrate the importance of time series features using Algorithm 2 in Section IV-D.

Sections IV-B and IV-C are structured as follows. We begin by assessing the performance of our models trained on different sets of features inspired by prior work. These experiments not only give us an insight into features integral for the LDP and BEP tasks but also allows us to view our work and feature set, in the context of listener disengagement and backchanneling literature. Next, we experiment with different characteristics of time series as features for Random Forests. In order to use a set of time series $\mathcal{D} = \{T_i\}_{i=1}^{n}$ as features for supervised learning algorithms such as Random Forests, one must first map $\mathcal{D}$ into a feature vector $\mathcal{X} = \{x_1, x_2, ...x_m\}$, where each feature x_j is a scalar value and a characteristic of

Features	Backchanneling		Listening	
	Storyteller	Listener	Storyteller	Listener
Gender	**0.40** *	0.20	0.17	0.22
Participants have same gender	0.13	0.13	**0.32** *	**0.32** *
Mothers' Highest education	0.21	**0.57** *	**0.29** *	0.18
ASQ Score (Thresh = 20)	–	**0.32** *	0.20	0.15
ASQ Score (Thresh = 30)	**0.27** *	**0.28** *	0.27	0.17
Uses words to describe feelings	**0.34** *	**0.32** *	**0.50** *	**0.53** *
Friendly with strangers	0.15	**0.28** *	0.17	0.14
Talks with adults (s)he knows well	0.22	0.20	0.18	0.19
Looks at parent when talking	0.27	**0.59** *	0.25	0.17
Has siblings	0.24	0.23	**0.34** *	0.28
Total household income	0.18	**0.27** *	**0.46** *	0.21

TABLE II
K-S STATISTICS: * INDICATE SIGNIFICANT DIFFERENCES AT 5% SIGNIFICANCE LEVELS.

Metric	Backchanneling		Listening	
	Low	High	Low	High
P	0.66	0.73	0.60	0.55
R	0.75	0.72	0.60	0.58
F1	0.66	0.69	0.57	0.53
AUC	0.84		0.74	

TABLE III
EXTENT OF BACKCHANNELING AND LISTENING USING RANDOM FORESTS WITH SOCIO-DEMOGRAPHIC FEATURES.

Importance socio-demographic features	
Backchanneling	**Listening**
Mother's highest education (L)	Gender (S)
Gender (S)	Mother's highest education (S)
ASQ Score (L)	Household income (S)
Gender (L)	Gender (L)
Household income (L)	Mother's highest education (L)

TABLE IV
MOST IMPORTANT FEATURES THAT PREDICT BACKCHANNELING AND LISTENING ((L): LISTENER, (S): SPEAKER).

some time series $T_i \in \mathcal{D}$. Time series can be effectively and efficiently characterised based on the distribution of their values, correlation properties, stationarity, entropy etc. [Fulcher, 2018, Christ et al., 2018]. In our prior work [Jindal et al., 2020], we had used the arithmetic means of all time series features at the window-level to predict listener disengagement and the extent of backchanneling using Random forests. However, in the present work, we experimented with various different characteristics of time series. In particular, we used two different sets of time series characteristics (or aggregates) for our prediction tasks: `Basic` and `Tsfresh`. The `Basic` set included the arithmetic mean (`M`), standard deviation (`SD`), minimum (`Min`), maximum (`Max`), median (`Md`), first (Q_1) and the third quartile (Q_3) of each time series feature at the window-level.

The `Tsfresh` set on the other hand, included a 'relevant subset' of 63 different time series characteristics such as absolute entropy, kurtosis, skewness etc.[§] and a total of 794 characterizations[¶] (or aggregates) for each time series feature at the window-level. Each aggregate is a time series characteristic with a specific configuration of its parameters. For instance, `approximate_entropy` is a time series characteristic, while `approximate_entropy(m, r)` is an aggregate for an integer `m` (length of compared run of data) and a positive float `r` (filtering level). Since too many irrelevant features may impair the quality of models, `tsfresh` also selects relevant features from the exhaustive set of aggregates it creates based on statistical hypothesis tests [Christ et al., 2018]. The final set of 'relevant' `tsfresh` features included 496 features for the LDP task and only 6 features for the BEP task in total.

Finally, we conclude each section by examining and comparing important features for both our Random Forest and ResNet models. In our earlier work [Jindal et al., 2020], we had only reported important features for the Random Forest models. However, in this paper, we also analyze features which contribute most to the predictive capability of our ResNet models using Algorithm 1.

[§]The interested reader may refer to https://tsfresh.readthedocs.io/en/latest/text/list_of_features.html for a complete list of time series characterization techniques.

[¶]`tsfresh` computes 794 features by default.

In our previous work [Jindal et al., 2020] we had examined the performance of LDP and BEP models across three experimental settings inspired by the hierarchical nature of our data. We had categorised our experiments as "Leave-One Subject Out" or LOSO, "Leave-One Episode Out" or LOEO, and "Random stratified". In the LOSO experiment for instance, a model (LDP or BEP) is trained on data from multiple subjects while it is tested on the left-out subject in each fold. We had observed an expected degradation of our models' performance from random stratified to the LOSO setting due to a reduction in the amount of information sharing between the training and testing sets. Since the LOSO experiment predicts on data from an unseen child, it is the most practically useful experimental setting. Performance in the LOSO experiments represent a lower bound and have been widely used in different studies due to their insights into the practical utility of models [Goswami et al., 2020, Agarwal et al., 2020]. In the following sections, we only report results under the LOSO experimental setting unless stated otherwise. We had previously discussed the impact of the experimental setting in great detail, and we urge the interested reader to refer to [Jindal et al., 2020] for more information.

A. Socio-Demographic Analysis

Table II presents the results of the Kolmogorov-Smirnov (K-S) and randomization tests on all socio-demographic factors. Those factors which led to a significant difference in backchanneling and listening proportion are highlighted and indicated with a *. It can be seen that factors such as how often listeners look at their parents when talking ($\mathcal{D} = 0.591$) and their mother's education ($\mathcal{D} = 0.569$) gave rise to the highest difference in the distributions of backchanneling proportion among all the examined socio-demographic factors. On the other hand, factors such as whether children use words to describe their feelings (Storyteller: $\mathcal{D} = 0.5$, Listener: $\mathcal{D} = 0.527$) and the storyteller's household income ($\mathcal{D} = 0.458$) influence the distribution of listening proportion the most.

To examine the direction of the difference *i.e.* to discover which factors cause an increase in the extent of time children spend listening and backchanneling, we plotted the Empirical Cumulative Distributions (ECDFs) of all socio-demographic factors which yielded significant differences. Figure 2 illustrate

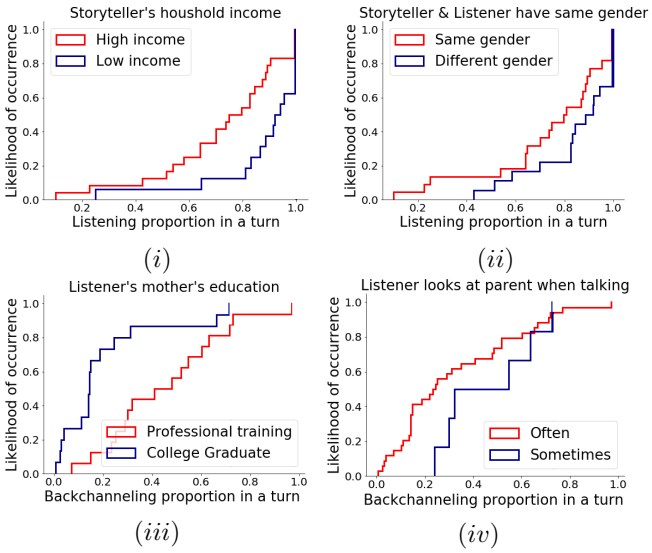


Fig. 2. Empirical CDFs of *listening proportion* corresponding to the (i) storyteller's total household income and (ii) whether both the storyteller and listener have the same gender. ECDFs of *backchanneling proportion* corresponding to the (iii) listeners' mothers' education and (iv) whether they look at their parents when talking to them.

ECDFs of some important socio-demographic factors which influence listening and backchanneling, respectively. Given an empirical cumulative distribution and a backchanneling proportion ($\mathcal{P}_{bc}$), listeners are much more likely to backchannel (frequently) if they `often` look at their parents while talking and their mothers have undergone `graduate or professional training`. Similarly, given a listening proportion ($\mathcal{P}_l$), listeners actively listen more often if they are of the same gender as their partners or belong to a high-income household.

Some other socio-demographic factors which influence the distribution of backchanneling proportion are the storyteller's gender and the listener's household income. The ASQ scores of the storyteller and the listener also yield significant differences. Similarly, responses to questions in the ages and stages questionnaire (ASQ), such as how often does the child use words to describe feelings and whether the child is friendly with strangers also impact the backchanneling proportion.

Correspondingly, aspects of the storytellers' socio-demography such as their household income, mother's education, whether they have siblings and how often they describe their feelings affect listening proportion. It is interesting to note that whether the storyteller and listener are of the same gender influences listening, but it has no bearing on the extent of backchanneling. In contrast, while ASQ scores and responses influence the backchanneling proportion, they result in insignificant differences in the listening proportion. Another exciting yet inexplicable finding is that listeners' socio-demography affects backchanneling more often in comparison to the storyteller's factors which influence the listening proportion more frequently.

Since we found that socio-demographic features indeed influence backchanneling and listening proportion, we trained Random Forest classifiers on these features to predict the extent of backchanneling and listening (‘`High`’ or ‘`Low`’) in a storytelling episode. These experiments allowed us to which discover socio-demographic and economic features best predict the amount of time children spent backchanneling and listening in an episode. To train the Random Forest models, we used all socio-demographic features mentioned in Table II. We labelled episodes with a BC proportion greater than median BC proportion ($\tau_{bc} = 0.23$) as having ‘`High backchanneling`’. Rest of the episodes were labelled as ‘`Low backchanneling`’. Similarly for the listening experiment, we labelled episodes having listening proportion greater than its median ($\tau_l = 0.86$) as ‘`High listening`’, and the rest as ‘`Low listening`’.

The results of these experiments are summarized in Tables III and IV. Table III summarizes the results of these experiments averaged over 5 folds of random stratified cross-validation. These results suggest that the Random Forests models can predict the extent of backchanneling and listening reasonably well, despite having access to a limited set of social and demographic features only.

Furthermore, Table IV lists important features which contribute most to the predictive capabilities of the Random Forest models, found using permutation feature importance (or Mean Decrease in Accuracy, Section III-F). It can be seen that the genders of both the partners in a dyad and the listener's mother's education are the most critical socio-demographic features which predict both, the extent of backchanneling and listening. This may imply that children behave differently depending on the gender of their partners. Moreover, the impact of mother's education on the child is natural, given that mothers have a profound influence on a child's overall development [Tamis-LeMonda and Rodriguez, 2008]. Studies have shown that educated mothers not only read to their children more often but often demonstrate sophisticated language skills which improve the quality of their child's verbal interactions [Scarborough and Dobrich, 1994, Raikes et al., 2006, Rowe et al., 2005]. Another important feature that influences backchanneling and listening is the household income of a child. This finding too follows from prior work which has shown that household income (and poverty) is strongly associated with less stimulating home environments which in turn adversely affects the cognitive and verbal abilities of children [Tamis-LeMonda and Rodriguez, 2008, Garrett et al., 1994].

B. Predicting Listener Disengagement

Table V summarizes results obtained for the Listener Disengagement Prediction task by our Random Forest and ResNet models under different experimental settings. Since we are more interested in accurately predicting disengagement, we only report performance metrics for the not-listening (negative) class. Below we discuss some key findings.

First, we discuss our observations from the feature ablation experiments. We found that adding prosodic features to the annotated ones (gaze, posture, nods etc.) significantly increases the `precision` of Random Forests [`Model 2`] while marginally decreasing its `F1` and `recall`. For the

M.No.	Reference	Features				Aggregates	P	R	F1	AUC
		Annotated	Prosodic	Visual	Demographic		Random Forests			
1	Park *et al.*, 2017 [Park et al., 2017a]	✓				Mean	0.80	0.57	0.65	0.89
2	Lee *et al.*, 2017 [Lee et al., 2017]	✓	✓			Mean	0.96	0.51	0.64	0.92
3		✓	✓	✓		Mean	**0.63**	**0.78**	**0.67**	**0.85**
4		✓	✓	✓		Mean & Stdev	0.96	0.35	0.49	0.93
5		✓	✓	✓		Basic	0.66	0.23	0.33	0.93
6		✓	✓	✓		Tsfresh	0.56	0.15	0.23	0.89
7		✓	✓	✓	✓	Mean	0.92	0.46	0.60	0.94
							ResNet			
8	Park *et al.*, 2017 [Park et al., 2017a]	✓				–	0.75	0.58	0.64	0.77
9	Lee *et al.*, 2017 [Lee et al., 2017]	✓	✓			–	0.35	0.92	0.35	0.64
10		✓	✓	✓		–	**0.67**	**0.79**	**0.70**	**0.89**

TABLE V
LISTENER DISENGAGEMENT PREDICTION: WE REPORT PRECISION (P), RECALL (R), F1-SCORE (F1) AND AUC FOR THE "NOT-LISTENING" CLASS.

ResNet however, the combination results in sharp drop in the F1 and precision, and a significant increase in recall. Therefore, while adding prosodic features did not result in a significant boost performance, their predictive utility cannot be ruled out. In fact, prosodic features such as F0 (Pitch), mfcc[1], mfcc[7] and mfcc[9] are amongst the most important features for ResNet [Figure 3 *(iii)*]. Moreover, including visual features along with the annotated and prosodic features, results in the best model for both Random Forests and ResNet [Model 2 & 9]. This stems from the fact that cognitive states such as engagement often manifest in facial expressions [Goswami et al., 2020]. Finally, the addition of socio-demographic features led to mixed results. While our Random Forest model reported an increase in precision and AUC, there was also a sharp drop in recall [Model 3 vs 7].

Next, we examine the impact of different sets of time series characteristics (or aggregates) on the performance of Random Forests [Models 3 to 6]. Given the same set of features (Annotated, Prosodic & Visual), we found that the arithmetic mean of time series outperforms other sets of aggregates by a substantial margin. Nevertheless, mean and standard deviation resulted in the most precise model (precision $= 0.96$). The reason behind the unsatisfactory performance of the basic and tsfresh feature sets could be that adding numerous highly correlated or irrelevant features result in poor quality models. Besides, tsfresh independently tests each time series characterization for its relevance [Christ et al., 2018]. However, some features may be excellent predictors of the target under investigation in combination with other features. Therefore, while univariate significance tests may be useful in limiting irrelevant features, they may also get rid of individually useless and collective useful features.

In conclusion, while Random Forests trained on different sets of features and time series characteristics gives promising results, ResNet with all the features (Model 10) yields the best listener disengagement prediction model (highest F1 and AUC). This is primarily because ResNet is able to capture the intricacies and dynamics of listener behaviour by learning from time series features.

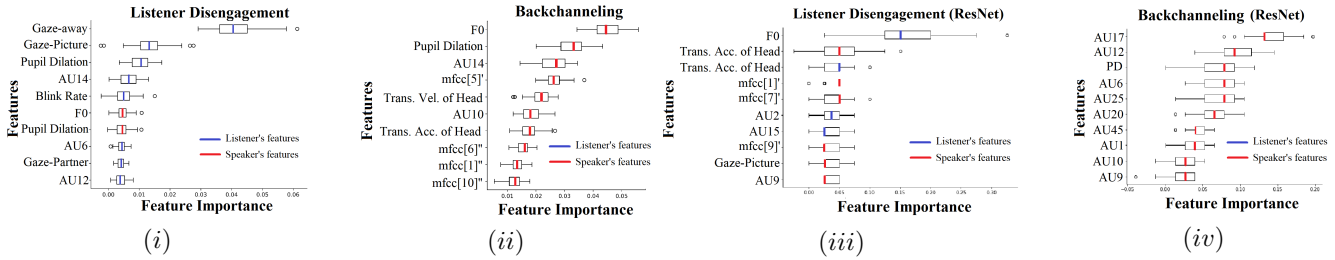
Figure 3 (i) and (iii) show the most important features for the best performing Fandom Forest [Model 3] and ResNet [Model 10]. We used the vanilla MDA algorithm [Breiman, 2001] for Random Forests, and Algorithm 1 for ResNet to compute feature importance across 50 independent runs. The annotated features, Gaze-Away and Gaze-Picture were the most important features for our Random Forest model. This is because listeners gazing-away may be perceived as being distracted [Lee et al., 2017]. The 3-D PDP [Figure 4 (ii)] also supports this conclusion. The PDP plot also reveals that children looking towards the picture are regarded as being attentive. We also found that pupil dilation and blink rate of listeners are important indicators of disengagement. To the best of our knowledge, these features have not been used in predicting disengagement before. But not all annotated features were useful. Features such as laugh, eyebrows, posture and nod were in fact the least important.

For our ResNet model, prosodic features of the speaker (F0 (Pitch), mfcc[1]' and mfcc[7]') turned out to be the most important. This is intuitive because hearing a monotonous speech may be boring. On the other hand, storytellers who modulate their voice may generate excitement and engagement. FAUs (AU02 (*outer brow raiser*) and AU15 (*lip corner depressor*)), and the acceleration of head were some other features which our ResNet model relied on to make predictions. Again, these features have never been used in prior work investigating disengagement and analysing social interactions.

C. Predicting the Extent of Backchanneling

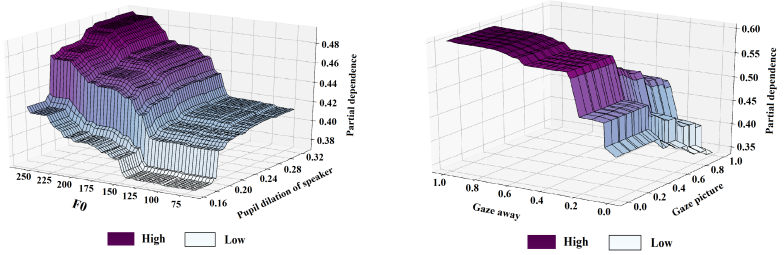
Table VI summarizes the results of our BEP experiments. We record the models' performance in terms of various performance metrics obtained for the *high backchanneling* (positive) class, averaged across all LOSO test splits. In the next few paragraphs, we discuss our key observations from these results.

First, we compare the performance of our Random Forest models for different values of the backchanneling threshold parameter *i.e.* for $\tau = 0.25$ and 0.50 [Models 1 to 8]. We observed an interesting trend in their precision and recall values. Specifically, most models (6 out of 8) trained for $\tau = 0.25$, had lower precision and higher recall in comparison to corresponding models trained for $\tau = 0.50$ *i.e.* $P_{\tau=0.25} < P_{\tau=0.50}$, and $R_{\tau=0.25} > R_{\tau=0.50}$. Thus, when τ drops from 0.50 to 0.25, there is an increase in the number of true positives, but at the cost of an increase in false positives. We can explain this trend based on the average backchanneling proportion of each subject (Figure 5) and our models' expectations. From Figure 5 it can be clearly seen that



(i) (ii) (iii) (iv)

Fig. 3. Permutation Feature Importance (MDA): Top 10 most important features for (i, ii) Random Forests and (iii, iv) ResNet.



(i) (ii)

Fig. 4. Partial Dependence Plots: (i) Combined effect of speaker's F0 and Pupil Dilation on backchanneling signals (ii) Impact of the proportion of time when the listener is gazing towards the picture (Gaze-Picture) and away (Gaze-Away) on listener's disengagement.

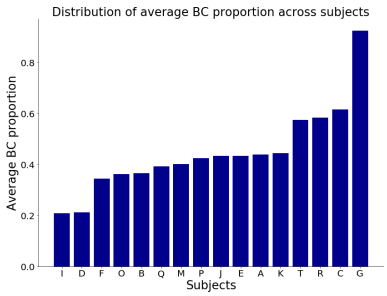


Fig. 5. Average *backchanneling proportion* for each subject across different episodes.

nearly 75% of our subjects spent less than 40% of an episode backchanneling on an average. In fact, only one subject (Subject *G*) had an average backchanneling proportion well in excess of 0.50. On the other hand, all children spent more than 25% of an episode backchanneling, except subjects *I* and *D*. Therefore, when a model is trained with the backchanneling threshold set to 0.25, it expects children to backchannel a lot, and occasionally predicts "*low backchanneling*" as "*high backchanneling*". Nevertheless, these models always predict "*high backchanneling*" windows accurately and hence have high `recall`. On the other hand, when $\tau = 0.50$, the model expects children to rarely backchannel to a great extent. Thus, models trained with $\tau = 0.50$ are more likely to predict "*high backchanneling*" as "*low backchanneling*", especially when the BC proportion is closer to 0.50. However, these models predict "*high backchanneling*" windows with a reasonably high BC proportion with high `precision`.

However, we expect our models to have a perfect balance between `precision` and `recall` regardless of the backchanneling threshold. In this regard, ResNet performed equally well on both the thresholds. Nonetheless, the model performed slightly better in terms of `precision` and `recall` at $\tau = 0.25$. The superior performance of ResNet may be attributed to the fact that it uses more information in the form of time series dynamics to predict the extent of backchanneling.

Next, we focus on our feature ablation experiments wherein we compared the performance of our models using different sets of features inspired by prior work. We first examine the performance of Random Forests models [`Models` 1 to 8]. For both the thresholds ($\tau = 0.25$ and 0.50), all the feature ablations showed similar variations. We observed that the performance of our model improved when trained on all prosodic features, in addition, to pitch [`Model` 1 vs 2]. Other prosodic features like `mfcc[5]'`, `mfcc[6]''`, `mfcc[1]''` and `mfcc[10]''` are in fact, some of the most important features for our Random Forests (Figure 3 (ii)). Adding annotated gaze features to vocal prosody also led to a slight improvement in performance [`Model` 3]. Finally, the best Random Forest model was the one trained on all visual and prosodic features [`Model` 5 for $\tau = 0.25$ and `Model` 4 for $\tau = 0.50$]. The addition of socio-demographic features to visual and prosodic features results in a boost in performance, especially in the case of $\tau = 0.25$. However, `Model` 8 is only of theoretical interest. It has limited practical utility since socio-demographic features may not always be available and their use may raise privacy concerns.

Again for ResNet, results for both values of τ showed the same trend. An interesting observation was that pitch alone was a reasonable predictor of the backchanneling extent [`Model` 9]. Using all other prosodic as well as annotated gaze features results in a slight albeit insignificant improvements [`Model` 10 & 11]. This is because prosodic and annotated gaze features are not as crucial for ResNet as they are for Random Forests, as shown in Figure 3 (iv). Like Random Forest, ResNet trained on all the visual and prosodic features performed the best overall [`Model` 12]. In fact, we found that visual features and especially FAUs were some of the most important for ResNet.

In addition to different feature sets, we also experimented with multiple sets of time series characteristics for Random Forests. The results are consistent with our findings from corresponding LDP experiments discussed in the last section. Given the same set of features (Visual & Prosodic), when τ is set to 0.25, a combination of mean and standard deviation yields the best performance [`Model` 5]. On the other hand,

M.No.	Reference	Features				Aggregates	$\tau = 0.25$				$\tau = 0.5$			
		Annotated	Prosodic	Visual	Demographic		P	R	F1	AUC	P	R	F1	AUC
							Random Forests							
1	Ward, 1996 [Ward, 1996]		✓(Pitch)			Mean	0.61	0.71	0.65	0.66	0.62	0.45	0.32	0.66
2	Jiang, 2011 [Jiang et al., 2011]		✓			Mean	0.77	0.84	0.80	0.86	0.92	0.51	0.40	0.83
3	Lee *et al.*, 2017 [Lee et al., 2017]; Morency *et al.*, 2010 [Morency et al., 2010]	✓(Gaze)	✓			Mean	0.75	0.83	0.81	0.85	0.92	0.44	0.50	0.84
4			✓	✓		Mean	0.90	0.57	0.68	0.87	**0.66**	**0.98**	**0.78**	**0.86**
5			✓	✓		Mean & Stdev	**0.71**	**0.97**	**0.81**	**0.88**	0.90	0.55	0.58	0.86
6			✓	✓		Basic	0.60	0.25	0.30	0.88	0.97	0.58	0.63	0.88
7			✓	✓		Tsfresh	0.66	0.83	0.73	0.73	0.64	0.65	0.55	0.74
8			✓	✓	✓	Mean	0.83	0.88	0.85	0.92	0.85	0.51	0.62	0.88
							ResNet							
9	Ward, 1996 [Ward, 1996]		✓(Pitch)			–	0.52	0.96	0.67	0.55	0.47	0.70	0.49	0.55
10	Jiang, 2011 [Jiang et al., 2011]		✓			–	0.57	0.70	0.68	0.62	0.60	0.69	0.51	0.58
11	Lee *et al.*, 2017 [Lee et al., 2017]; Morency *et al.*, 2010 [Morency et al., 2010]	✓(Gaze)	✓			–	0.55	0.67	0.70	0.63	0.45	0.66	0.55	0.52
12			✓	✓		–	**0.72**	**0.91**	**0.83**	**0.90**	**0.70**	**0.89**	**0.80**	**0.78**

TABLE VI
BACKCHANELING EXTENT PREDICTION: WE REPORT THE PRECISION (P), RECALL (R), F1-SCORE (F1) AND AUC FOR "HIGH BACKCHANNELING" CLASS.

for $\tau = 0.50$ arithmetic mean outperforms all other sets of aggregates [Model 4].

Next, we compare the performance of the Random Forests and ResNet trained on the same features. We observed that ResNet performed better than Random Forests in most settings, except for models trained on prosodic features [Model 2], and a combination of gaze and prosodic features [Model 3] for $\tau = 0.25$. This demonstrates that modelling the dynamics of time series features indeed helps in predicting the extent of backchanneling.

Finally, Figure 3 (ii) and (iv) present the distribution of *importance scores* of the 10 most important features for our Random Forests and ResNet model, respectively, found using Permutation Feature Importance (MDA). For Random Forests, the pitch (F0) of the speaker is the most important factor influencing listener backchannels. This also follows from prior work in predicting listener backchannels [Ward and Tsukahara, 2000, Noguchi and Den, 1998]. Besides, we also observe that the speaker's *pupil dilation*, AU10 (*upper lip raiser*), AU14 (*dimpler*), and *translational head acceleration* are strong predictors of backchanneling responses. To the best of our knowledge, none of these features have been used to predict backchanneling before. To further analyze the combined influence of F0 and *pupil dilation* (best predictors of backchanneling) on the extent of backchanneling, we plotted the 3-D PDP shown in Figure 4 (i). It is evident that both these features have a positive impact on seeking backchannel responses. It is interesting to note that around the range of $0.20 - 0.22$, the speaker's *pupil dilation* causes a sharp increase in the listener's backchanneling response. In the case of ResNet, the speaker's FAUs and *pupil dilation* are the most important features. Unlike Random Forests, prosodic features such as mfcc′ are not strong predictors of backchanneling for ResNet.

D. Importance of time series features

We conclude this section by presenting the results of a short yet important study that validates our hypothesis that the dynamics of time series features are rich sources of information to predict disengagement and backchanneling. Specifically, we used Algorithm 2 to find the importance of time series features on the predictive utility of our LDP and BEP models. We ran the algorithm for 20 independent iterations for both the tasks, and in each iteration, the ResNet model was trained and evaluated using LOSO cross-validation on the permuted dataset (Section III-F). Overall, we observed that there was a significant decrease in the performance of both the LDP and BEP models when trained and tested on the permuted data, as compared to the baseline results [Models 10 & 12] discussed in the previous sections. The LDP model reported an average drop of 29.8% in F1 and 28.6% in AUC. On the other hand, the average drop in F1 and AUC for the BEP model was 36.4% and 25.7%, respectively. This establishes the fact that modeling the temporal characteristics of the features significantly improved the performance of the ResNet model for both the tasks.

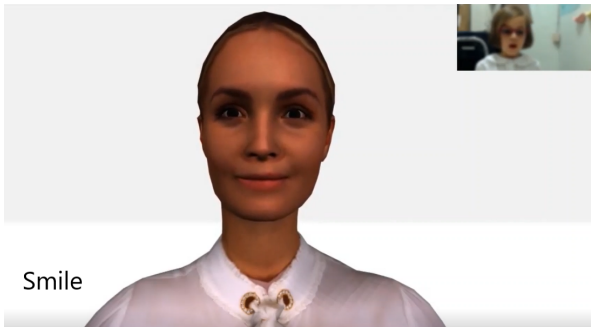


Fig. 6. *Alice, a social listener.* We deployed our BEP model on a 3-D avatar called *Alice*. Alice uses the model to identify backchanneling opportunities.

V. CONCLUSION

In this work, we proposed two models to help peer learning robots come across as active listeners and engaging storytellers to foster early language development in children. Specifically, we formulated Listener Disengagement (LDP) and a Backchanneling Extent Prediction (BEP) models. Extensive experiments with different sets of features and time series characteristics yielded exciting results. To lend interpretability to our models, we used Permutation Feature Importance in addition to Partial Dependency Plots to find important features for our models. We found that visual and vocal features such as *blink rate*, *pupil dilation*, *velocity & acceleration of head*, *FAUs*, and MFCC coefficients were extremely useful in

predicting disengagement and backchanneling extent. To the best of our knowledge, these features have not been used by studies in the past. We also established the utility of time series features in predicting disengagement and the extent of backchanneling. Further, we also examined the influence of socio-demographic features on the amount of time children spent backchanneling and listening to their peers.

To evaluate our models as potential real-time systems, we deployed the best performing LDP and BEP models to predict disengagement and backchanneling during some episodes in the P2PSTORY dataset. We deployed our backchanneling extent prediction model as a lightweight asynchronous python-flask application on a 3-D avatar[‖], called *Alice*. Alice uses our BEP model to identify backchanneling opportunities and chooses a backchannel response (`smile`, `nod`, `head shake`, `eyebrow raise`) at random[**].

A limitation of our work is that we did not evaluate our models on a real social robot, and believe that further experiments are essential to corroborate their effectiveness. However, since children consider robots as social beings [Kahn Jr et al., 2012] and readily learn new information from them [Movellan et al., 2009], we believe that our findings from analysing child-child dyadic interactions will provide useful lessons in improving child-robot interactions.

Acknowledgment

The authors would like to thank Prof. Rajni Jindal for her insightful feedback.

References

Mansi Agarwal, Maitree Leekha, Ramit Sawhney, Rajiv Ratn Shah, Rajesh Kumar Yadav, and Dinesh Kumar Vishwakarma. Memis: Multimodal emergency management information system. In *European Conference on Information Retrieval*, pages 479–494. Springer, 2020.

G Aist and J Mostow. When speech input is not an afterthought: A reading tutor that listens. In *Workshop on Perceptual User Interfaces*, 1997.

G Aist, J Mostow, B Tobin, P Burkhead, A Corbett, A Cuneo, B Junker, and MB Sklar. Computer-assisted oral reading helps third graders learn vocabulary better than a classroom controlabout as well as one-on-one human-assisted oral reading, 2001.

Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. Openface: an open source facial behavior analysis toolkit. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–10. IEEE, 2016.

Shalini Bhatia, Roland Goecke, Zakia Hammal, and Jeffrey F Cohn. Automated measurement of head movement synchrony during dyadic depression severity interviews. In *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, pages 1–8. IEEE, 2019.

Dan Bohus and Eric Horvitz. Models for multiparty engagement in open-world dialog. In *Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 225–234. Association for Computational Linguistics, 2009.

Cynthia Breazeal, Kerstin Dautenhahn, and Takayuki Kanda. Social robotics. In *Springer handbook of robotics*, pages 1935–1972. Springer, 2016.

[‖]We obtained a 3-D model of a middle-aged female from https://renderpeople.com/free-3d-people/ for our demo.

[**]Predicting the type of backchannel response is beyond the scope of our work.

Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

Leo Breiman. Manual on setting up, using, and understanding random forests v3.1, 2002. URL http://oz.berkeley.edu/users/breiman/Using_random_forests_V3.

Hennie Brugman and Albert Russel. Annotating multi-media/multi-modal resources with ELAN. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal, May 2004. European Language Resources Association (ELRA). URL http://www.lrec-conf.org/proceedings/lrec2004/pdf/480.pdf.

Maximilian Christ, Nils Braun, Julius Neuffer, and Andreas W Kempa-Liehr. Time series feature extraction on basis of scalable hypothesis tests (tsfresh–a python package). *Neurocomputing*, 307: 72–77, 2018.

Herbert H Clark, Susan E Brennan, et al. Grounding in communication. *Perspectives on socially shared cognition*, 13(1991):127–149, 1991.

Stephanie M. Curenton. Narratives as learning tools to promote school readiness. *Early Education and Development*, 21(3):287–292, 2010. doi: 10.1080/10409289.2010.485532. URL https://doi.org/10.1080/10409289.2010.485532.

Iwan de Kok and Dirk Heylen. The multilis corpus–dealing with individual differences in nonverbal listening behavior. In *Toward Autonomous, Adaptive, and Context-Aware Multimodal Interfaces. Theoretical and Practical Issues*, pages 362–375. Springer, 2011.

Alan R Dennis and Susan T Kinney. Testing media richness theory in the new media: The effects of cues, feedback, and task equivocality. *Information systems research*, 9(3):256–274, 1998.

Allen T Dittmann. Developmental factors in conversational behavior. *Journal of Communication*, 22(4):404–423, 1972.

Sidney D'Mello, Rick Dale, and Art Graesser. Disequilibrium in the mind, disharmony in the body. *Cognition & emotion*, 26(2): 362–374, 2012.

Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 33(4):917–963, 2019.

Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

Ben D Fulcher. Feature-based time-series analysis. In *Feature engineering for machine learning and data analytics*, pages 87–116. CRC Press, 2018.

Patricia Garrett, Nicholas Ng'andu, and John Ferron. Poverty experiences of young children and the quality of their home environments. *Child development*, 65(2):331–345, 1994.

Mononito Goswami, Lujie Chen, and Artur Dubrawski. Discriminating cognitive disequilibrium and flow in problem solving: A semi-supervised approach using involuntary dynamic behavioral signals. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34. AAAI, 2020.

Zeeshan Ali Haq and Ziaul Hasan. Eye-blink rate detection for fatigue determination. In *2016 1st India International Conference on Information Processing (IICIP)*, pages 1–5. IEEE, 2016.

Dongmei Jiang, Yulu Cui, Xiaojing Zhang, Ping Fan, Isabel Ganzalez, and Hichem Sahli. Audio visual emotion recognition based on triple-stream dynamic bayesian network models. In Sidney D'Mello, Arthur Graesser, Björn Schuller, and Jean-Claude Martin, editors, *Affective Computing and Intelligent Interaction*, pages 609–618, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg. ISBN 978-3-642-24600-5.

Rajni Jindal, Maitree Leekha, Minkush Manuja, and Mononito Goswami. What makes a better companion? towards social & engaging peer learning. In *Proceedings of the Twenty-forth European Conference on Artificial Intelligence*. IOS Press, 2020.

Peter H Kahn Jr, Takayuki Kanda, Hiroshi Ishiguro, Nathan G Freier, Rachel L Severson, Brian T Gill, Jolina H Ruckert, and Solace Shen. robovie, you'll have to go into the closet now: Children's social and moral relationships with a humanoid robot. *Developmental psychology*, 48(2):303, 2012.

Mariska E Kret. The role of pupil size in communication. is there room for learning? *Cognition and Emotion*, 32(5):1139–1145, 2018.

Jin Joo Lee, Cynthia Breazeal, and David DeSteno. Role of speaker cues in attention inference. *Front. Robotics and AI*, 4:47, 2017. ISSN 2296-9144. doi: 10.3389/frobt.2017.00047. URL https://www.frontiersin.org/article/10.3389/frobt.2017.00047.

Zofia Malisz, Marcin Włodarczak, Hendrik Buschmeier, Joanna Skubisz, Stefan Kopp, and Petra Wagner. The alico corpus: analysing the active listener. *Language resources and evaluation*, 50(2):411–442, 2016.

Andrew A. McReynolds, Sheba P. Naderzad, Mononito Goswami, and Jack Mostow. Toward learning at scale in developing countries: Lessons from the global learning xprize field study. In *Provisionally accepted to the Seventh Annual ACM Conference on Learning at Scale*. ACM, 2020.

Marek P Michalowski, Selma Sabanovic, and Reid Simmons. A spatial model of engagement for a social robot. In *9th IEEE International Workshop on Advanced Motion Control, 2006.*, pages 762–767. IEEE, 2006.

Louis-Philippe Morency, Iwan de Kok, and Jonathan Gratch. A probabilistic multimodal approach for predicting listener backchannels. *Autonomous Agents and Multi-Agent Systems*, 20(1):70–84, 2010.

Jack Mostow, Greg Aist, Paul Burkhead, Albert Corbett, Andrew Cuneo, Susan Eitelman, Cathy Huang, Brian Junker, Mary Beth Sklar, and Brian Tobin. Evaluation of an automated reading tutor that listens: Comparison to human tutoring and classroom instruction. *Journal of Educational Computing Research*, 29(1):61–117, 2003.

Javier Movellan, Micah Eckhardt, Marjo Virnes, and Angelica Rodriguez. Sociable robot improves toddler vocabulary skills. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 307–308. ACM, 2009.

Hiroaki Noguchi and Yasuharu Den. Prosody-based detection of the context of backchannel responses. In *Fifth International Conference on Spoken Language Processing*, 1998.

Behnaz Nojavanasghari, Tadas Baltrušaitis, Charles E Hughes, and Louis-Philippe Morency. Emoreact: a multimodal approach and dataset for recognizing emotional responses in children. In *Proceedings of the 18th acm international conference on multimodal interaction*, pages 137–144, 2016.

Hae Won Park, Mirko Gelsomini, Jin Joo Lee, and Cynthia Breazeal. Telling stories to robots: The effect of backchanneling on a child's storytelling. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, pages 100–108. IEEE, 2017a.

Hae Won Park, Mirko Gelsomini, Jin Joo Lee, and Cynthia Breazeal. Telling stories to robots: The effect of backchanneling on a child's storytelling. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, pages 100–108. IEEE, 2017b.

Carole Peterson, Beulah Jesso, and Allyssa McCabe. Encouraging narratives in preschoolers: An intervention study. *Journal of child language*, 26(1):49–67, 1999.

Ronald Poppe, Khiet P Truong, Dennis Reidsma, and Dirk Heylen. Backchannel strategies for artificial listeners. In *International Conference on Intelligent Virtual Agents*, pages 146–158. Springer, 2010.

Helen Raikes, Barbara Alexander Pan, Gayle Luze, Catherine S Tamis-LeMonda, Jeanne Brooks-Gunn, Jill Constantine, Louisa Banks Tarullo, H Abigail Raikes, and Eileen T Rodriguez. Mother–child bookreading in low-income families: Correlates and outcomes during the first three years of life. *Child development*, 77(4):924–953, 2006.

James Rehg, Gregory Abowd, Agata Rozga, Mario Romero, Mark Clements, Stan Sclaroff, Irfan Essa, O Ousley, Yin Li, Chanho Kim, et al. Decoding children's social behavior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3414–3421, 2013.

Meredith L Rowe, Barbara Alexander Pan, and Catherine Ayoub. Predictors of variation in maternal talk to children: A longitudinal study of low-income families. *Parenting: Science and Practice*, 5(3):259–283, 2005.

Evangelos Sariyanidi, Hatice Gunes, and Andrea Cavallaro. Automatic analysis of facial affect: A survey of registration, representation, and recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(6):1113–1133, 2014.

Hollis S Scarborough and Wanda Dobrich. On the efficacy of reading to preschoolers. *Developmental review*, 14(3):245–302, 1994.

Candace L Sidner, Christopher Lee, Cory Kidd, Neal Lesh, and Charles Rich. Explorations in engagement for humans and robots. *arXiv preprint cs/0507056*, 2005.

Nikhita Singh, Jin Joo Lee, Ishaan Grover, and Cynthia Breazeal. P2pstory: Dataset of children as storytellers and listeners in peer-to-peer interactions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 434:1–434:11, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-5620-6. doi: 10.1145/3173574.3174008. URL http://doi.acm.org/10.1145/3173574.3174008.

Catherine S Tamis-LeMonda and Eileen T Rodriguez. Parents role in fostering young childrens learning and language development. *Encyclopedia on early childhood development*, 1:1–11, 2008.

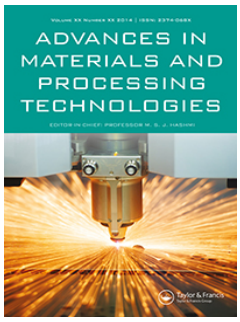
Keith Topping and Stewart Ehly. *Peer-assisted learning*. Routledge, 1998.

Peter Tymms, Christine Merrell, Allen Thurston, John Andor, Keith Topping, and David Miller. Improving attainment across a whole district: school reform through peer tutoring in a randomized controlled trial. *School Effectiveness and School Improvement*, 22(3):265–289, 2011.

Nigel Ward. Using prosodic clues to decide when to produce back-channel utterances. In *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96*, volume 3, pages 1728–1731. IEEE, 1996.

Nigel Ward and Wataru Tsukahara. Prosodic features which cue back-channel responses in english and japanese. *Journal of pragmatics*, 32(8):1177–1207, 2000.

Victor H Yngve. On getting a word in edgewise. In *Chicago Linguistics Society, 6th Meeting, 1970*, pages 567–578, 1970.



Advances in Materials and Processing Technologies

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tmpt20

Tribological characterisation of thermal sprayed CrC alloyed coating – A review

Navendu Gupta , Shailesh Kumar Singh & Shailesh Mani Pandey

To cite this article: Navendu Gupta , Shailesh Kumar Singh & Shailesh Mani Pandey (2020): Tribological characterisation of thermal sprayed CrC alloyed coating – A review, Advances in Materials and Processing Technologies, DOI: 10.1080/2374068X.2020.1793262

To link to this article: https://doi.org/10.1080/2374068X.2020.1793262

Published online: 15 Jul 2020.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tmpt20

ADVANCES IN MATERIALS AND PROCESSING TECHNOLOGIES
https://doi.org/10.1080/2374068X.2020.1793262

Taylor & Francis
Taylor & Francis Group

Check for updates

Tribological characterisation of thermal sprayed CrC alloyed coating – A review

Navendu Gupta[a], Shailesh Kumar Singh[b] and Shailesh Mani Pandey[c]

[a]Department of Mechanical Engineering, Delhi Technological University, New Delhi, India; [b]Department of Mechanical Engineering, Indian Institute of Technology, Dhanbad, India; [c]Department of Mechanical Engineering, National institute of Patna, Bihar, India

ABSTRACT

High-Velocity Oxygen Fuel (HVOF) and Plasma spraying coating procedure are utilised to enhance the properties wear, COF, and corrosion of different mechanical parts. Further, the aim of this review is to the assessment of microstructure and mechanical properties of CrC based compound coatings sprayed by HVOF and plasma spraying strategy onto distinctive substrate materials and to see the capability of CrC based coatings for the piston rings application. The wear and corrosion resistance has expanded because of dispersed components such as Ni and Cr carbide phases present in the coated materials. The HVOF procedure particularly cermet coatings in which powder materials needed to accelerate to store on substrate materials brought high wear and destructive resistance. It was additionally observed that the coating qualities were to a great extent influenced by the properties of the substrate material. The few blends of CrC-based compounds can create great outcomes than WC-Co powder coating. The thermal spray coating procedures and CrC based combinations for the piston ring application have been accounted for like weight reduction and wear rate and COF. It can be expressed that the properties of plasma-sprayed and HVOF sprayed CrC based composite coatings have comparative properties for piston ring application.

ARTICLE HISTORY
Accepted 5 July 2020

KEYWORDS
HVOF; plasma spray; CrC; microstructure; wear; COF

1. Introduction

The resistance for movement, Friction has been a battle all over history for humanity. Essential creations in man's soonest push to beat friction were to utilise water and later normal oils to grease up moving contacts. Friction, wear, and oil were distinguished to impact the effectiveness and lifetime of apparatus in the business and in this manner the world economy [1,2]. These components drove an administration council in the United Kingdom in 1966 led by H.P. Jost to explore the matter. The 'Jost Report (1965)' assessed that 515 million pounds sterling could be spared every year in the United Kingdom by completing progressed tribological solutions [2,3]. The report coined the word 'tribology' for the science and technology of interacting surfaces in relative motion and included various suggestions for corrective actions. A strategy for energy conservation is necessary

CONTACT Navendu Gupta navendug19@gmail.com Delhi Technological University, House No.- 16-17 Block No. I- 11 Upper Ground, Sector 16 Rohini, New Delhi - 110089, India

© 2020 Informa UK Limited, trading as Taylor & Francis Group

through tribology in the areas of road transportation, power generation, industrial machinery that accounts for 80% of the total power consumption [4]. The petroleum and automotive industries are facing fierce international competition, government regulations, and rapid technological changes. Ever-increasing government regulations require improved fuel economy and lower harmful emissions from the lubricant systems as well as automotive fuel. Better fuel-efficient vehicles and higher energy-conserving engine oils are becoming important in the face of both reductions of engine fuel consumption and preservation of environmental resources. Only 21.5% of the total fuel energy is available for car's motion [5,6,7]. The major part of the fuel energy is consumed to overcome friction based on the data available in the literature 28% (excluding rolling resistance) of the total fuel energy is used to overcome engine friction losses including piston skirt friction, piston rings, and bearings. Out of which 68% loss in friction between the piston ring and cylinder liner contact only. Reducing this friction and wear may lead to decrease in fuel consumption hence increase efficiency and reliability of the vehicle.

The various ways to reduce piston ring friction and wear is to use better lubricants with friction modifiers. The addition of friction modifiers lubricants has been shown to significantly reduce the mechanical losses of critical components in internal combustion engine; thereby improving fuel economy. Such as al_2O_3 and TiO_2 nano-lubricants reduced the friction by 40–50% and wear rate by 21–29% [8–10]. Although spherical nanoparticles can act like nano-size bearings and abrasive particles between the sliding interfaces. Though they reduce friction, they cause abrasive wear on the surface. The other way is to use Functional coatings and surface treatments that offer several possibilities to improve the sliding properties of metal surfaces. In IC engines, coatings and surface treatments are commonly applied to one or several components of the piston, piston ring and cylinder liner system [11]. There is a range of coatings and surface treatments currently used in engine components in the reciprocating system that covers a variety of different coating compositions and deposition techniques. Two different groups of coating application techniques are commercially used for the deposition of wear-resistant coatings on piston rings, namely thermal spraying and galvanic coating. A third, widely used surface treatment method for piston rings is surface nitriding [12,13].Nowadays, due to the rapid technological developments, metallic coatings need, to maintain their competitiveness and to find new possible applications, to be environmentally friendly and to enhance their properties both chemical and mechanical. In particular, galvanic processes can be considered highly polluting. Undoubtedly, from an environmental point of view, one of the most critical galvanic processes is the hard chromium plating. Hard chromium plating technology is used today in the industry to coat different types of critical mechanical components (valves, pistons, piston rings, rods, hydraulic components) due to the good wear and corrosion resistance of this coating. Though chromium-plating causes effects on human health because of the use of substances (hexavalent chrome ($Cr6^+$)) in the galvanic coating process whose toxicological features have been recognised. Carcinogenic effects were observed; a common example of a professional disease in the galvanic field is the nasal septum perforation [14,15,16].Therefore, intensive researches have been attracted to the possible alternatives for replacing the electroplated hard chromium coatings on piston ring. Although nowadays thermally sprayed coating are proven to successful for lowering the COF and wear rate for piston rings as a potential replacement of electroplated chromium coatings [17,18,19].

Various methods of thermal coating on piston rings being used today are high-velocity oxy-fuel, plasma spraying, physical vapour deposition, chemical vapour deposition [20,21,22].Thermal spraying technologies are important in a variety of different industries for the deposition of antiwear coatings based on containing metallic, carbide, ceramic or composite.various composite coating have proven to beneficial for reducing COF and wear rates.TiSiCN nanocomposite coatings deposited by enhanced plasma magnetron sputtering (PEMS) showed excellent adhesion and dry coefficient of friction in the range of 0.17 to 0.2 using a ball-on-disk tribometer. The gasoline engine test showed that the uncoated piston rings resulted in 25% to 34% of the frictional losses in two separate baseline engine tests. On the contrary, the coated rings contributed to 18% of the total frictional loss in the engine test. The diesel engine endurance test showed a 28% and 40% lower ring weight loss for the coated top and second rings, respectively, when compared to the uncoated baseline. Also, the cylinder liner, which was coated, showed an average 50% lower wear than that in the uncoated baseline engine test [23,24,25]. An investigation shows that PVD CRN hard coatings can reduce wear rates very effectively. The results shown that the coating's exhibit higher hardness and lower COF and wear rate than conventional coatings [26–28].Another research shows that molybdenum composite coatings can reduce wear loss on piston rings. Mo composition in the composite coatings significantly contributed to reduce friction and wear at high temperature 450°C and increased hardness than conventional coatings [29–31]. Another type of coating that exhibits excellent tribological characteristics are diamond-like carbon (DLC) coatings. Which usually show excellent wear, high hardness, high corrosion resistance, high chemical and thermal stability, low friction and excellent running property [32–34].

The development of physical vapour deposition (PVD) coatings of low friction, high hardness, and thin diamond-like carbon (DLC) has offered new possibilities for friction reduction in cars there is 30% decrease in friction in lubricated contacts by applying DLC compared to a steel contact [35,36,37].The utilisation CrC-NiCr based composite coatings likewise demonstrated a diminishment of wear and COF, high hardness and possibly high-temperature applications. Thin film coating is one of the most common coatings to be used in the tribological improvements of piston rings, a hydraulic piston, bearing shaft, etc. CrN Coating of a single layer by the PVD often containing one structure phase for piston rings is most common and commercially introduced [38,39]. As per the concern of applications of the study, the researcher and industry persons are looking towards more reliable solutions in the automobile industry.

This paper displayed around assessing of thermally sprayed CrC based composite coatings and their attributes and their field of utilisation and potential use for piston ring coating. The coating of piston rings is taken for instance of the tribological framework. It is as yet a subject of the modern portrayal of mechanical segments in contact in regard to coating examination. This paper will review various parameters effecting the coating, feedstock powder, process parameters and various other parameters and their effect on the coating performance and properties.

2. Effect of feedstock powder

Coating execution is firmly subject to the surface microstructure, which like this is dependent on the attributes of the beginning powder from which the coating is framed

and the spray procedure parameters employed. The work is done by [40]to look at the impact of powder qualities on the microstructure of sprayed coatings and in this way on rough wear conduct of the coatings. Three sorts of CrxCy–NiCr powders given various assembling strategies, specifically mixed, sintered and pulverised, and composite powders, were contemplated. The work has tried to comprehend the microstructures of the coatings on the qualities of the powders from which they were sprayed and to clarify the wear conduct of the coatings in regards to their microstructures. The wear resistance of the blended coating is altogether lower than that of the other two coatings because of its low volume division of carbide and coarse level of inhomogeneity, both of which result straight forwardly from the blended way of the powder feedstock. The other two coatings, where the carbide and matrix phase are intimately bonded in the powder feedstock, demonstrate significantly higher wear resistance. The sintered and crushed coating exhibits the greatest degree of microstructural reaction during spraying; however, it possesses a high hardness and wear resistance. The composite coating, which shows much less decomposition during spraying, has a lower hardness than the sintered and crushed coating but its wear resistance is up to 22% better than that of the sintered and crushedcoating. This may be due to the short of a reacted matrix and the oxidation associated with the sintered and crushedcoating. Another significant commercial advantage of the composite powder is its markedly higher deposition efficiency than that of the other powders [41]. A commercial Ni- Cr-Si-B powder (EWAC 1004) was modified by adding 10 and 20 wt.% CrC particles. These modified powders were used for development of coatings. The addition of CrC particles in a commercial hard facing EWAC 1004 powder successfully controlled the structure and properties. An increase in the microhardness of the coating was found with the addition of CrC particles although heat treatment of the coating deteriorated the hardness. The addition of CrC particles lowers the wear rate of modified powder coating three to eightfold compare to that of unmodified powder coatings.

HVOF sprayed coatings exhibited more microstructural alterations than HVAF sprayed ones when compared feedstock powders (Fe-31Cr-12Ni-3.6B-0.6 C), (Fe-31 Cr-12Ni-2Mo-3.6B- 0.6 C) in wt% onto low carbon steel. Precipitate-free particles with high aspect ratio were deposited due to the higher particle temperature during HVOF spraying (~1800°C), compared to the HVAF process (~1400°C), which led to carbide/boride dissolution [42]. Based on SEM analysis the porosity level of the cross-section of NC WC, Ni/MC coating onto St37 steel substrates (ASTM A283 grade B) was measured to be less than 1% (~0.8%), (~0.3%) respectively. It is investigated that the coatings comprise of ultrafine and nanosized WC particles, with the majority of particle size in the range of ~50~100nm, within the Co(Ni) matrix and no any significant pieces of evidence of decarburization was observed in the microstructure [43].

In the study [44] the powders obtained by agglomeration and sintering have an improved flowability due to round shaped powder particles with uniform microstructure and prevalent equilibrium phases. Decarburisation and oxidation are reduced in the HVAF coating due to lower process temperatures which are confirmed by the XRD pattern confirming the presence of WC, Cr3C2 and Ni phases in both coatings, but W2 C and Cr2O3 have been identified in the HVOF. There is approximately no difference between the HVAF sprayed coating and the feedstock powder which means that no major transformation took place during the spraying process.

Chromium carbide-based materials are employed to produce a hard coating for elevated temperature wear application including sliding, fretting, abrasion, and erosion. It is usually considered that the wear resistance of cermet coatings is predominately subjective by their microstructures, e.g., carbide particle size carbide content and carbide distribution within splats, and porosity, etc., the results of investigation done on Cr3C2–NiCr coating were characterised shown in research [45]. The abrasive wear of the HVOF Cr3C2–NiCr coating is mainly related to its microstructure parameters particularly the carbide particle size, and the bonding of carbide particles to the NiCr matrix, carbide content. It was confirmed that the wear loss of the coating was absolutely about the square root of the average visible diameter of carbide particle size and was reversely about the carbide volume fraction. A relatively finer carbide size and higher volume fraction in the even dispersed state may lead to an improved bonding of carbide particles to the matrix phase and consequently an improved wear performance. The precipitation of Cr23C6phase may modify the wear resistance by improving the resilience of the matrix to gouging.

High energy milling of Cr3 C2 – 25(Ni20Cr) powders in gaseous nitrogen and liquid hexane reduced the average particle size. The decrease in particle size with enlarging in milling time was more pronounced in nitrogen compared with that in liquid hexane. The microstructure of HVOF sprayed coatings of Cr3C2-25(Ni20Cr) prepared with milled powders was uniform compared with those made with as-received powders. The hardness and fracture robustness of the nanostructured Cr3C2-25(Ni20Cr) coatings were higher compared with coatings prepared with the as-received powders [46].

The as-sprayed coating (agglomerated and sintered Cr3C2-25(Ni20Cr)) feedstock powder on grid blasted 1.4828 steel substrate yields porosity value below 1% as determined from carefully polished cross-sections using optical microscopy and digital image analysis [47]. The SEM images from the cross-section of ST-Glazing are demonstrated that a highly denseglazing layer with negligible porosity of ~0.3% and a thickness of 208 ± 32 μm was developed onto the 316 L stainless steel disk substrate of commercial Stellite 6 (C-1.15 W-6.50Si-1.15Ni<3.0Cr25.75Fe<3.0Co-bal. in wt% & size 15–45 μm) using ST-HVOF (as-sprayed) coating. Besides, the glazed layer was well bonded to the underlying coating, and no significant evidence of pores or cracks can be found at their interface [48].

FeBCr alloy coating is less porous below 1% (ASTM B276) than the FeSiNiCr coating alloy on gray cast iron with particle size 15–50 μm. The grains of FeBCr coating are found to the much finer size than FeSiNiCr coating. The microstructure of surface and cross-section is alike and is replete with uniformly distributed fine grains like powder deposit with prominent broad boundaries, suggesting a consistent coating through HVOF process [49].Some commercially available feedstock powder like WC-Co, WC-CoCr, WC-12 Co, WC-(W,Cr), Cr3C2–NiCr, TiC-Based Materialsareused for coat the substrate materials. Very high spraying speed of the HVOF spraying and the powder particles staying shortly in flame stream using Cr3C2–NiCr based coating, the decarburisation may not be occurred [50].

The Cr3C2-NiCr alloy coating on steel substrates by HVOF coating has been reported that the fully amorphous ribbons formed can produce good mechanical properties and wear resistance at ambient condition. This paper presents work undertaken in the production of Cr3C2-NiCr alloy powders that were designed for use in thermal spraying as shown in Figure 1

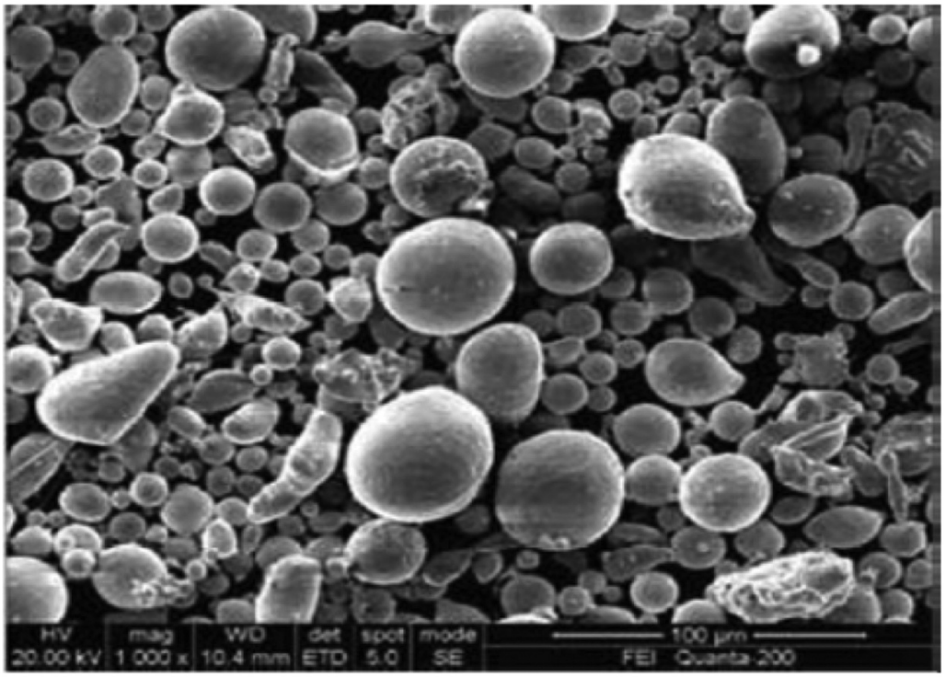


Figure 1. Cr3C2-NiCr powder image by SEM [78].

HVOF thermal spraying with oxygen and liquid petroleum gas as the fuel gas have been successful to deposit WC-CrC-Ni alloy coatings on boiler tube steels [51]. However, the quality of coating is closely related to the thermal spraying conditions and feedstock powder size [52]. The level of decarburisation of Cr3C2 has also been found to play a significant role about both hardness and wear properties of the coating since a high hardness is required to improve the coating wear performance [53]. Moreover, also the substrate temperature was carefully controlled to avoid overheating of the substrate by using compressive air cooling arrangement and by adopting an inter-pass delay during spray deposition [54].

Therefore, individualised process optimisation is necessary for different feedstocks due to the extrinsic feedstock characteristics such as density, morphology, heat transfer properties [55]. For better understanding, the flame and particle behaviour in an HVOF thermal spray process the investigation for gas temperature, gas velocity, gas mach number, gas pressure, gas component were analysed. Usually, high hardness corresponds to high abrasive wear and corrosion resistance as well as compact microstructure (low porosity) for liquid fuel HVOF sprayed WC-based coating deposited using WC_12 Co powder [56]. A promising oxidation/corrosion/erosion resistance was achieved by composite coating like HP2 [Cr3C2 – NiCr +25%(WC-Co)] deposited on T22 boiler steel using HVOF spray process [57]. The reduced toughness of the coating surface could be reasonably due to a combination of factors: the environmental interaction, microstructural instability (leading to hardness increase) and the reduction of compressive residual stresses [51]. HVOF Cr2 C3-NiCr Coating produced by using high kinetic; thermal spray process can be considered a very promising solution for improving fatigue performance of structural steel components in several industrial applications.

3. Coating parameters and porosity

The current section gives us a clear understanding of the coating process optimisation by Taguchi technique which is one of the optimisation and well recognised methods for the proper reappearance of experiments concerned only with the effects of design parameters. As per the suggested automobile components applications, the commercially available WC–CrC–Ni powder (TAFA 1356VM) is coated on the substrate IN 718 by JK3500 HVOF equipment. There are mainly three levels of four parameters hydrogen flow rate, oxygen flow rate, powder feed rate and spray distance are considered. As per the concern of HVOF WC–CrC–Ni is obtained by optimising hardness (38 FMR oxygen flow rate, 53 FMR hydrogen flow rate, 25 g/min powder feed rate and 7 in. spray distance).The sequence of the process parameters is oxygen flow rate is followed by hydrogen flow rate, feed rate and spray distance. It is also observed in the literature that the hardness of the coating is also decreasing with the increasing porosity. It can be seen that the influence of each parameter for hardness is different and oxygen flow rate affects hardness mostly when changing the levels [58].

Some of the researchers also clearly reported the porosity level with the hardness. The flow rates of oxygen (500 L/min) and the career gas methane (160 L/min) with corresponding stoichiometry (1.56) the porosity level was decreased to 1.3% and the hardness increased up to 1286 HV [52]. The Ni-P coating on ST 37 steel disk achieves extremely low porosity percentage of 0.3%, while higher porosity level of 1.6% and 1.3% was measured in the case of conventional WC-12Co and WC-17Co coatings [53].

Optical microscope image of the coating cross-section confirms that the thickness of SHS9172+ Co coating onto 304 stainless steel substrate is thicker than that of SHS9172 coating onto a substrate, under the same coating parameters for HVOF process. It suggests that adding Co to the Fe-based alloy coating can improve the deposition rate. The minimum average value (ASTM E2109-01) of porosity (~0.8%) for SHS9172+ Co is obtained after annealing up to 900oC for 3 h [59]. HVOF sprayed carbide coating on stainless steel substrate posses, low porosity less than 1% for WC-17Co (0.81%), WC-10Co-4 Cr (0.35%), WC-12Co (0.31%), and Cr3C2 – NiCr (0.65%) due to high impact velocity [55]. WC- 10Co-4 Cr coatings on naval brass using HVOF spray parameters such as with an oxygen flow rate of 253 l pm, an LPG flow rate of 61lpm, a spray distance of 227 mm, and a powder feed rate of 35 g/min, yield minimum porosity (1.38 vol.%) [56]. Porosity morphology has revealed by SEM as shown in Figure 2.

4. Effect of temperature & oxidation

Authors showed 88.5% increase in oxidation resistance of Ni/mc WC relative to mc-WC coating, and 89.3% improvement in oxidation resistance of Ni/nc-WC with respect to nc-WC coating after 3 h at 800°C. The SEM observation of Ni/mc-WC and Ni/nc-WC oxidised at 800°C revealed very dense and adhesive oxide layers with thickness of ~30–-40 μm, whereas highly porous oxide layers with of thickness of 210 and 300 μm were produced on the surface of mc-WC and nc-WC coatings. The kinetics analysis showed that oxidation of mc-WC and ncWC coatings obeys the linear law, respectively, with apparent activation energies of 90.4 and 78.9 kJ/mol [57].

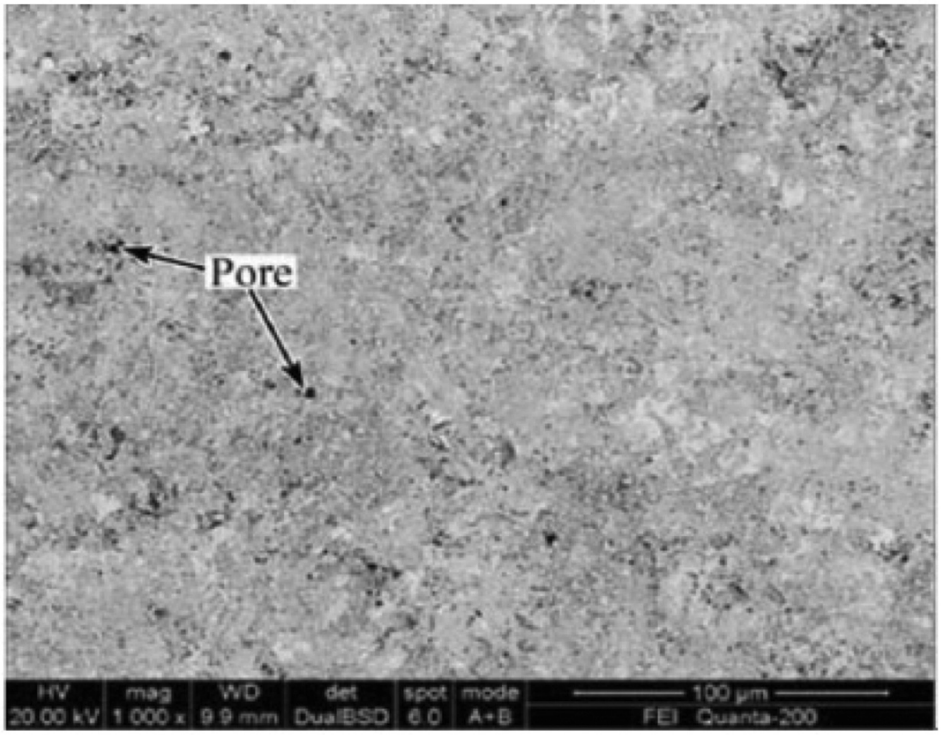


Figure 2. Porosity measurement by using SEM [78].

The Cr3-25(Ni-20Cr) and Ni-20Cr coatings on AISI 310 L sheet were more oxidation resistant compared with WC-20Cr-7Ni coatings, especially at temperatures >800°C. This is due to the higher rate of oxidation of WC in the latter coating. Besides, the oxidation rate of WC also increases forming WO3, which promotes cracking and subsequent spalling of the chromium oxide [60]. The HVOF sprayed NiCrAl coating was found to be effective in imparting hot corrosion resistance to Superfer 800 in the coal-fired boiler environment as compared to that of Ni-5Al coating. Due to the formation of a thin layer of oxides of chromium and nickel in the uppermost part of the scale as revealed by the cross-sectional X-ray mapping analysis [61].

The thick protective oxide scale developed on the surface which attributed to a superior oxidation resistance at 800°C of 25% (Cr3C2-25(Ni20Cr)) + 75% NiCrAlY on Ti-31 alloy substrate. The upper most layer of the oxide scale mainly consisted of a continuous film of Cr2O3 which have minimal reaction to air condition [62]. The coating microstructure of the ASTM A182 F51 duplex stainless steel (nominally Fe-25%,Cr-5%Ni-3%Mo-0.14 N alloy with less than 0.03%C) is characterised by a low porosity as well homogeneous carbide distribution and no carbide coarsening was observed after ageing up to 400°C [49].

5. Effect of hardness

[44] The microhardness values of the HVAF coating were 932 ± 47 HV0.3 and the HVOF coating showed somewhat higher hardness values, respectively, 1045 ± 65 HV0.3. The reason for increased hardness in the case of the HVOF coating is the presence of hard but brittle phases which formed during spraying due to the decarburisation process. The influence of abrasive media on the grinded coatings (Figure 3) proved that both coatings

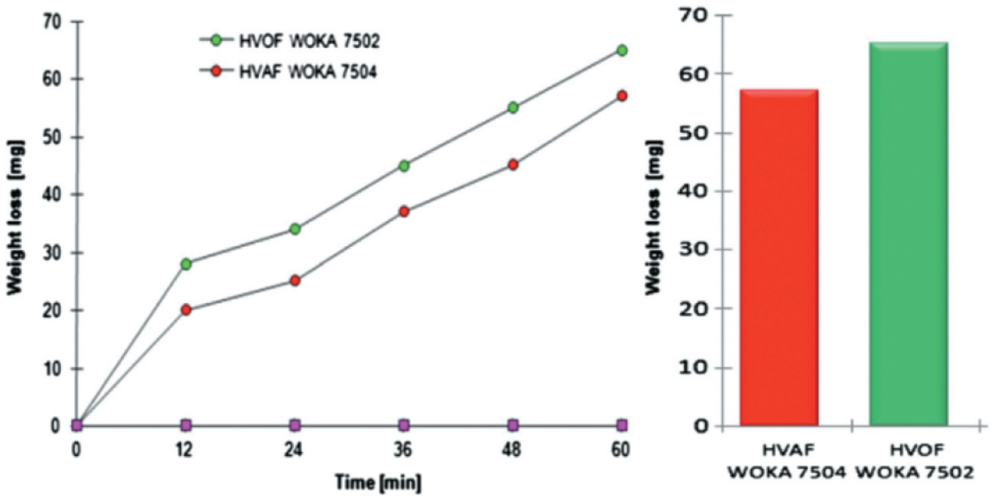


Figure 3. Abrasion wear results for the sprayed coating [44].

have good wear resistance with similar results, but the HVAF coating showed somewhat lower wear loss.

A slight strengthening effect was detected even for the F51 duplex stainless steel. In this case, the hardness rises of about 2% and 11% compared to the steel in the as received condition (227 ± 7 HV), respectively, after isothermal treatments of 100 h at 350°C and 100 h at 400°C [63]. A maximum microhardness of 1498 HV0.3 was measured for WC-F coating, whereas WC-C and WC- N coatings revealed lower values of 1305 and 1254 HV0.3 [64]. The iron-based alloy coatings have shown maximal hardness in the range of 800 to 900 HV. Furthermore, the microhardness of the coatings varies with the thickness of the coating- grey cast iron substrate interface [65].

The significant improvement in microhardness (726 HV0.1) TiO2 coated on mild steel is due to the phase uniformity of the coating, i.e., tetragonal hard rutile phase, which is dominating the amorphous anatase phase [66]. The great significance of stabilised hardness about 920 HV0.1, as it enhances the wear resistance capabilities of the FeBCr coating compared to the wear that can achieve in a system of gradient hardness. The greater hardness of the coating was attributed to solid-solution strengthening by the B and the Cr in Fe phase [67]. However, FeBCr (980 HV) coating exhibited superior hardness (6.5% higher) compared to FeSiNiCr (920 HV) coating as shown in Figure 4. Due to the low porosity accompanied by very fine grained structure [68].

In before the coating deposition of submicron-sized (~300 nm), WC particles were mechanically blended with agglomerated and sintered WC-12 wt%Co powder on 304 stainless steel plates. XRD spectra suggest WC particle leads to increase in microhardness (18.39% increases), from 1153 HV for the coating without of WC addition to 1365 HV for the coating with five wt% WC addition [70]. For lower indentation loads (below 10 N), the indentation mark was small, close to the thickness of the layer. So, the micro-hardness was mainly affected by the intrinsic hardness of TiN coating and elastic modulus of crystals inside the splats [71]. The contribution of high hardness (>1000 HV0.1) and high toughness (>80MPa) of HVOF sprayed carbide coating on stainless steel substrate, caused probably by less plastic deformation and fatigue due to the repeated action of the abrasive particles followed by the undermining of the carbide particles resulting in their eventual pullout [72].

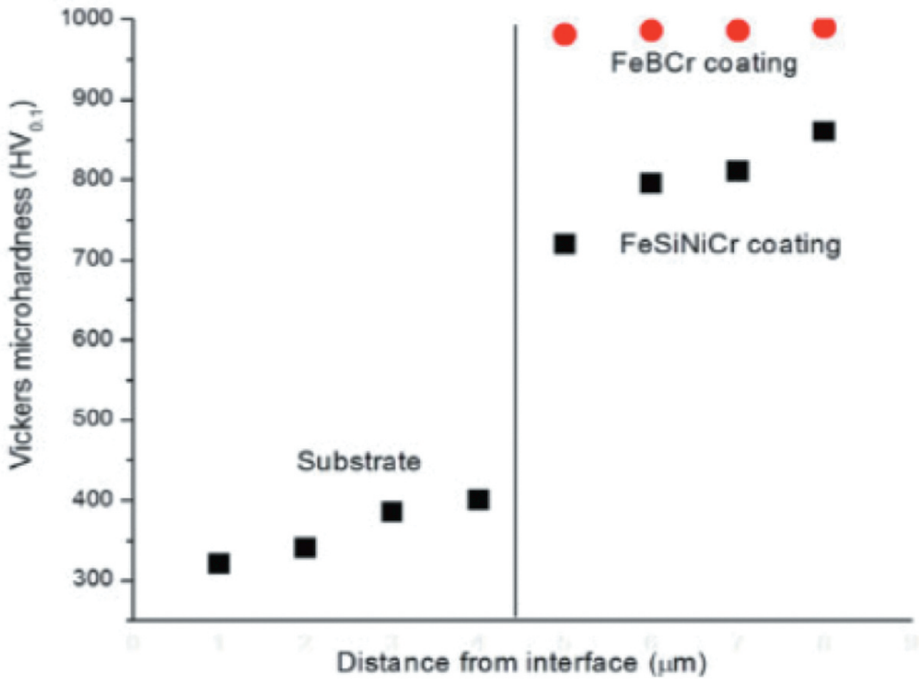


Figure 4. Microhardness profile of FeSiNiCr based alloy coating along cross-section [69].

[46] The hardness of thermal sprayed coatings is affected by the technique as well as characteristics of the feedstock. In the HVOF method, the particle velocity is high and thermal energy low, and this leads to high hardness. It has been reported that materials which are nanostructured often display higher hardness compared to similar conventional materials, although lower than that predicted values using the classical Hall-Petch equation. In this investigation, also the hardness of nanostructured Cr3 C2-25(20NiCr) coating was higher than that of the coating prepared with as-received powders. The higher hardness values of the nanostructured coatings are considered to be due to the homogeneous microstructure caused by the mechanical milling process to obtain feedstock and the intrinsically high hardness of the nanostructured material.

6. Effect of coefficient of friction

[73]Three different PVD deposited coatings were deposited CrN, (Ti, Al)N and CrC/C coatings. The wear of the coatings was found to be dependent on the substrate material as well as coating parameters such as hardness, microstructure, etc., the results showed that the CrC/C DLC-based coating showed significant lower wear and friction coefficient on eCrN and (Ti, Al)N coatings. Substrate hardness has a strong impact on the wear of the PVD coatings and consequently on the friction characteristics and galling resistance of the coating/substrate composite.

High-velocity oxy-fuel (HVOF) thermal spraying coating has been used widely in defence, power plants, and aerospace. Recently this coating technique is appraised as a favourable candidate for the replacement of the traditional electroplated hard chrome (EHC) plating which pollutes the environment and causes lung cancer by toxic hexavalent Cr6+ [58]. The HVOF sprayed CrC-based composite coatings were investigated coating processes optimised by Taguchi's method the friction and wear behaviours are analysed for the WC–CrC–Ni

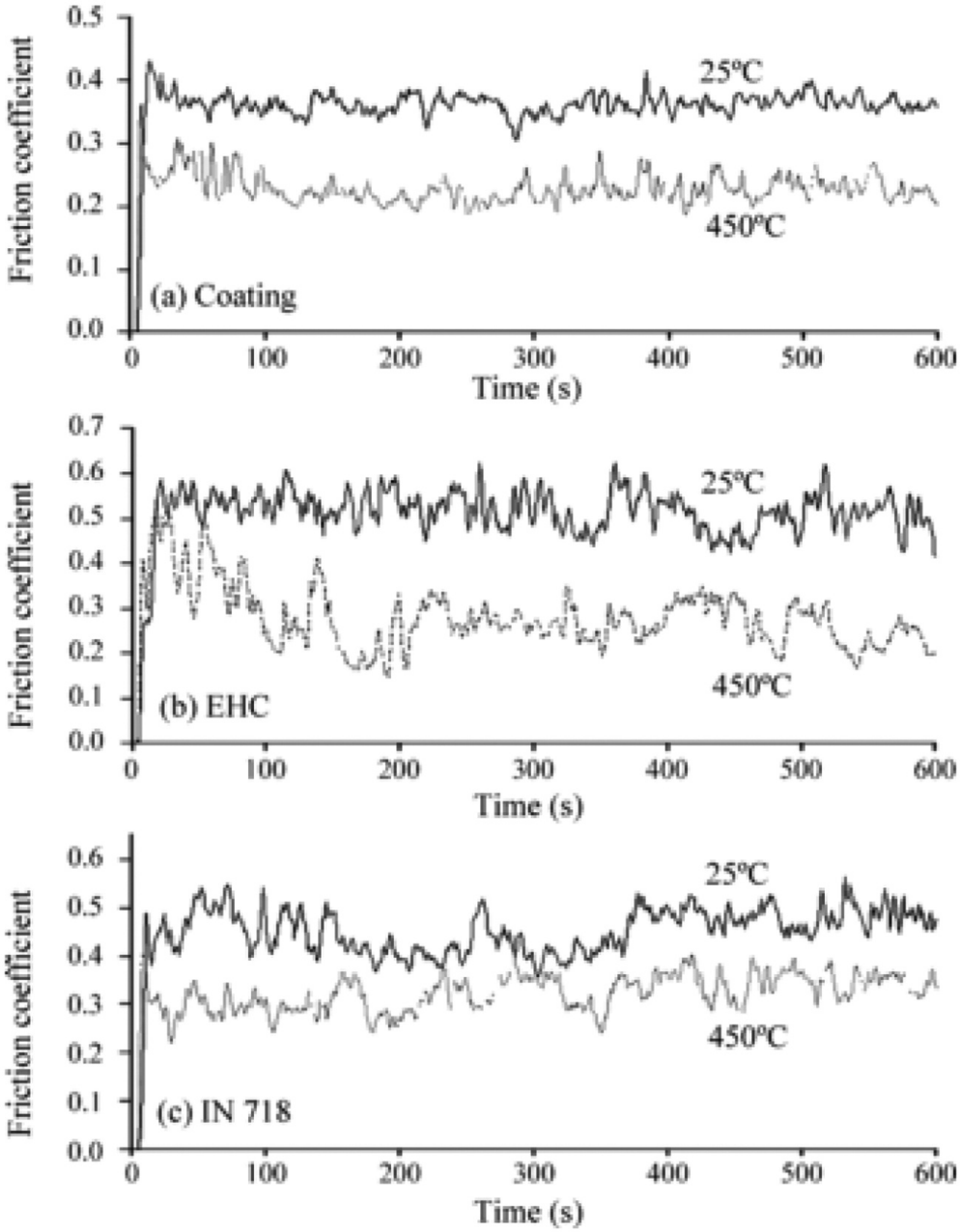


Figure 5. Friction coefficients vs. time at 25 and 450°C [58].

coatings, EHC and the substrate Inconel 718 (IN 718) both at 25 and 450◦C. The effect of temperature on friction and wear behaviours were evaluated. TheOCP coating showed significant lower Cof and wear then in718 and electrolytic hard chrome (EHC) plating Fig.5shows the Effect of COF on in718, EHCand WC-CrC-Ni.

7. Effect of wear mechanism

[74] The progressive deterioration of metallic surfaces due to various types of wear (abrasive, erosive, adhesive, corrosive and chemical wear) in various industries (coal and hydrothermal power plants, cement, automotive, chemical and cement industry) some statistical methods have recently been implemented in wear studies. These methods

share the advantage of facilitating research into the effects of different factors and their interactions (combined effect), by limiting the number of tests. Hence, in this study, an attempt has been made to study the independent as well as the combined effect of the factors using fractional factorial design (Response Surface Methodology) based on the experimental data obtained an abrasive wear model was developed to correlate the abrasive wear of the coatings regarding applied factors and their interactions. The validity of the abrasive wear model was evaluated under different abrasive wear conditions by comparing the experimental and modelled results. The hardness increases with the increase in chromium carbide concentration. The maximum hardness was obtained with 20 wt% chromium carbide. The boost in hardness is due to the formation of new phases and intermetallic compounds. Response Surface Methodology (RSM) with fractional factorial design approach is an excellent tool, which can be successfully used to develop an empirical equation for the prediction and understanding of wear behaviour of coatings in terms of individual factors (C, L, A, S and T) as well as in terms of the combined effects (CL, CT, LA, LS and AS) of various factors. The load and sliding distance have a more severe effect on the abrasive wear of the coating as compared to abrasive size. Interaction effects of various factors on abrasive wear are almost of same order less than their main factor effects. The interaction effect of abrasive size-sliding distance (AS) is considerably higher than load-abrasive size (LA). Increasing (%) CrC concentration; reducing load, abrasive size and sliding distance minimise the abrasive wear significantly. The increase in chromium carbide concentration increases the abrasive wear resistance of the coatings. Abrasive wear rate of 20 wt% chromium carbide coating is lower as compared to 0 wt% chromium carbide coatings.

An agglomerated and sintered WC-12Co powder exhibits different wear behaviours and is improves wear resistance with the uniform and homogeneous distribution of metal binder and carbides [75]. Whereas this paper a laminar splat – like microstructure was found in the as-sprayed coating, where the WC-Co phase was nearly uniformly distributed into the Cr3C2 – NiCr matrix. The composite carbide coating Cr3C2 – NiCr on ductile iron

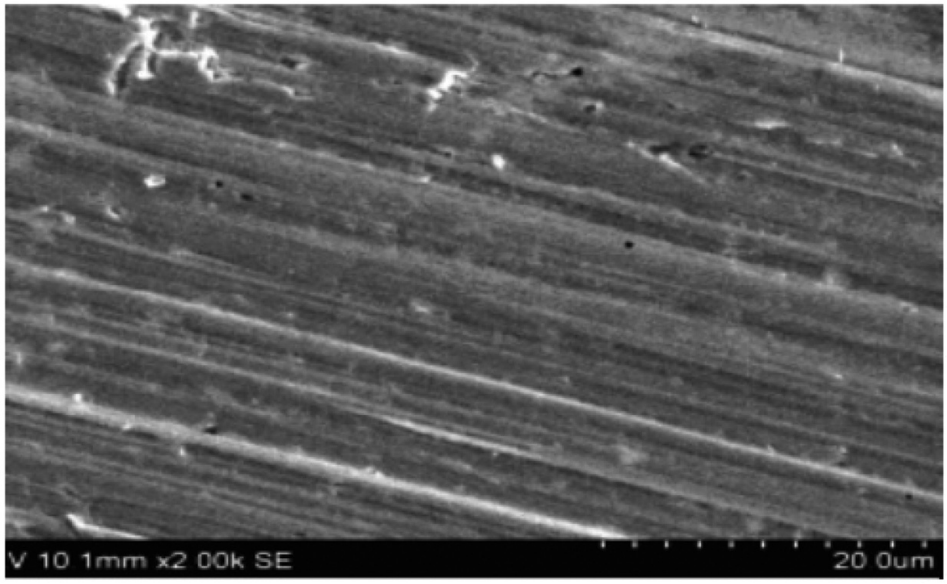


Figure 6. Wear groove of Fe-based alloy coating [76].

has good wear resistance associated with the effect of the plasticity of the coating addition of soft, metallic particles to the base ceramic powder. In adding together, better wear properties are obtained as the decarburisation degree decreases. The SEM image showed Fig.6 two body grooving morphology on Fe-based coating materials.

It evident for the resistance against low-stress abrasive and high-stress abrasive wear of coatings heat treated for 8 h at 800°C was reduced by roughly 10% and 30%, respectively, compared with the as-sprayed state. As a result of their tougher binder matrix, the wear tracks of the aged coatings are free of micro-cracks and thus exhibit less wear was presented [69]. Regarding the practical application of Cr3C2-NiCr coatings for high-temperature wear resistance, which the formation of a very dense coating with high inter-splat adhesion is of greater importance than the formation of a coating with minimal carbide dissolution, as long as it does not occur at the expense of excessive carbon loss. It appears that in-flight carbide dissolution can be tolerated in coatings for use at high temperature because of precipitation of the carbide grains. The significance of this effect increases with increasing temperature, particularly above 700°C [77]. The coefficient of friction profile of Cr3C2-NiCr coating at a different elevated temperature as shown in Figure 7.

Despite the high coefficient of friction, HVOF Fe(Mo) coating exhibited the lowest sliding wear rate accompanied by the lowest counterpart wear [79]. Under the dry sliding condition, HVOF sprayed FeSiNiCr alloy coating exhibit higher weight loss in this test duration at room temperature. The greater weight loss of as-sprayed coating was associated with removal of slurry particles [78].

A major role of wear takes place in testing that indicates the diamond abrasive particles of gray cast iron. An irrespective of change in the characteristic behavior of the subsurface at higher loads. Thus, above the 0.05 N load, the coating performs with consistently less wear resistance [80]. The wear rate was unaffected by sliding speed, while the properties of stainless steel substrate with WC-(W,Cr)2C-Ni coating have some influence only at the

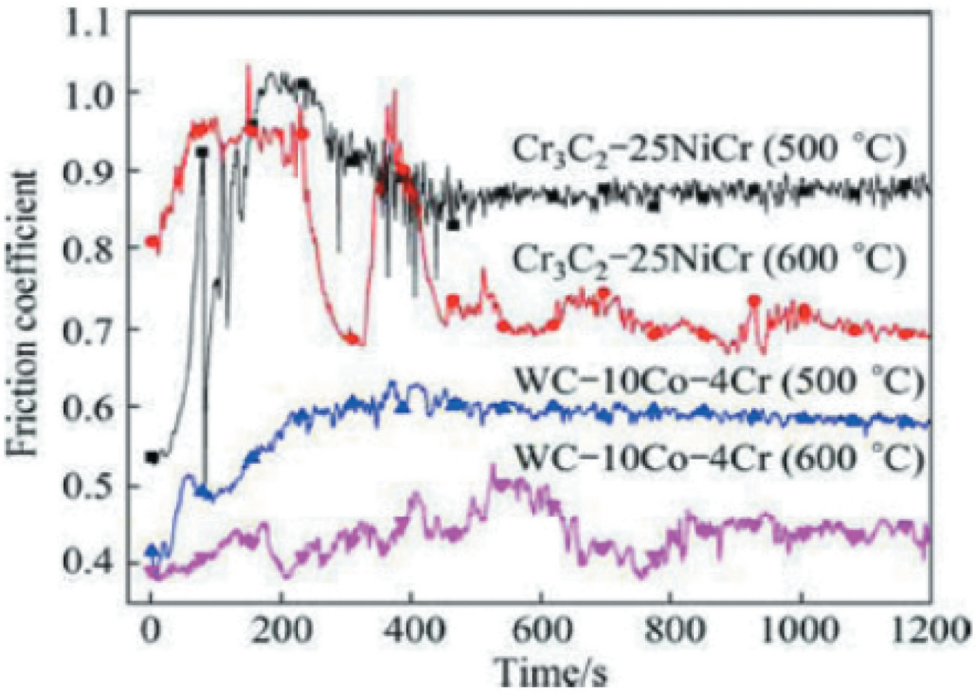


Figure 7. Coefficient of friction profile of Cr3C2-NiCr coating at elevated temperatures [78].

highest test temperature of 750°C. Wear rate values increase gradually from ~10-7 mm3/(Nm) at room temperature to 1 × 10–5 mm^3/(Nm) at 750°C [81].

F-Cr-Ni-Si-b-C alloy coatings, pure and strengthened with 20 wt% and 40 wt% WC-12 wt% Co, were deposited onto carbon steel substrates by HVOF thermal spray process. The sliding wear performance of the Fe-based coating is almost always superior to that of the HVOF sprayed [82]. HVOF sprayed NiMoAl-Cr3C2-Ag composite coating that was composed of γ-Ni, orthorhombic Cr3C2 and Ag phases exhibit a highly dense structure and well bonds with the stainless steel substrate (ASTM C633-79). The addition of Cr3C2 effectively increases the wear resistance of the coating while silver plays a major role in reducing friction coefficient [83].

The wear rate of Ni-P modified coating on ST 37 steel disk was ~3.2 × 10–4 mg/m indicating approximately 68% and 72% improvement in wear resistance, on the conventional WC-12Co and WC-17Co coatings [84]. Based on dry sand testing, the best anti-wear properties are obtained when cobalt is added to HVOF SHS9172 coatings and annealed at a high temperature of 700°C. Firstly, the wear volume loss for both SHS9172 and SHS9172+ Co coatings decrease at less than 500°C. It suggests that there are precipitates of α-Fe in amorphous/disordered phase. Secondly, the increasing trend of the wear volume loss probably is related to the B2Fe15Si3 metastable phase precipitate at more than 500°C. The wear loss was due to more carbides, borides, and silicides (i.e., Fe3Si) [85].

[58]the oxygen content of wear trace at 450°C has increased by 1.96%, 42.27% and about 11 times compared to those at 25°C for WC–CrC–Ni coatings, EHC and IN 718, respectively. It illustrates that the effect of temperature on oxidation for EHC and IN 718 is much stronger than that for the WC–CrC–Ni coating since they are reactive metal and alloy. Wear trace of the coating is shallow compared with EHC and IN 718 both at 25°C and 450°C, which indicates the excellent anti-wear performance of HVOF WC–CrC–Ni coating. Therefore the HVOF WC–CrC–Ni coating is very protective for alloy surface.

8. Characterisation of coating

This section of the research deals with the study and characterise of the coating by scanning electron microscopy (SEM) and energy dispersive spectroscopy (EDS). The study also includes that possibility of higher weight due to pores, voids, splat boundaries present in the coating microstructure [86]. Especially in the boiler environment, coal-fired plant even after1000 h of exposure, the prevention of corrosion was by preventing the penetration of oxygen into the substrate. Due to the formation of oxides on the surface and the splat boundaries was due to penetration of oxygen develop a weight gain [87].

To avoid the defects, change of properties and behaviour of hard metal, the coating technique is recommended. Accordingly, there was a development in spray coating process, i.e., HVOF method. It is reported that the HVOF process exhibits the lowest porosity (Figure 8) due to the high impact velocity compared to other spraying techniques such as electric arc, flame, plasma spraying, and detonation gun [88]. To reduce the extent of decomposition, proper optimisation of the HVOF flame jet is recommended. Multilayer coatings show better anti-wear properties as compared to monolayer ones. The effect of the architecture of the multilayer coating on these properties is still investigated. The change of the adhesion and wear was observed for multilayer coatings with a different thickness ratio of the CrCN and CrN layers [89].

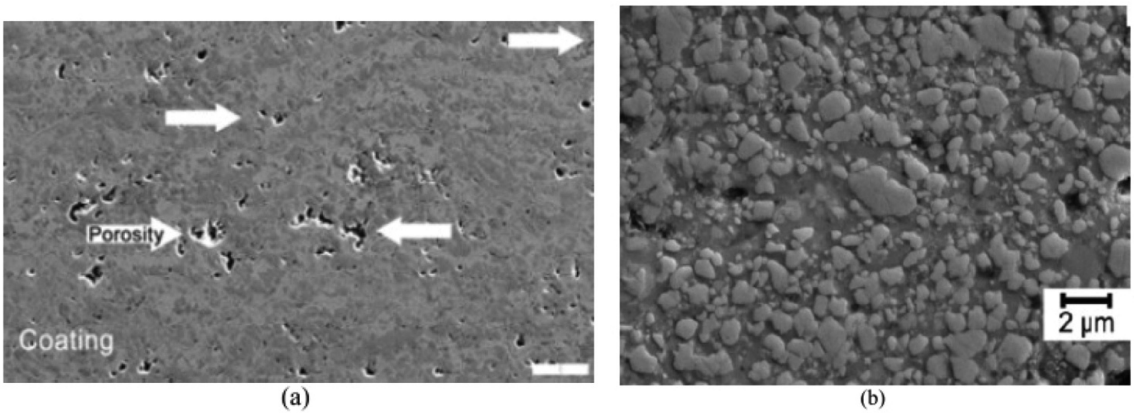


Figure 8. Surface morphology of NiCr coating [60].

8.1. Characterisation of CrC-NiCr

Thermal sprayed Cr3 C 2–NiCr coatings are being used to improve the wear resistance and decrease the friction coefficient between various sliding components in automotive applications. On the contrary, these coatings can be used as a replacement of the electroplated hard chromium coatings. The mechanical and wear behaviour of HVOF Cr 3C2 – NiCr coatings, obtained from sintered nano-composite Cr3C2 – NiCr powders, to improve the lifetime of coated materials [90].

The coatings studied in this work were Cr2 C3 75%+NiCr20 25% weight and Cr2 C3 60% +NiCr20 40%. weight, deposited on a steel substrate with a thickness of approximately 30 mm, using the HVOF thermal spray process. The nanostructured CrC–NiCr coatings present a 'bimodal structure', formed by CrC particles with a size greater than 1 mm, together with sub-micron carbides in a Ni–Cr matrix. The nano-CrC75–NiCr25 coatings present a lower adherence between splats, than the Nano-CrC60–NiCr40 coating, which appears to be more dense and homogeneous.volvesthe reduction of the superficial roughness. This fact that could reduce the grinding and blasting operations, necessary to obtain a very smooth surface, finish of the coated components and decreases the manufacturing time to produce cheaper products (the hardness of the CrC–NiCr coatings increases the cost of grinding operations) In the tribological tests, carried out under dry conditions, the specific wears rate results and the friction coefficient of nanostructured CrC60–NiCr40 coatings are lower than nanostructured CrC75–NiCr25 coating. This fact could be related to the different wear mechanism, probably as consequence of the bad adherence, between splats in nanostructured CrC75–NiCr25 coating that promotes the fast detachment of the coating at the pressures generated in dry contact.

The decomposition process is limited, for higher velocity flame and lower flame temperature [91]., as well as a good candidate for hard chrome replacement on sealing and bearing surfaces in fluid power components. The Fe-based alloy powder coatings will meet the thermal stability criteria for automotive applications. The coating is stable up to 900°C, without any phase transformations. TGA-DTA curves show that negligible (0.2%) weight loss could be due to the formation of a few oxides. So, it was clear that no significant weight loss in coated substrate [92]. Whereas WC-10Co-4Cr coating has a more dense structure which forms W2C phase and Cr7, Cr23C6 phase form in the Cr3C2-25NiCr coating mainly due to the decarburisation during spraying agreed with XRD results [93]

Hard coatings are also used to prevent seizure. At 600°C some coatings like CrC (chromium carbide) and CrN (chromium nitride) offer high wear, abrasion, and oxidation resistant layers.

Their development led to the replacement of less efficient titanium-based ceramic coatings (TiC or TiN) and carcinogenic electroplated hard chromium coating. CrC and CrN behaviour is similar (low friction coefficient, high hardness, high toughness, high corrosion resistance and good wear resistance and under extreme temperature conditions) are discussed in scientific literature [73,94,95]. [96] Tribological response of four different configurations was used to an uncoated reference contact, a bonded coating (industrial reference) and a WS2 ion implantation layer deposited either on raw Inconel 718 substrate or top of a CrC sublayer. The results show that at ambient temperature, the bonded coating shows good durability associated with a low, stable friction (0.25). Both WS2 and WS2 + CrC will not resist for more than a few thousand cycles on top of showing high COF before failing (above 0.6). At high temperature, however (600°C), the bonded coating COF increases to 0.6 while WS2 and WS2 + CrC COF falls under 0.4. Improved high-temperature performances were linked to the generation of a cohesive glaze layer on the CrC layer. A similar phenomenon is suspected for the much thinner WS2 layer. The CrC sublayer increases the load bearing ability and the lifespan of the component.

8.2. Phase analysis of CrC-NiCr

The addition of a thin immediate layer of NiCrAl coating could also be effective to enhance the bonding strength. The FeSiNiCr coatings on grey cast iron exhibit excellent ability to resist localised corrosion when treated with 20 wt% H2SO4, due to the presence of high chromium content, it was suitable for automotive application [97]. A NiCrBSiWFeCoC alloy coating with a thickness of 200 μm was prepared onto AISI 1045 steel substrate by using HVOF thermal spraying process an alternative for corrosion application due to the

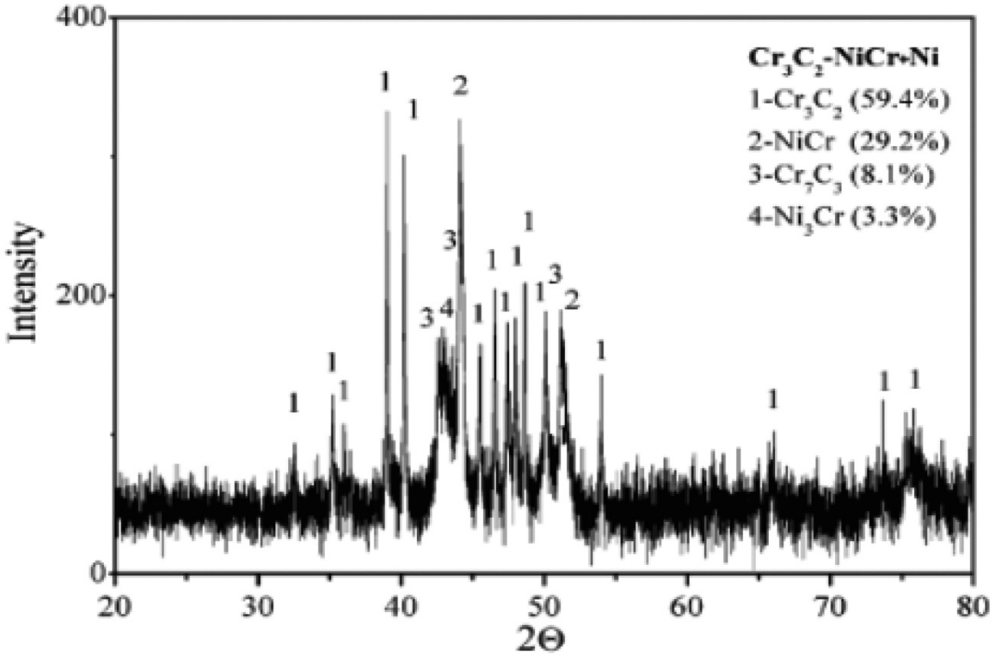


Figure 9. XRD analysis of Cr3C2-NiCr powder coating [61].

presence of amorphous phase and low porosity [85]. TiN-matrix composite coating on medium carbon steel, it was found that its corrosion resistance was strictly dependent on coating porosity [98]. NiMoAl-Cr3C2-Ag composite coating that performs an extremely dense structure, high bonding strength as well as excellent lubrication function and wears resistance has a potential to be used as a high-performance protective coating in harsh environments such as elevated temperature, high speed and heavy load [99].

The HVOF deposited coating has a thickness of about 270 μm, and the HVAF deposited coating 160 μm. The HVAF coating presents a lower degree of porosity due to lower process temperature and high particle velocities. According to the porosity quantification, the HVOF coating has presented somewhat a higher degree of porosity with 1.63% in comparison with the HVAF coating with 0.92% porosity. This might be caused by the different particle size distributions of the powder and different spraying processes. The EDX spectra and high magnification SEM micrographs in literature proved that the HVAF process did not affect the phase compositions found in the feedstock powder. The presence of WC phases and Cr2C3 phases can be noticed in both coatings [100].

The added Fe-based alloy powders transformed into the amorphous alloy. Thus, the composite coating comprises WC and Fe-based amorphous phases. To protect the nuclear fuel cladding, the dense Cr3C2 – NiCr coating was fabricated using HVOF spraying on the Zr-2.5Nb alloy. The as-prepared were examined in high-temperature environment and autoclave. During exposure to high-temperature air and steam, the Cr3C2 – NiCr coating could provide protection for the zirconium alloy coating. After heat treatment, the coating was still well bonded on the substrate [101]. The minimum carbon loss is obtained for WC-F (WC-FeCrAl) (16%), while WC-C (WC-Co) and WC-N (WC-NiMoCrFeCo) coatings undergo more severe decarburisation of 30% and 36%, respectively. Besides, no crystalline peaks related to the binder phases can be observed in the XRD patterns of the coatings [102]. WC particles with bright contrast and the metallic binder phases with the dark contrast of two phases with different contrast are visible on the cross-sectional BSE images of powder particle at high magnification. It is also obvious that the carbide grains are surrounded by metallic binders [103].

TiNi powder particles deposited on a Q235 steel substrate by HVOF technique. The TEM image shows the deformation of the particle, and the substrate increased, the adiabatic shear instability happened. As a result, the grain was heavily deformed. In this condition of high strain rate, subgrains with large angular grain boundary were developed, and local grain boundary migration driven by differences in dislocation energy occurred at this stage [104]. The addition of Al2O3-13% TiO2 nanoparticles on mild steel as the reinforcing materials significantly improved the quality of the HVOF coatings and the overall coating microstructure was found to be homogeneous and uniform long with the lamella boundaries [105].

9. Effect of corrosion behaviour

The composite coating (Wt% of WC-10Co-4cr) on the mild steel achieved a higher corrosion resistance (pitting and corrosion cracks) in simulated seawater than WC coating. However, there was no any change/improvement in the wear resistance of composite and WC coating [106]. EPMA and EDAX analysis along the cross-section of the corroded coating reveal that only the external surface of the coating has been

oxidised to form a thick oxide scale and below this, coating remains unoxidised. HVOF system with a gas-fuelled gun is used for coat TiO2 on the mild steel substrate, which attributes an excellent corrosion resistance in salt spray tests compared to the uncoated mild steel substrate [107]. The corrosion resistance of AISI 321 MASS has revealed an interesting trend with varying grain sizes. However, since grain refinement has also enhanced its mechanical properties, the extent to which prior deformation and the rate at which the deformation occurred has affected corrosion resistance will also be examined [108].

TiO2 coating offers better resistance to the impinging slurry particles avoiding loss of material [109]. The composite coating (Fe48Mo14 C-r15Y2C15B6 Fe-based amorphous powder) on 316 L stainless steel exhibited, higher corrosion weight loss than the monolithic amorphous coating [110]. Based on electrochemical testing, the best anti-corrosion properties in 5% NaCl solution are obtained when cobalt was added to HVOF SHS9172 as-sprayed coatings [61].

10. Effect of substrate material

The investigation shows that the hardness of substrate material significantly influences the coating characteristics. The wear of the diverse coatings was found to be reliant on the type of substrate material, T33, and T44, respectively. In common the coatings deposited on the softer T33 substrate are worn by a brittle like fracture mechanism while constant wear wears the coatings deposited on the harder T44 substrate, the COF values were also lower on T44.the observed result is due to the lower load-bearing capacity of the softer T33 substrate [73]. In general three different shapes were chosen in order to analyse how the substrate curvature affects the coating microstructure. A flat, square sample, a cylindrical sample and a rod sample can be easily sprayed. The substrates were prepared using three different surface treatments. The substrates denominated 'as-machined' were machined in order to have a roughness close to Ra = 0.1 μm. The two other series of substrates were grit-blasted with two aluminium oxide (Al2O3) grit sizes (around 100 μm and 200 μm) producing two different surface roughness conditions: Ra ≈ 2 μm, denominated as 'fine'; and Ra ≈ 5 μm, and as 'coarse', respectively. The surface roughness of the coating was measured with the help of 3D surface profilometer [111,112].

11. Conclusions

From the condensed results and talk, thermally sprayed cermet coatings have shown up as a practical answer for a wide scope of wear resistance applications to enhance the administration life of machine parts. Tungsten carbide and chromium carbide-based coatings are more than once utilised for a large number of the application in a gas turbine, steam turbine, and air motor to improvise the imperviousness to sliding, grating and erosive wear. The HVOF and plasma sprayed coatings that have been the subject of the work like microstructure, wear, COF and corrosion conduct. Thermally sprayed HVOF and plasma sprayed coatings are the most widely recognised being the subject of corrosion, and sliding wear examinations while HVOF coatings have likewise gotten some consideration for sliding wear–corrosion applications. Most analyses test existing

coatings that have been produced for either wear resistance or corrosion resistance. The audit has recognised the important cooperations amongst wear, and corrosion components with COF that happen on an extensive variety of coating frameworks alongside somewhere in the range of two bodies and three bodies scraped area procedure that intends to educate coating decision and foresee execution. It was inferred that the CrC-based amalgam coating could be delivered with different blend is one of them being NiCr which created significant outcomes contrasted with other feedstock materials. The expansion of WC with CrC likewise decreased COF of the coating when contrasted with the substrate and electroplated hard chromium. The other alloyed coatings likewise demonstrated comparable outcomes. It was additionally watched that the coating could be utilised to supplant cancer-causing electroplated hard chromium coating on cylinder rings. CrC alloyed covered structure accomplishes more exceptional wear resistance and exceptional corrosion resistance and lower estimations of COF utilised with an impact of both formless/or crystalline microstructure initiated in the covered materials. The Fine CrC–NiCr coatings exhibit better execution than hard chrome concerning mechanical and tribological properties. For future fruitful utilisations of CrC–NiCr HVOF coatings as a contrasting option to hard chromium, many variables like wear resistance, friction coefficient, costs and fundamental issues must be considered on the whole.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

[1] Holmberg K, Andersson P, Erdemir A. Global energy consumption due to friction in passenger cars. Tribol Int. 2012;47:221–234.
[2] Jost HP. James Clayton lecture energy saving through tribology : a techno-economic study. Proceedings - Institution of Mechanical Engineers. 1981;195: 151–173. https://doi.org/10.1243/PIME_PROC_1981_195_016_02.
[3] Jost HP. Tribology - Origin and future. Wear. 1990;136(1):1–17.
[4] Tung SC, McMillan ML. Automotive tribology overview of current advances and challenges for the future. Tribol Int. 2004;37(7):517–536.
[5] Priest M, Taylor CM. Automobile engine tribology—approaching the surface. Wear. 2000;241(2):193–203.
[6] Andersson BS. Paper XVIII (iii) Company Perspectives in Vehicle Tribology - Volvo. Tribol Ser. 1991;18(C):503–506.
[7] K BS, Jadhav NP. Friction losses between piston ring-liner assembly of internal combustion engine: a review Online]. Available. Int J Sci Res Publ. 2013;3(6):2250–3153. : www.ijsrp.org.
[8] Arumugam S, Sriram G. Preliminary study of nano- and microscale TiO 2 additives on tribological behavior of chemically modified rapeseed oil. Tribol Trans. 2013;56(5):797–805.
[9] Gara L, Zou Q. Friction and wear characteristics of water-based ZnO and Al 2 O 3 nanofluids. Tribol Trans. 2012;55(May):345–350.
[10] Davis FA, Eyre TS. The effect of a friction modifier on piston ring and cylinder bore friction and wear. Tribol Int. 1990;23(3):163–171.
[11] Andersson P, Tamminen J, Sandström CE. Piston ring tribology: A literature survey. In: *VTT Tied. - Valt. Tek. Tutkimusk.*. Vol. 2178. 2002. p. 3–105.

[12] Shi PZ, Wang J, Tian CX, *et al*. Structure, mechanical and tribological properties of CrN thick coatings deposited by circular combined tubular arc ion plating. Surf. Coatings Technol. 2013;228(SUPPL.1):534–537.
[13] Legg KO, Graham M, Chang P, et al. The replacement of electroplating. Surf Coat Technol. 1996;81(1):99–105.
[14] Fedrizzi L, Rossi S, Cristel R, Bonora, PL. Corrosion and wear behaviour of HVOF cermet coatings used to replace hard chromium. Electrochimica Acta. 2004;49:2803–2814. https://doi.org/10.1016/j.electacta.2004.01.043
[15] Rastegar F, Richardson DE. Alternative to chrome: HVOF cermet coatings for high horse power diesel engines. Surf Coat Technol. 1997;90(1–2):156–163.
[16] von Burg R, Liu D. Chromium and hexavalent chromium. J Appl Toxicol. 1993;13(3):225–230.
[17] Vetter J, Barbezat G, Crummenauer J, et al. Surface treatment selections for automotive applications. Surf Coat Technol. 2005;200(5–6):1962–1968.
[18] Barbezat G. Advanced thermal spray technology and coating for lightweight engine blocks for the automotive industry. Surf Coat Technol. 2005;200(5–6):1990–1993.
[19] Kumar V, Kandasubramanian B. Processing and design methodologies for advanced and novel thermal barrier coatings for engineering applications. Particuology. 2016;27:1–28.
[20] Houdková Š, Zahálka F, Kašparová M, et al. Comparative study of thermally sprayed coatings under different types of wear conditions for hard chromium replacement. Tribol Lett. 2011;43(2):139–154.
[21] Limpichaipanit A, Wirojanupatump S, Jiansirisomboon S. Fabrication and properties of thermal sprayed stainless steel-based nanocomposite coatings. Surf Coat Technol. 2015;272:96–101.
[22] Picas JA, Forn A, Matthäus G. HVOF coatings as an alternative to hard chrome for pistons and valves. Wear. 2006;261(5–6):477–484.
[23] Lin J, Wei R, Bitsis DC, et al. Development and evaluation of low friction TiSiCN nanocomposite coatings for piston ring applications. Surf Coat Technol. 2016;298:121–131.
[24] Zabala B, Igartua A, Fernández X, *et al*. Friction and wear of a piston ring/cylinder liner at the top dead centre: experimental study and modelling. Tribol. Int. 2017;106(October 2016):23–33.
[25] Li Q, Jiang FQ, Leng YX, et al. Microstructure and tribological properties of Ti(Cr)SiCN coating deposited by plasma enhanced magnetron sputtering. Vacuum. 2013;89(1):168–173.
[26] Friedrich C, Berg G, Broszeit E, et al. PVD Cr N coatings for tribological application on piston rings x. Surf. Coatings Technol. 1997;97(1–3):661–668.
[27] Lorenzo-Martin C, Ajayi O, Erdemir A, et al. Effect of microstructure and thickness on the friction and wear behavior of CrN coatings. Wear. 2013;302(1–2):963–971.
[28] Broszeit E, Friedrich C, Berg G. Deposition, properties and applications of PVD Cr N coatings. Surf. Coatings Technol. 1999;115(1):9–16.
[29] Hazar H. Characterization of MoN coatings for pistons in a diesel engine. Mater Des. 2010;31(1):624–627.
[30] Ahn HS, Lyo IW, Lim DS. Influence of molybdenum composition in chromium oxide-based coatings on their tribological behavior. Surf Coat Technol. 2000;133–134:351–361.
[31] Eyre TS. Wear characteristics of metals. Tribol Int. 1976;9(5):203–212.
[32] Vetter J. 60years of DLC coatings: historical highlights and technical review of cathodic arc processes to synthesize various DLC types, and their evolution for industrial applications. Surf Coat Technol. 2014;257:213–240.
[33] Keunecke M, Bewilogua K, Becker J, et al. CrC/a-C:H coatings for highly loaded, low friction applications under formulated oil lubrication. Surf Coat Technol. 2012;207:270–278.
[34] Haque T, Morina A, Neville A, et al. Non-ferrous coating/lubricant interactions in tribological contacts: assessment of tribofilms. Tribol Int. 2007;40(10–12 SPEC. ISS.):1603–1612.
[35] Hill F, "ICEF2012-92139," 1–6, 2016.
[36] Piston-ring coating. Seal Technol. 2011;2011(7):13. https://doi.org/10.1016/s1350-4789(11)70270-7.

[37] Aouadi SM, Luster B, Kohli P, et al. Progress in the development of adaptive nitride-based coatings for high temperature tribological applications. Surf Coat Technol. 2009;204 (6–7):962–968.
[38] Singh SK, Chattopadhyaya S, Pramanik A, et al. Effect of lubrication on the wear behaviour of CrN coating deposited by PVD process. Int J Surf Sci Eng. 2019;13(1):60–78.
[39] Singh SK, Chattopadhyaya S, Pramanik A, et al. Wear of chromium nitride coating under high loads and speeds. Int. J. Surf. Sci. Eng. 2019;4:263. DOI:10.1504/IJSURFSE.2019.103929
[40] Wirojanupatump S, Shipway PH, Mccartney DG. The influence of HVOF powder feedstock characteristics on the abrasive wear behaviour of Cr x C y – niCr coatings. Wear. 2001;249 (9):829–837. https://doi.org/10.1016/S0043-1648(01)00821-3
[41] Harsha S, Dwivedi DK, Agarwal A. Influence of CrC addition in Ni-Cr-Si-B flame sprayed coatings on microstructure, microhardness and wear behaviour. Int J Adv Manuf Technol. 2008;38(1–2):93–101.
[42] Fu D, Xiong H, Wang Q. Microstructure evolution and its effect on the wear performance of HVOF-sprayed conventional WC-co coating. J Mater Eng Perform. 2016;25(10):4352–4358.
[43] Singh H, Kaur M, Prakash S. High-temperature exposure studies of HVOF-Sprayed Cr3C2-25(NiCr)/(WC-Co) coating. J Therm Spray Technol. 2016;25(6):1192–1207.
[44] Hulka I, Şerban VA, Secoşan I, et al. Wear properties of CrC-37WC-18M coatings deposited by HVOF and HVAF spraying processes. Surf Coat Technol. 2012;210:15–20.
[45] Ji GC, Li CJ, Wang YY, et al. Microstructural characterization and abrasive wear performance of HVOF sprayed Cr3C2-NiCr coating. Surf Coat Technol. 2006;200(24):6749–6757.
[46] Alvares C, *et al*. Microstructure and mechanical properties of thermal sprayed nanostructured Cr 3 C 2 -Ni20Cr coatings 2. Mater. Methods. 2008;11(2):137–143.
[47] Pan J, Hu S, Yang L, et al. Numerical analysis of flame and particle behavior in an HVOF thermal spray process. Mater Des. 2016;96:370–376.
[48] Guo SF, Pan FS, Zhang HJ, et al. Fe-based amorphous coating for corrosion protection of magnesium alloy. Mater. Des. 2016;108:624–631.
[49] Chen W, Mao T, Zhang B, et al. Designs and preparation of advanced HVOF-PVD duplex coating by combination of HVOF and arc ion plating. Surf Coat Technol. 2016;304:125–133.
[50] Jafari M, Enayati MH, Salehi M, et al. High temperature oxidation behavior of micro/nanostructured WC-Co coatings deposited from Ni-coated powders using high velocity oxygen fuel spraying. Surf Coat Technol. 2016;302:426–437.
[51] Gariboldi E, Rovatti L, Lecis N, et al. Tribological and mechanical behaviour of Cr3C2?NiCr thermally sprayed coatings after prolonged aging. Surf Coat Technol. 2016;305:83–92.
[52] Wang G, Xing C, Tao F, et al. Enhancement in the corrosion resistance of WC coatings by adding a Fe-based alloy in simulated seawater. Surf Coat Technol. 2016;305:62–66.
[53] Varis T, Suhonen T, Calonius O, et al. Optimization of HVOF Cr3C2[Formula presented] coating for increased fatigue performance. Surf Coat Technol. 2016;305:123–131.
[54] Jian Y, Huang Z, Xing J, et al. Effects of Mn addition on the two-body abrasive wear behavior of Fe-3.0??wt% B alloy. Tribol Int. 2016;103:243–251.
[55] Kunioshi CT, Correa OV, Ramanathan LV. High temperature oxidation and erosion–oxidation behaviour of HVOF sprayed Ni–20Cr, WC–20Cr–7Ni and Cr3C2–Ni–20Cr coatings. Surf Eng. 2006;22(2):121–127.
[56] Mahesh RA, Jayaganthan R, Prakash S. Evaluation of hot corrosion behaviour of HVOF sprayed NiCrAl coating on superalloys at 900 °C. Mater Chem Phys. 2008;111 (2–3):524–533.
[57] Somasundaram B, Somasunderbgmailcom E, Kadoli R, Ramesh, MR, Ramesh, CS. High temperature corrosion behaviour of HVOF sprayed WC-CrC-Ni coatings. International Journal of Surface Science and Engineering. 2016;10(4):400–413. https://doi.org/10.1504/IJSURFSE.2016.077542
[58] Fang W, Cho TY, Yoon JH, *et al*. Processing optimization, surface properties and wear behavior of HVOF spraying WC-CrC-Ni coating. J. Mater. Process. Technol. 2009;209 (7):3561–3567.

[59] Wu YY, Tsui WC, Liu TC. Experimental analysis of tribological properties of lubricating oils with nanoparticle additives. Wear. 2007;262(7–8):819–825.
[60] Ang ASM, Howse H, Wade SA, Berndt, CC. Manufacturing of nickel based cermet coatings by HVOF process. Surface Engineering. 2015;32(10): 713–724. https://doi.org/10.1179/1743294415Y.0000000031 .
[61] Ksiazek M, Boron L, Radecka M, et al. Mechanical and tribological properties of HVOF-Sprayed (Cr3C2-NiCr+Ni) composite coating on ductile cast iron. J Mater Eng Perform. 2016;25(8):3185–3193.
[62] Bolelli G, Milanti A, Lusvarghi L, et al. Wear and impact behaviour of high velocity air-fuel sprayed Fe-Cr-Ni-B-C alloy coatings. Tribol Int. 2016;95:372–390.
[63] Hussain T, Simms NJ, Nicholls JR, et al. Fireside corrosion degradation of HVOF thermal sprayed FeCrAl coating at 700-800??C. Surf Coat Technol. 2015;268:165–172.
[64] Janka L, Norpoth J, Trache R, et al. Influence of heat treatment on the abrasive wear resistance of a Cr3C2NiCr coating deposited by an ethene-fuelled HVOF spray process. Surf Coat Technol. 2016;291:444–451.
[65] Jin D, Yang F, Zou Z, *et al*. A study of the zirconium alloy protection by Cr3C2-NiCr coating for nuclear reactor application. Surf. Coatings Technol. 2016;287:55–60.
[66] Matthews S, Berger LM. Long-term compositional/microstructural development of Cr3C2-NiCr coatings at 500 ??c, 700 ??c and 900 ??c. Int J Refract Met Hard Mater. 2016;59:1–18.
[67] Milanti A, Matikainen V, Bolelli G, et al. Microstructure and sliding wear behavior of Fe-based coatings manufactured with HVOF and HVAF Thermal Spray Processes. J Therm Spray Technol. 2016;1–16. DOI:10.1007/s11666-016-0410-z
[68] Nahvi SM, Jafari M. Microstructural and mechanical properties of advanced HVOF-sprayed WC-based cermet coatings. Surf Coat Technol. 2016;286:95–102.
[69] Priyan MS, Hariharan P. Wear and corrosion resistance of Fe based coatings by HVOF sprayed on gray cast-iron for automotive application. Tribol Ind. 2014;36(4):394–405.
[70] Priyan MS, Hariharan P. Micro abrasive wear resistance of FeSiNiCr alloy coating deposited by HVOF coating. Int J Surf Sci Eng. 2013;7(3):250–268.
[71] Ramesh CS, Sekhar N, Keshavamurthy R, et al. A study on slurry erosion and corrosion behaviour of HVOF sprayed titania coatings. Int J Surf Sci Eng. 2015;9(1):55–68.
[72] Shoja-Razavi R. Laser surface treatment of stellite 6 coating deposited by HVOF on 316L alloy. J Mater Eng Perform. 2016;25(7):2583–2595.
[73] Eriksson J, Olsson M. Tribological testing of commercial CrN, (Ti,Al)N and CrC/C PVD coatings - Evaluation of galling and wear characteristics against different high strength steels. Surf Coat Technol. 2011;205(16):4045–4051.
[74] Sharma S. Parametric study of abrasive wear of Co-CrC based flame sprayed coatings by response surface methodology. Tribol Int. 2014;75:39–50.
[75] Priyan MS, Hariharan P. Abrasive wear modes in ball - Cratering test conducted on Fe 73 Si 15 Ni 10 Cr 2 alloy deposited specimen. Tribol Int. 2014;36(1):97–106.
[76] Priyan MS, Hariharan P, Azad A, et al. Microstructure and wear analysis of FeBCr based coating deposited by HVOF method. Tribol Ind. 2014;36(2):134–143.
[77] Wesmann JAR, Espallargas N. Effect of atmosphere, temperature and carbide size on the sliding friction of self-mated HVOF WC-CoCr contacts. Tribol Int. 2016;101:301–313.
[78] Zhang WC, LIU L-B, ZHANG M-T, *et al*. Comparison between WC-10Co-4Cr and Cr3C2-25NiCr coatings sprayed on H13 steel by HVOF. Trans Nonferrous Met Soc China. 2015;25(11):3700–3707. (*English Ed*.). .
[79] Yuan J, Ma C, Yang S, et al. Improving the wear resistance of HVOF sprayed WC-Co coatings by adding submicron-sized WC particles at the splats' interfaces. Surf Coat Technol. 2016;285:17–23.
[80] Guo X, Planche MP, Chen J, et al. Relationships between in-flight particle characteristics and properties of HVOF sprayed WC-CoCr coatings. J Mater Process Technol. 2014;214(2):456–461.

[81] Bolelli G, Berger LM, Bonetti M, et al. Comparative study of the dry sliding wear behaviour of HVOF-sprayed WC-(W,Cr)2C-Ni and WC-CoCr hard metal coatings. Wear. 2014;309 (1–2):96–111.
[82] Bolelli G, Börner T, Milanti A, *et al*. Tribological behavior of HVOF- and HVAF-sprayed composite coatings based on Fe-Alloy + WC-12% Co. Surf. Coatings Technol. 2014;248:104–112.
[83] Chen J, Zhao X, Zhou H, et al. Microstructure and tribological property of HVOF-sprayed adaptive NiMoAl-Cr3C2-Ag composite coating from 20??C to 800??C. Surf Coat Technol. 2014;258:1183–1190.
[84] Hong S, Wu Y, Li G, et al. Microstructural characteristics of high-velocity oxygen-fuel (HVOF) sprayed nickel-based alloy coating. J Alloys Compd. 2013;581:398–403.
[85] Jafari M, Enayati MH, Salehi M, et al. Improvement in tribological properties of HVOF sprayed WC-Co coatings using electroless Ni-P coated feedstock powders. Surf Coat Technol. 2013;235:310–317.
[86] Jegadeeswaran N, Ramesh MR, Bhat KU. Oxidation resistance {HVOF} sprayed coating 25% (Cr3C2-25(Ni20Cr)) + 75%NiCrAlY on titanium alloy. Procedia Mater Sci. 2014;5 (1995):11–20.
[87] Ma J, Yan DQ, Hu JW, et al. Reactive HVOF sprayed TiN-matrix composite coating and its corrosion and wear resistance properties. Trans Nonferrous Met Soc China. 2013;23 (4):1011–1018.
[88] Lin QS, Zhou KS, Deng CM, et al. Deposition behavior and microstructural development of TiNi powder particles in low temperature-HVOF spraying process. Appl Surf Sci. 2013;283:352–359.
[89] Gilewicz A, Warcholinski B. Tribological properties of CrCN/CrN multilayer coatings. Tribol Int. 2014;80:34–40.
[90] Picas JA, Forn A, Igartua A, et al. Mechanical and tribological properties of high velocity oxy-fuel thermal sprayed nanocrystalline CrC-NiCr coatings. Surf Coat Technol. 2003;-174–175(3):1095–1100.
[91] Liu L, Zhang C. Fe-based amorphous coatings: structures and properties. Thin Solid Films. 2014;561:70–86.
[92] Liu W-H, Shieu F-S, Hsiao W-T. Enhancement of wear and corrosion resistance of iron-based hard coatings deposited by high-velocity oxygen fuel (HVOF) thermal spraying. Surf Coat Technol. 2014;249:24–41.
[93] Xie M, Zhang S, Li M. Comparative investigation on HVOF sprayed carbide-based coatings. Appl Surf Sci. 2013;273:799–805.
[94] Vera EE, Vite M, Lewis R, et al. A study of the wear performance of TiN, CrN and WC/C coatings on different steel substrates. Wear. 2011;271(9–10):2116–2124.
[95] Nilsson M, Olsson M. Tribological testing of some potential PVD and CVD coatings for steel wire drawing dies. Wear. 2011;273(1):55–59.
[96] Fortes Da Cruz J, Da Silva Botelho T, Lemaire-Caron I, et al. Role of WS2, WS2+CrC and bonded coatings on damage and friction of Inconel 718 flat rough surfaces at high temperature. Tribol Int. 2016;100:430–440.
[97] Murugan K, Ragupathy A, Balasubramanian V, et al. Optimizing HVOF spray process parameters to attain minimum porosity and maximum hardness in WC-10Co-4Cr coatings. Surf Coat Technol. 2014;247:90–102.
[98] Murugan. K, Balasubramanian RA, Sridhar V. Developing empirical relationship to predict hardness in WC- 10Co-4Cr HVOF sprayed coatings. Procedia Mater Sci. 2014;5:918–927.
[99] Myalska H, Moskal G, Szymański K. Microstructure and properties of WC-Co coatings, modified by sub-microcrystalline carbides, obtained by different methods of high velocity spray processes. Surf Coat Technol. 2014;260:303–309.
[100] Bolelli G, Berger L-M, Börner T, *et al*. Sliding and abrasive wear behaviour of HVOF- and HVAF-sprayed Cr3C2-NiCr hard metal coatings. Wear. 2016;358–359:32–50.
[101] Oksa M, Varis T, Ruusuvuori K. Performance testing of iron based thermally sprayed HVOF coatings in a biomass-fired fluidised bed boiler. Surf Coat Technol. 2014;251:191–200.

[102] Oladijo OP, Venter AM, Cornish LA. Correlation between residual stress and abrasive wear of WC-17Co coatings. Int J Refract Met Hard Mater. 2014;44:68–76.
[103] Ozdemir I, Ogawa K, Sato K. Iron boron based powders sprayed by high velocity spray processes. Surf Coat Technol. 2014;240:373–379.
[104] Picas JA, Rupérez E, Punset M, et al. Influence of HVOF spraying parameters on the corrosion resistance of WC-CoCr coatings in strong acidic environment. Surf Coat Technol. 2013;225:47–57.
[105] Puchi-Cabrera ES, Staia MH, Santana YY, *et al*. Fatigue behavior of AA7075-T6 aluminum alloy coated with a WC-10Co-4Cr cermet by HVOF thermal spray. Surf. Coatings Technol. 2013;220:122–130.
[106] Rafiei M, Salehi M, Shamanian M, et al. Preparation and oxidation behavior of B4C–Ni and B4C–TiB2–TiC–Ni composite coatings produced by an HVOF process. Ceramics International. 2014;40(8, Part B):13599–13609.
[107] Sun B, Fukanuma H, Ohno N. Study on stainless steel 316L coatings sprayed by a novel high pressure HVOF. Surf Coat Technol. 2014;239:58–64.
[108] Tiamiyu AA, Eduok U, Szpunar JA, et al. Corrosion behavior of metastable AISI 321 austenitic stainless steel: investigating the effect of grain size and prior plastic deformation on its degradation pattern in saline media. Sci. Rep. 2019;9(1). DOI:10.1038/s41598-019-48594-3
[109] Venter AM, Oladijo OP, Luzin V, et al. Performance characterization of metallic substrates coated by HVOF WC-Co. Thin Solid Films. 2013;549:330–339.
[110] Verdian MM, Raeissi K, Salehi M. Fabrication and corrosion resistance of HVOF-sprayed Ni2Si intermetallic compound. Appl Surf Sci. 2013;273:426–431.
[111] Caio F, Moreau C. "Influence of substrate shape and roughness on coating microstructure. Spray," SP Coatings. 2019;9(11):746. .
[112] Singh SK, Chattopadhyaya S, Pramanik A, et al. Wear behavior of chromium nitride coating in dry condition at lower sliding velocity and load. Int J Adv Manuf Technol. 2017;1–11. DOI:10.1007/s00170-017-0796-x

Ageing Research Reviews 61 (2020) 101078

Contents lists available at ScienceDirect

Ageing Research Reviews

journal homepage: www.elsevier.com/locate/arr

Review

Ubiquitin biology in neurodegenerative disorders: From impairment to therapeutic strategies

Dhiraj Kumar, Rashmi K. Ambasta, Pravir Kumar*

Molecular Neuroscience and Functional Genomics Laboratory, Delhi Technological University (Formerly Delhi College of Engineering), Delhi 110042, India

ARTICLE INFO

Keywords:
Ubiquitin E3 ligase
Ubiquitin proteasome system
Ubiquitination
Neurodegenerative diseases
Signaling
Therapeutics

ABSTRACT

The abnormal accumulation of neurotoxic proteins is the typical hallmark of various age-related neurodegenerative disorders (NDDs), including Alzheimer's disease, Parkinson's disease, Huntington's disease, Amyotrophic lateral sclerosis and Multiple sclerosis. The anomalous proteins, such as Aβ, Tau in Alzheimer's disease and α-synuclein in Parkinson's disease, perturb the neuronal physiology and cellular homeostasis in the brain thereby affecting the millions of human lives across the globe. Here, ubiquitin proteasome system (UPS) plays a decisive role in clearing the toxic metabolites in cells, where any aberrancy is widely reported to exaggerate the neurodegenerative pathologies. In spite of well-advancement in the ubiquitination research, their molecular markers and mechanisms for target-specific protein ubiquitination and clearance remained elusive. Therefore, this review substantiates the role of UPS in the brain signaling and neuronal physiology with their mechanistic role in the NDD's specific pathogenic protein clearance. Moreover, current and future promising therapies are discussed to target UPS-mediated neurodegeneration for better public health.

1. Introduction

According to the united nation's 2019 statistical reports of world population, a steady rise from one in eleven people (9%) to one in six people (16 %) with age of over 65 years by 2050. Further, it is projected to become triple, from 143 million in 2019 to 426 million in 2050 with 80 years aged individuals (United Nations, 2019). While the aging figures imply an improved life expectancy, but neurodegenerative pathologies remained an unresolved mystery for the aged people. Here, age and sex are the two important variables affecting human health and diseases, one being the main risk factor, while other impacting the disease incidences, severity, and the treatment responses. In fact, post mortem and living brain studies have revealed neuro-epigenetic signatures of age and sex in our brain (Gilbert et al., 2019). Moreover, timely research has demonstrated numerous neurodegenerative mechanisms in the brain, including mitochondrial dysfunctions, ROS production, protein oxidation, lipid peroxidation, ER stress, nitrosative stress, protein aggregation (amyloid, tau, synuclein, Lewy bodies, and

Abbreviations: AD, Alzheimer's disease; ALS, amyotrophic lateral sclerosis; ASK1, apoptosis signal-regulating kinase 1; AMPA, α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid; AMPAR, α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor; Aβ, amyloid beta; AKT, protein kinase B; ARE, adenylate-uridylate-rich element; ATP, adenosine triphosphate; CaMKII, calcium-calmodulin-dependent protein kinase II; cAMP, cyclic adenosine monophosphate; CD28/36, cluster of differentiation 28/36; CLIPTAC, click-formed PROTAC; cPLA2, cytosolic phospholipases A2; CREB, cAMP response element-binding protein; CREM, cAMP-responsive element modulator; DNA, deoxyribonucleic acid; DUBs, deubiquitinating enzymes; ERK1/2, extracellular signal-regulated protein kinases 1 and 2; E1s, ubiquitin-activating enzymes; E2s, ubiquitin-conjugating enzymes; E3s, ubiquitin-ligating enzyme; ER, endoplasmic reticulum; ERAD, endoplasmic reticulum associated degradation; FOXO, forkhead family of transcription factors; GABA, gamma-aminobutyric acid; GluA1, glutamate ionic receptor A1; GluR, glutamate receptor; HD, Huntington's disease; HSR, heat stress response; IFN-γ, interferon gamma; JNK, c-jun N-terminal kinase; KEAP1, Kelch-like ECH-associated protein 1; LBD, Lewy body disease; MAPK, mitogen-activated protein kinase; MDM2, mouse double minute 2 homolog; mHtt, mutant huntingtin; MHC, major histocompatibility complex; MS, multiple sclerosis; NDDs, neurodegenerative disorders; NEDD8, neural precursor cell expressed, developmentally down-regulated 8; NF-kB, nuclear factor kB; NMDA, N-methyl-D-aspartate; NMDAR, N-methyl-D-aspartate receptor; Nrf2, nuclear factor erythroid 2–related factor 2; NUB1, negative regulator of ubiquitin like proteins 1; PA28, proteasome activator; PD, Parkinson's disease; PI3K, phosphoinositide 3-kinase; PKA/C, protein kinase A/C; PROTAC, proteolysis targeting chimera; PRU, pleckstrin-like receptor for ubiquitin domain; ROS, reactive oxygen species; SOD1, superoxide dismutase 1; SUMO1, small ubiquitin-like modifier 1; TCR, T-cell receptor; TDP-43, TAR DNA-binding protein 43; TUBE, tandem ubiquitin binding entity; UBA, ubiquitin associated domain; UBB^{+1}, mutant ubiquitin; USP9X, ubiquitin specific peptidase 9 X; Ubl, ubiquitin like protein; UPS, ubiquitin proteasome system

* Corresponding author at: Molecular Neuroscience and Functional Genomics Laboratory, Department of Biotechnology, Delhi Technological University (Formerly Delhi College of Engineering), Room # FW4TF3, Mechanical Engineering Building, Shahabad Daulatpur, Bawana Road, Delhi 110042, India.
E-mail address: pravirkumar@dtu.ac.in (P. Kumar).

https://doi.org/10.1016/j.arr.2020.101078
Received 6 November 2019; Received in revised form 24 March 2020; Accepted 22 April 2020
Available online 12 May 2020
1568-1637/ © 2020 Elsevier B.V. All rights reserved.

inclusion bodies), DNA damage, microglial activation, neuro inflammation, and proteasomal dysfunction. Ubiquitin proteasome system (UPS) dysfunction is central to all the neurodegeneration mechanisms, since it regulates the proteostasis in nearly all these processes (Douglas and Dillin, 2010). Interestingly, disturbed proteostasis as a consequence of aging have also reported to cause circadian and sleep disturbances in the patients and are believed to drive disease pathogenesis early in the neurodegeneration (Musiek and Holtzman, 2016). Altogether, they affect specific neurons in distinct parts of the brain, i.e. selective neuronal and regional vulnerability in different neurodegenerative disorders (NDDs), including Alzheimer's disease (AD), Parkinson's disease (PD), Huntington's disease (HD), Amyotrophic lateral sclerosis (ALS), and Multiple sclerosis (MS). In addition, their progression in the brain occurs in a stereotypical and predictable fashion, whose mechanism is still unknown (Fu et al., 2018). Additionally, these pathologies are governed by a systemic response that includes glial cells, vascular cells and peripheral cells, apart from neurons as a sophisticated network in the brain. Therefore, integrative biology of systemic proteostasis equilibrium in the human brain needs to be addressed for unraveling the potential therapeutic modalities to target the neurodegenerative pathologies.

Like every cell, neuron has an efficient system to clear toxins and to maintain cellular-integrity with help of a well-coordinated ubiquitin coding system. Here, ubiquitin coding is a type of post-transnational modification of a protein that regulates its stability, longevity, functionality and localization in a cell (Hershko, 2005). The functional UPS machinery is comprised of multiple components, including ubiquitin protein, ubiquitin activating enzyme (E1), ubiquitin conjugating enzyme (E2), ubiquitin ligating enzyme (E3), deubiquitinating enzymes (DUBs), and the proteasome. Initially, ubiquitin is activated by E1 enzyme via ATP hydrolysis, i.e. adenylation of the ubiquitin C-terminal carboxyl group and its transfer to a cysteinyl residue on E1 enzyme (Rose, 2005). It is followed by the conjugation of ubiquitin, i.e. transfer of activated ubiquitin from E1 to a cysteinyl residue of E2 enzyme. The final step of ubiquitin transfer to the substrate protein's lysine residue with isopeptide bond takes place either directly from E2 or after thioester formation of ubiquitin with the E3 ligase enzyme. These three main steps of ubiquitin-activation, conjugation, and ligation are known as Ubiquitination. The subsequent ubiquitin attachment at the different lysine site on substrate protein (multi-monoubiquitination) or on the previously attached ubiquitin (polyubiquitination) determines the cellular fate depending upon the state and conformations of polyubiquitin chains, including methionine (M1), and lysine [K6, K11, K27, K29, K33, K48, and K63; (Callis, 2014)]. These ubiquitination signals regulate diverse biological functions in neurons, such as DNA damage responses, endocytosis, neuronal-development, neural network morphogenesis, synapse remodeling, ribosome biogenesis, cell adaptation and signal transduction pathways (Kwon and Ciechanover, 2017). Interestingly, ubiquitination is a cross-road between cell survival and death due to its potent ability to clear malfunctioned or cytotoxic proteins that enable cell to overcome proteotoxic stress. Besides, cell adopts any of these mechanisms, including proteasomes and autophagy to target the ubiquitinated protein for its selective degradation. In fact, investigations on the model substrates revealed Lys48 polyubiquitination as a predominant signal for proteasomal degradation while others known to direct protein-protein interactions to form multi-protein complex and modulate the activities of their substrates. Since, Lys-48 linked poly ubiquitin chain form a closed conformation where hydrophobic patches of ubiquitin molecules sequestered at the ubiquitin-ubiquitin interfaces, thereby disallowing them to interact with other target receptors. While, Lys-63 linked polyubiquitin chain, forms an extended structure exposing their hydrophobic patches thereby allowing their interaction with the receptor proteins. In this way, Lys-48 and Lys-63 polyubiquitin chains are the predominant markers for proteasomal and autophagosomal protein clearance from the cell (Fig. 1), and affirm ubiquitination as a shared signal for protein quality control in neurons either through their targeting for diverse cellular functions or targeting them for degradation via proteasomal or autophagosomal/lysosomal pathway (Ji and Kwon, 2017).

2. How ubiquitination is crucial for neuronal signaling and brain homeostasis?

Over hundred billion interconnected wire-frame networks of neurons transmit well-coordinated electrochemical signals to regulate complex physiological functions in our brain. In general, these functions include Pre- and Post-synaptic regulation, neurotransmitter's vesicle transport and release, synaptic density and plasticity, and neuronal receptor modulations (DiAntonio et al., 2001; Lin and Man, 2013). Moreover, ubiquitination determines learning and memory functions via regulating the abundance of selected proteins involved in the synaptic plasticity (Caroni et al., 2012). Interestingly, ubiquitination is also associated with drug addiction mediated behavioral changes and neuronal adaptations via targeting drug induced proteins for proteasomal degradation (Massaly et al., 2015). Furthermore, it also controls the transcription process in neurons by regulating the degradation of transcriptional repressors like IκB and CREM (cAMP-responsive element modulator) (Liu and Chen, 2011). Alike neurons, ubiquitination also acts in glial cells like astrocytes and microglia to maintain protein homeostasis, cytokine balance and healthy environment in the brain. Therefore, it is necessary to discuss the ubiquitin system's biology in the brain for proper understanding of complex biophysiological functions to regulate brain homeostasis.

2.1. UPS mediated synaptic plasticity: A gateway for learning and memory functions in the brain

The substantial research in several model systems has forced us to believe that synapses are the principal site for information storage in the brain. However, memory consolidation relies on the strength of the synapses in different parts of the brain, including their initial storage in the hippocampus followed by their transfer to the cortex region (Abraham et al., 2019). Interestingly, UPS plays a vital role in controlling these synaptic connections during memory and learning processes of synaptic plasticity via regulating protein synthesis and degradation (Bingol and Sheng, 2011). Both short-term and long-term synaptic plasticity's regulation included the turnover of diverse protein kinases- cAMP-dependent protein kinase A (PKA), calcium-calmodulin-dependent protein kinase II (CaMKII), protein kinase C (PKC), and mitogen-activated protein kinase (MAPK). For instance, NMDA receptor's stimulation triggered CaMKII autophosphorylation and consequent redistribution of proteasomes from dendritic shaft to the spine and its proteasomal activity in hippocampal neurons (Bingol et al., 2010). Similarly, serotonin stimulated sensory neurons exhibited proteasomal degradation of inhibitory R subunit of PKA and subsequent PKA and PKC mediated degradation of CREB repressor- CREB1b to regulate long-term facilitation in *Aplysia* model (Upadhya et al., 2004). Likewise, a recent cross-talk between GluA1 phosphorylation and ubiquitination demonstrated that GluA1 ubiquitination causes enhanced protein kinase A (PKA) phosphorylation at Ser-845 (Guntupalli et al., 2017; Widagdo et al., 2017). The former study indicates the potential of ubiquitination in PKA mediated phosphorylation of AMPA receptor subunits (GluR1 and GluR4) to trigger their synaptic incorporations in neuronal membrane during long-term potentiation (Esteban et al., 2003; Zheng and Keifer, 2009). On contrary, MAPK family proteins such as ERK1/2 is responsible for gene transcription, protein translation and post-translational modification during information processing and memory formation while others, including ERK5, p38 or JNKs have a distinct role in learning and memory (Medina and Viola, 2018). Besides, proteasomal degradation of NF-kB regulators- IjB and IkB kinases regulated gene transcription and its associated formation and stability of fear memory in the amygdala (Jarome et al., 2011). Moreover, UPS

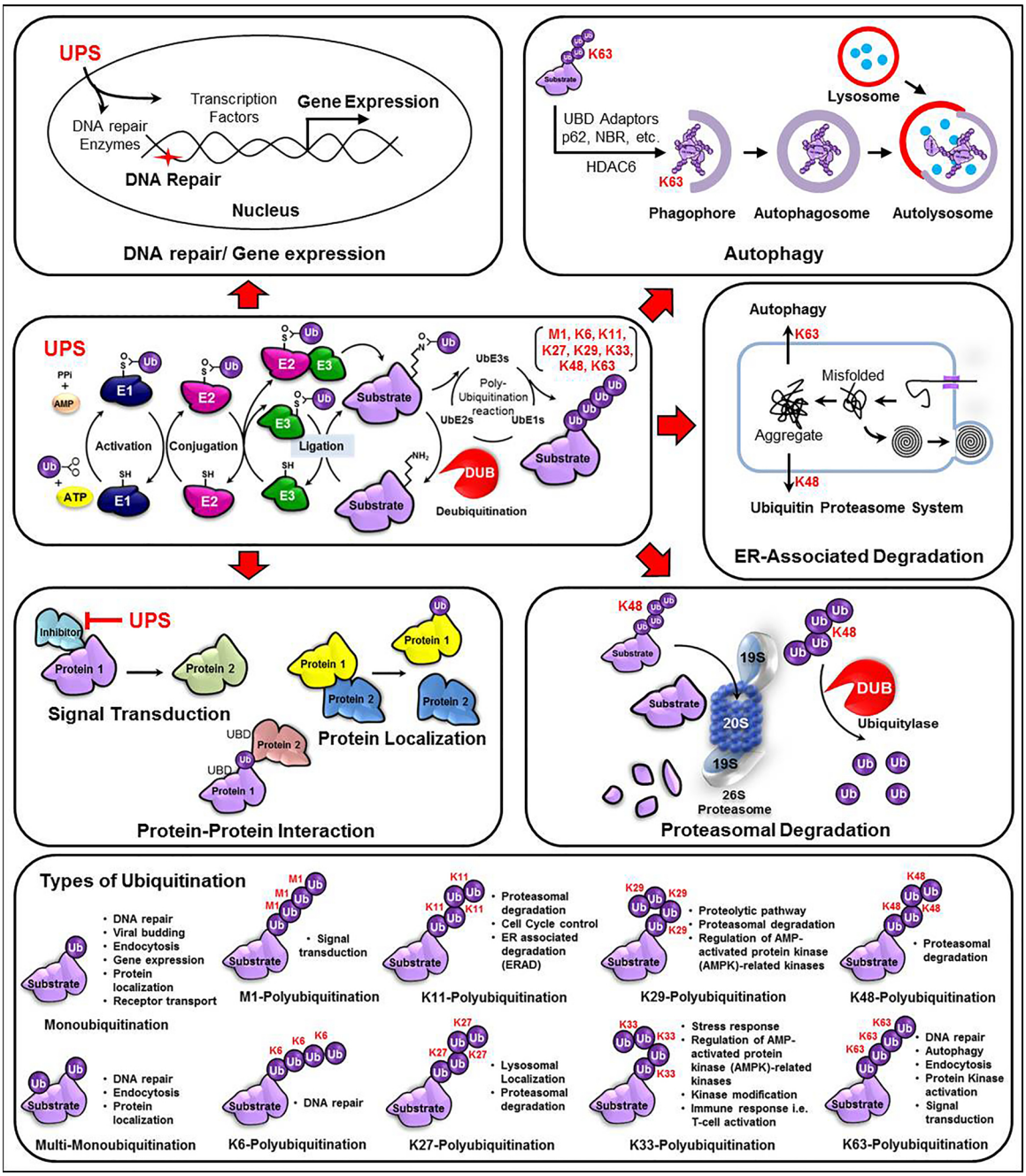


Fig. 1. The ubiquitination process and their cellular responses based on the ubiquitin chain architecture: The ubiquitination involves the complex interplay of ubiquitin activating enzymes (E1s), ubiquitin conjugating enzymes (E2s), ubiquitin ligases (E3s) and deubiquitinases (DUBs) to form a variety of ubiquitin tags, including mono-ubiquitin, multi-mono-ubiquitin and poly-ubiquitin chains (M1, K6, K11, K27, K29, K33, K48, K63) that perform a variety of functions in a cell. UPS- Ubiquitin proteasome system; Ub- Ubiquitin, UBD- Ubiquitin binding domain; p62- Nucleoprotein 62, NBR- Neighbor of BRCA1 gene; HDAC6- Histone deacetylase 6.

also regulates both pre- and post-synaptic protein densities and receptor shuttling to govern synaptic transmission activity during memory and learning process. For instance, NMDAR activity dependent proteasomal degradation of MOV10 resulted in translational activation of alpha-CaMKII, Limk1, and the depalmitoylating enzyme lysophospholipase1 (Lypla1) during synaptic plasticity (Banerjee et al., 2009). Other synaptic proteins like Dunc-13 (pre-synaptic vesicle priming protein), Syntaxin 1 (synaptic vesicle exocytosis protein), and RIM1α (pre-synaptic scaffolding protein) are proteasomally degraded by different E3 ligases namely Fbxo45, STARING (Syntaxin 1-interacting RING finger protein) and SCRAPPER respectively (Chin et al., 2002; Hsieh et al., 2018; Yao et al., 2007). Similarly, UPS also modulates structural proteins, neurotransmitter receptors and regulatory proteins in postsynaptic terminal. For instance, proteasome inhibitor MG132 prevented internalization of membrane receptors, including ionotropic GLR-1 glutamate receptors and NMDA-induced AMPA receptors (Widagdo et al., 2015). In addition, postsynaptic scaffold protein PSD-95 stabilized the surface expression of AMPA receptors, while its proteasomal degradation by Mdm2 (E3 ligase) triggered AMPA receptor internalization (Colledge et al., 2003). Similarly, NMDA receptors also retro translocated and degraded by UPS with the help of SCF-type ligase and contribute towards short-term synaptic plasticity regulation (Kato et al., 2005). Likewise, the glycine and GABA receptors also internalized and degraded upon ubiquitination in the neurons (Lin and Man, 2013). Furthermore, other post-synaptic scaffolding proteins such as GRIP1, GKAP, Shank, SPAR, and Arc have a role in synaptic strength

Table 1
Key role of E3 ligases in synaptic plasticity mediated memory and learning processes within the brain.

S.No.	E3 ligases	Role in Memory and Learning	Target Protein	Experimental Model	References
1	APC^{Cdh1}	Interacts with EphA4 receptor to mediate ephrin-induced proteasomal degradation of AMPARs	AMPAR	EphA4 knockout mice	(Fu et al., 2011)
2	CBX4	Contextual fear conditioning memory Formation	HDAC7	C57BL6J mice	(Jing et al., 2017)
3	CRL4	Target voltage- and calcium-activated BK channel for ER retention	Cereblon, BK channel	CRBN mutant mice	(Song et al., 2018)
4	Fbx2	Controls synaptic NR1and contribute to homeostatic regulation of NMDA receptor	NR1	Hippocampal neurons	(Kato et al., 2005)
5	MDM2	Glutamate receptor trafficking	PSD-95	Rat forebrain crude synaptosomes	(Colledge et al., 2003)
6	MGRN1	Potentiates the transcriptional cellular response to proteotoxic stress	–	C57BL/6J mice and Brain lysate	(Benvegnù et al., 2017a)
7	Nedd4	Role in trafficking of GluA1-containing AMPAR	GluA1, AMPAR	Nedd4(+/-) mice	(Camera et al., 2016)
8	Parkin	Role in short term memory through regulation of hippocampal synaptic plasticity	Synphilin, α-synuclein, synaptotagmin XI, Pael receptor, tubulin, p38/JTV-1, mitofusin	Parkin-deficient mice	(Rial et al., 2014)
9	RNF167	Regulates GluA2 associated synaptic AMPAR currents	GluA2, AMPAR	Sprague–Dawley rats	(Lussier et al., 2012)
10	$SCF^{\beta\text{-}TRCP}$	Regulate postsynaptic density of spine-associated Rap GTPase activating protein (SPAR) in association with Polo-like Kinase 2	SPAR	E19 Long Evans rat hippocampi	(Ang et al., 2008)
11	SCRAPPER	Role in RIM1 (Rab3-interacting molecule-1) mediated synaptic vesicle release and neurotransmission	RIM1	SCR (+/−) mice	(Yao et al., 2011)
12	TRIM3A	Role in γ-actin turnover at hippocampal synapses regulating spine densities, long-term potentiation, and short-term contextual fear memory consolidation	γ-actin	Trim3 −/− mice	(Schreiber et al., 2015)
13	TRIM67	Role in appropriate brain development and behavior	Netrin receptor DCC	Trim67-/- mice	(Boyer et al., 2018)
14	TRIM9	Role in neuronal morphogenesis and spatial learning and memory	SNAP25, Netrin-1	Trim9-/- mice	(Winkle et al., 2016)
15	UBE3A	Regulate postsynaptic SK2 levels for NMDA receptor activation. Also regulate glutamate receptor trafficking and CaMKII abundance	SK2, Arc	UBE3A-deficient mice	(Mabb and Ehlers, 2010; Sun et al., 2015)

regulation by their own proteasomal degradation to control the surface expression of diverse receptors, including AMPA (GluR2) and NMDA by different E3 ligases- TRIM3, SCFβ-TRCP, and Triad3A (Ferrara et al., 2019; Lin and Man, 2013; Mabb et al., 2014). There is growing list of E3 ligases regulating the synaptic proteins and receptors to modulate learning and memory process in the brain (Table 1). In summary, kinase action contributes towards memory formation by two ways, including the regulation of ion channel density/conductivity, i.e. regulation of membrane receptor trafficking and regulation of gene transcription/ local translation that governs synaptic transmission and synaptogenesis respectively.

2.2. Redox regulated UPS activity governs lipid metabolism, inflammatory responses and cellular immunity in the brain

The brain is main metabolizers of oxygen, but with relatively delicate defensive antioxidant mechanisms. Here, mitochondrial machinery metabolizes oxygen to produce ATP that yields several reactive oxygen species (ROS) including oxygen ions, free radicals, and peroxides in neurons. In fact, low level of ROS ensures the activation of various signaling pathways through transcriptional activation, while its abundance causes oxidative or nitro-oxidative stress in neurons. In fact, various studies have demonstrated the role of oxidative stress in UPS impairment thereby aggravating the pathogenesis of NDDs. For instance, High levels of many polyubiquitinated proteins belonging to protein synthesis, quality control system, energy metabolism, and cytoskeleton network are reported in the human AD brain with respect to the age-matched controls. Moreover, the identification of quality control system proteins, including PPIA, HSP-90β, and UCH-L1 in the pool signifies oxidative damage linked UPS impairment and proteotoxicity in AD (Tramutola et al., 2018). Similarly, Nitrosative stress is found to S-nitrosylate E3-ligase Parkin and consequent downregulation of E3 ligase-UPS pathway providing a link between free radical toxicity and protein accumulation in PD (Chung et al., 2004). Likewise, another study has revealed defective form of UCH-L1, a major target of oxidative damage, supporting a direct link between oxidative stress and UPS dysfunction in the pathogenesis of AD and PD (Choi et al., 2004). Further, proteomic analysis has identified other unconventional ubiquitin chain targets like Ubc6 that are critical for proteasomal function (Xu et al., 2009). In addition, they trigger severe inflammatory responses against the degenerated neurons and muscles, oxidized proteins, and lipid peroxidation, or glycated products in the brain (Popa-Wagner et al., 2013). For instance, ROS mediated phosphorylation of NF-κB inhibitor by IκB kinase (IKK) may regulate proteasomal degradation of IκB and consequent release of NF-κB (Alkalay et al., 1995; Kanarek and Ben-Neriah, 2012; Rajakumar et al., 2013). Thus, rescued NF-κB migrates into the nucleus and promotes the transcription of pro-inflammatory mediators along with c-Fos/Activator protein 1 (Collins et al., 2016). Another regulator of redox signaling; NRF2 remains bound to its adaptor KEAP1 under basal conditions and are maintained at a low level in the cytoplasm through ubiquitination. Though in the diseased state it migrates into the nucleus and binds to the ARE sequence and consequently, trigger transcriptions of multiple genes involved in the expression of antioxidants, efflux transporters, glutathione, ATP synthesis, and several proteasomal subunits, etc. (Sivandzade et al., 2019; Changjiang et al., 2018). Furthermore, oxidative stress initiates the release of pleiotropic cytokines like IFN-γ that induces the proteasomal activator PA28 (11S multimeric complex) to trigger over expression of immunoproteasomes and consequent MHC class- I mediated antigenic peptide's presentation in astrocytes and microglial cells (Jansen et al., 2014; Launay et al., 2013). Besides, oxidative stress as well linked with the PI3K/AKT pathway mediated over expression of the FOXO transcription factors that guide neuronal survival and stress responses in nervous system (Lefaki et al., 2017; Uranga et al., 2013). In addition, UPS also protects against lipid peroxidation and consequent toxin production, which may otherwise cause neural apoptosis

(Nagamine et al., 2010). For instance, a UPS component, Parkin monoubiquitinates CD36 and stabilize it at plasma membrane to facilitate the uptake of fatty acids stored in lipid droplets or oxidized by mitochondria. Moreover, under cellular stress, CD36 triggers Ca2+- and ERK1/2-mediated activation of cPLA2 to release arachidonic acid from ER and nuclear membrane phospholipids that aids in synaptic transmission and its conversion into bioactive eicosanoids by microglia to have pleiotropic effects on inflammation (Abumrad and Moore, 2011). UPS also regulates both T-cell receptor (TCR) and co-stimulatory CD28 signaling via Cbl family ubiquitin ligases and USP9X deubiquitinase mediated regulation of pro-inflammatory cytokines and their receptor's expression to direct effective immunity and self-tolerance (Naik et al., 2014). These scientific advancements have improved our understanding about the ubiquitin system mediated redox biology, lipid metabolism, inflammation and immune system regulation in the brain cells by checking the levels of ROS, oxidized proteins, transcriptional factors, and pro-inflammatory agents within limits to maintain cellular homeostasis.

This way, ubiquitination system regulates protein turnover in neurons and glial cells to facilitate multifaceted processes of neuronal signaling, synaptic regulation and neuronal adaptations to maintain brain homeostasis (Fig. 2).

3. Intervening roles of UPS in the neurodegenerative pathology

3.1. UPS in drug addiction mediated behavioral changes and neuronal adaptations

Drug addiction is a compulsive pattern of drug seeking and enduring propensity to relapse it, in effect, of the neuro-adaptations in specific brain regions analogous to long-term memory consolidation. Latest research has defined drug addiction mechanism to be associated with alterations in synaptic transmission in mesocorticolimbic and corticostriatal pathways and the transcriptional changes in neurons for their durable changes in plasticity and behavior. In fact, UPS is evident to control the trafficking or internalization of various receptors, including Glutamate (mGluRs, AMPARs, NMDARs) and non-glutamate receptors (GABA, nAChRs) for maintaining proper synaptic transmissions. For instance, cocaine addiction mediated synaptic plasticity changes have been observed in the nucleus accumbens (NAcc) (Massaly et al., 2013). Similarly, proteasomal degradation of N-ethylmaleimide-sensitive fusion (NSF) protein (a post synaptic density protein), modulate AMPA glutamate receptor's expression and involved in cocaine addiction mediated memory reconsolidation (Ren et al., 2013). In addition, repeated exposure to both amphetamine and cocaine altered glutamatergic neurotransmission in NAcc and potentiated AMPAR synaptic transmission, while synaptic de-potentiation was only associated with cocaine-induced protein synthesis and metabotropic glutamate receptor-5 (mGluR5) activation in NAcc medium spiny neurons (Jedynak et al., 2016). Likewise, proteasomal degradation of GKAP and Shank destabilized PSD-95 and consequently, down regulated NR2B subunit of glutamate N-methyl-Daspartate receptors (NMDARs) to trigger behavioral sensitization in amphetamine addictions in NAcc (D'Souza, 2015; Hung et al., 2010). Another instance where protein ubiquitination is involved in the neuroplastic changes in nAChR's associated synaptic transmission underlying alcohol and nicotine addictions (Feduccia et al., 2012). Furthermore, morphine, cocaine, and nicotine evoked synaptic plasticity was reported with nitric oxide and BDNF induced ubiquitination and reduction in GABA and GABARs in ventral tegmental area and prefrontal cortex of the brain (Lüscher and Malenka, 2011). These studies have highlighted the importance of investigating diverse ubiquitination enzymes involved in the regulation of drug-induced neuroplastic changes to mimic behavioral pathology in human addicts.

3.2. Intricacies of UPS in the pathophysiology of neurodegenerative disorders

A prominent feature of neurodegenerative disorders includes the accumulation of misfolded proteins in response to overwhelming effect of degradative pathways in neurons causing cytotoxicity and neurodegeneration. For instance, the histopathology of AD brain is marked by the presence of neurofibrillary tangles and amyloid containing senile plaques throughout the brain. Moreover, they exhibit progressive atrophy of cortical and subcortical structures, affecting hippocampus and consequent cognitive impairment in the AD patients (Deture and Dickson, 2019). While PD exhibits α-synuclein containing cytoplasmic inclusions called Lewy bodies in the brain resulting in the selective degeneration of dopaminergic neurons in the *substantia nigra pars compacta* and associated motor symptoms in PD (Mhyre et al., 2012). Besides, HD is characterized by the deposition of poly glutamine expansion containing mutant huntingtin protein. That results in the degeneration and atrophy of the striatum, cerebral cortex and other subcortical regions of HD brain with the advancement of disease (Sari, 2011). Additionally, ALS also exhibit inclusion bodies containing SOD1, FUS, Optineurin, and TDP-43, etc., and trigger degeneration in the motor cortex, brain stem and central spinal cord (Ragagnin et al., 2019). On the other hand, MS is an inflammatory demyelinating disease, where auto-antigens and inflammatory agents trigger inflammation, neurodegeneration and tissue damage (Peterson and Fujinami, 2007). These anomalies trigger multiple pathological events, including mitochondrial dysfunction, ER stress, oxidative stress, and inflammation in vulnerable neurons. Importantly, misfolded protein degradation is associated with the principal action of ubiquitin-proteasome system involving the explicit set of ubiquitination enzymes (E1, E2, E3 and DUBs) especially ubiquitin E3 ligase for targeting a specific protein for degradation. Therefore, recent advancements in ubiquitination research have shifted its paradigm from being a prime suspect for neurodegenerative pathophysiology to an attractive therapeutic target for clearing disease-related proteins.

3.2.1. Alzheimer's disease: clearance of Aβ, AβPP, Tau and α-synuclein

Alzheimer's disease (AD) acknowledged as the world's leading cause of dementia, is characterized by extracellular amyloid-β (Aβ) and intracellular tau deposits as senile plaque and neurofibrillary tangle in the brain respectively. Accumulating evidence signifies the alteration in proteostasis network, i.e. impaired proteasome as a salient feature in AD, involving ER stress, oxidative stress, mitochondrial dysfunction, and inflammation mediated neurodegeneration (Fig. 3A) (Keller et al., 2000a). For instance, various studies have shown oligomeric Aβ deposition in the lumen of ER, in effect, of AβPP mutation along with elevated ER stress markers (eIF2α, GRP78) and ubiquitin E3 ligase-HRD1 (Gerakis and Hetz, 2018; Umeda et al., 2011). Moreover, ER stress triggered Aβ neurotoxicity by E2-25K/Hip-2 mediated activation and stabilization of caspase-12 (Fonseca et al., 2014). Likewise, Aβ also induced ER stress mediated JNK3 activation and consequent AβPP phosphorylation and endocytosis for its proteolytic processing to form amyloid beta (Yoon et al., 2012). Similarly, a reduction in MGRN1 E3 ligase levels in hippocampal neurons alter AβPP's trafficking along the secretory pathway and leads to an increased Aβ40 and Aβ42 release (Benvegnù et al., 2017b). Importantly, Keller et al., has demonstrated the role of multicatalytic proteasome (MCP) inhibition in cellular dysfunction and associated neurodegeneration in PC6 cells and Fisher 344 rats to be sensitive to the oxidative stress (Keller et al., 2000b, 2000c). Moreover, UCH-L1, a deubiquitylase enzyme is also reported to be oxidatively modified and dysfunctional in AD brain by the Butterfield group and others (Castegna et al., 2002; Choi et al., 2004; Sultana et al., 2006). The oxidative modifications of UCH-L1 at cysteine (Cys^{152}) and methionine ($Met^{1, 6, 12,124,179}$) residues are responsible for the tau-immunoreactive tangles and Aβ overproduction in AD and PD brains (Choi et al., 2004; Liu et al., 2015). In addition, Guglielmotto et al.,

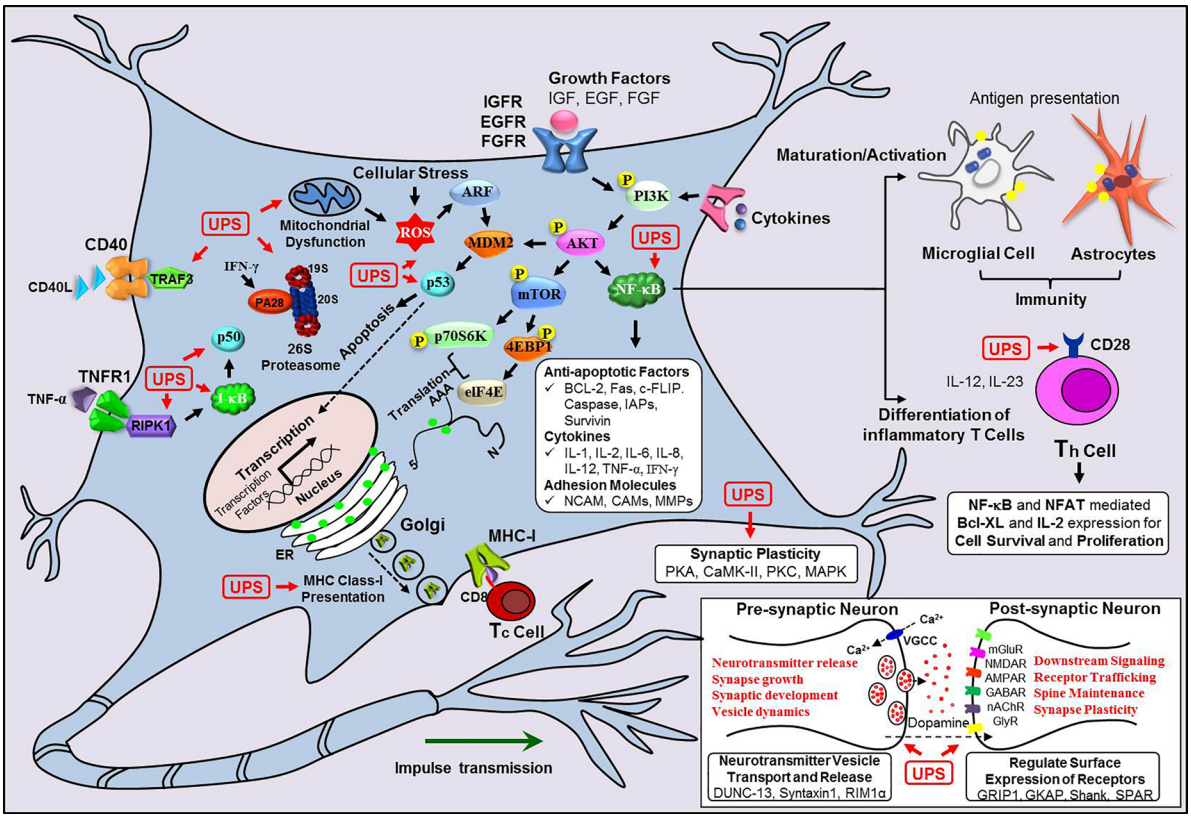


Fig. 2. Role of Ubiquitin Proteasome System in the Brain signaling, Synaptic regulation and Cellular homeostasis: UPS is involved in multifaceted functions, including MHC Class-I antigen presentation, Cellular immunity, T-cell receptor and CD28-signaling, transcription and translation process, and synaptic transmission within the brain. Moreover, it governs mitochondrial trafficking; ROS mediated signaling, and p53 linked apoptotic regulation; NF-kB mediated inflammatory responses (release of anti-apoptotic factors, cytokines and adhesion molecules), diverse receptor's expressions (CD40, TNFR1, Growth factor receptors) to maintain cellular-homeostasis. It also activates microglia and astrocytes to trigger immunoproteasome linked antigen presentation and subsequent T-cells response to generate various cytokines and associated immune defense under cellular stress. Additionally, UPS governs the synaptic transmission by regulating neurotransmitter release, Synapse growth, Synaptic development, Vesicle dynamics in the Pre-synaptic neurons while receptor trafficking, Downstream signaling spine maintenance, and Synaptic plasticity in Post-synaptic neurons. IGF-Insulin growth factor; EGF-Epidermal growth factor; FGF-Fibroblast growth factor; IGFR-Insulin growth factor receptor; EGFR-Epidermal growth factor receptor; FGFR-Fibroblast growth factor receptor; CD8-Cluster of differentiation 8; CD40-Cluster of differentiation 40; TNFα-Tumor necrosis factor alpha; TNFR1-Tumor necrosis factor receptor 1; MHC- Major histocompatibility factor; ER-Endoplasmic reticulum; TRAF3-TNF Receptor Associated Factor 3; RIPK1-Receptor-interacting serine/threonine-protein kinase 1; I-κB-Inhibitor of kappa B; NF-κB-Nuclear Factor kappa beta; IFNγ-Interferon gamma; PA28-Proteasome activator 28; ARF-ADP ribosylation factors; MDM2-Mouse double minute 2 homolog; PI3K-Phosphatidylinositol-4,5-bisphosphate 3-kinase; AKT-Protein kinase B; mTOR-mammalian target of rapamycin; eIF4E-Eukaryotic translation initiation factor 4E; p70S6K-Ribosomal protein S6 kinase beta-1; IAPs-Inhibitor of apoptosis; IL-Interleukins; NCAM-Neural cell adhesion molecule; CAMs-Cell adhesion molecules; MMPs-Matrix metalloproteinases; NFAT-Nuclear factor of activated T-cells; PKA/PKC/CaMK-II-Protein kinases; VGCC-Voltage gated calcium channel; DUNC-13- Drosophila homolog UNC-13; RIM1α-Regulating synaptic membrane exocytosis 1 alpha; GRIP1-Glutamate Receptor Interacting Protein 1; GKAP-Guanylate kinase-associated protein; SPAR-Spine-associated Rap-Gap; mGluR-Metabotropic glutamate receptor; NMDAR-N-methyl-D-aspartate receptor; AMPAR-α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor; GABAR-Gamma-Aminobutyric acid; nAChR-Nicotinic acetylcholine receptor; GlyR-Glycine receptor.

reported UCH-L1 inhibition-based induction of BACE1 up-regulation and consequent neuronal death in control and 5xFAD transgenic AD mouse signifying Aβ peptide accumulation (Guglielmotto et al., 2017). On contrary UCH-L1 overexpression have demonstrated BACE1 degradation and affected APP processing and Aβ production in both *in-vitro* and *in-vivo* studies thereby enhancing synaptic function and contextual memory (Gong et al., 2006; Zhang et al., 2014, 2012). However, Aβ inactivates ubiquitin proteasome system and subsequent impairment in multivesicular bodies (MVB) sorting pathway (Bustamante et al., 2018), nevertheless, UPS also play a decisive role in Aβ clearance. For instance, mutant ubiquitin and ubiquitin-like ubiquilin-1 proteins are linked with Aβ accumulation while up regulated proteasomal degradation of BACE1 and γ-secretase components leads to decrease in Aβ accumulation (Hong et al., 2014; Gong et al., 2010). Another E3 ligase-substrate adaptor protein- FBXO2 expedited the clearance of AβPP and regulate synaptic connections, which otherwise reduced in AD (Atkin et al., 2014). Similarly, over expression of other E3 ligases- Parkin and CHIP along with E2 conjugating enzyme- UbcH5B and heat shock proteins- Hsp70/90 curtailed AβPP expression, Aβ burden, hyper phosphorylated tau and inflammation in the AD mouse model (Gadhave et al., 2016; Sweeney et al., 2017). Moreover, multiple E3 ligase, including Parkin, SIAH1, MUL1, ARIH1, and SMURF1 is reported to regulate mitophagy by ubiquitinating outer mitochondrial membrane proteins (Jacoupy et al., 2019; Reddy and Oliver, 2019). Likewise, MDM2 and NUB1 facilitated tau degradation in the brain by disrupting GSK-3β's interaction with p53 and tau respectively (Proctor and Gray, 2010; Richet et al., 2012). Furthermore, inhibition of proteasomal ubiquitylase Usp14, UCHL1 and intrusion of proteasome activator Blm10 have also enhanced tau degradation studies (Boselli et al., 2017; Tramutola et al., 2016). On contrary very few studies have reported α-synuclein ubiquitination in AD (Mukaetova-Ladinska et al., 2000; Nemes et al., 2004) although numerous studies have shown its aggregation along with amyloid plaques and neurofibrillary tangles in AD (Twohig and Nielsen, 2019). Collectively, the above results depicted the role of UPS in both early and late-stage AD with characteristic synaptic dysfunction and neurodegeneration.

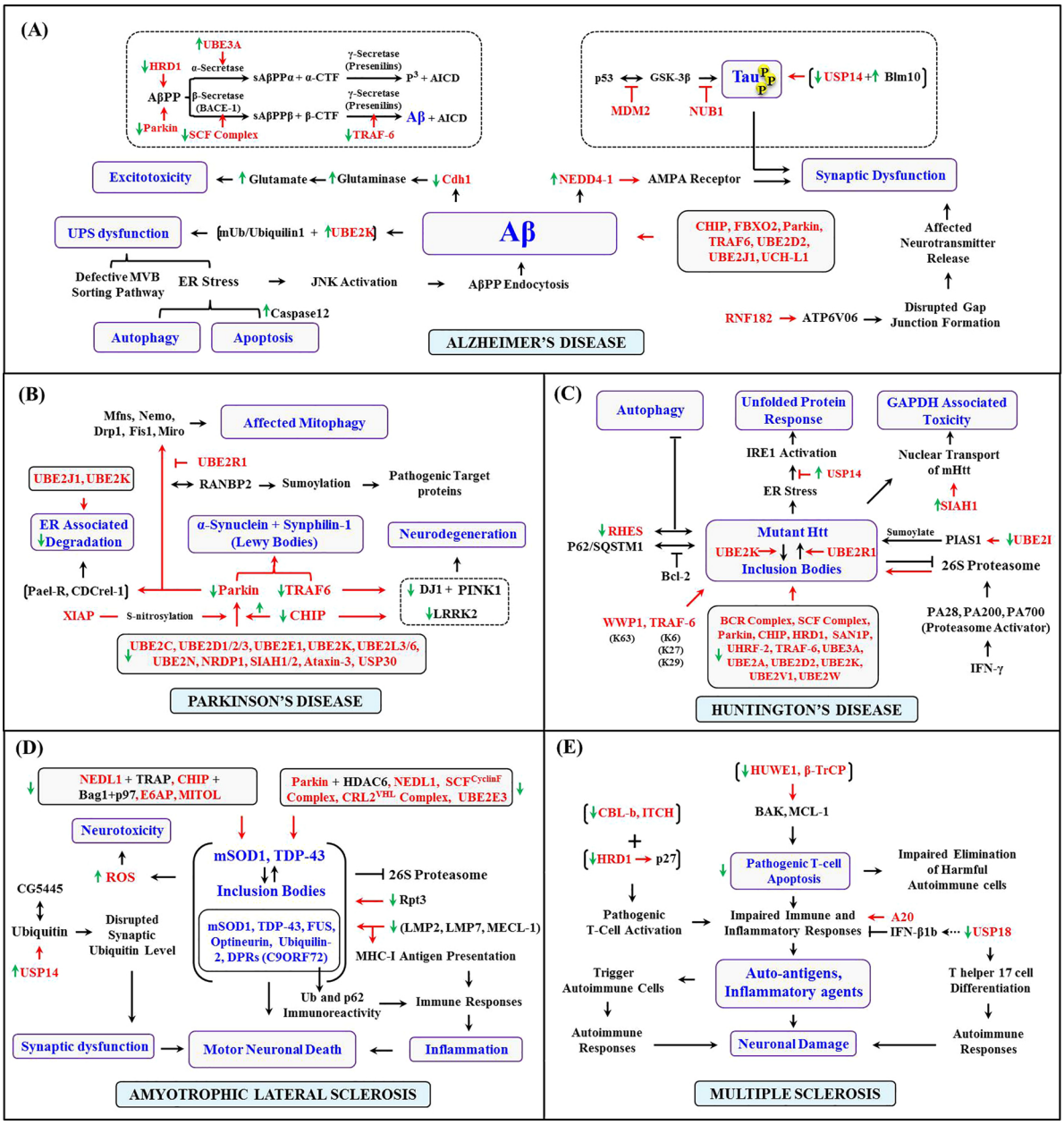


Fig. 3. Ubiquitination mediated neurodegeneration mechanism in neurodegenerative disorders: (A) Alzheimer's Disease- The AβPP is proteolytically processed to generate non-amyloid (p3) and amyloid (Aβ) peptide with help of α- and β-secretase respectively, followed by γ-secretase. Under pathogenic state, E3 ligases HRD1 and Parkin failed to regulate the level of AβPP and give rise to its amyloidogenic products. Similarly, down-regulated SCF complex, TRAF-6 and up regulated UBE3A triggered the selection of amyloidogenic pathways for Aβ synthesis. Moreover, disrupted Cdh1 activity, lead to glutaminase overload associated excited-toxicity in neurons. Furthermore, Neurotransmitter release and synaptic functions are compromised due to impaired AMPA signaling, and gap junction formation due to NEDD4-1 and RNF182 mediated proteasomal degradation of AMPA receptors and ATP6V06 (Endosomal proton pump member). Likewise, multi vesicular body-sorting pathway for lysosomal degradation is affected due to UBE2K mediated mutant ubiquitin or ubiquitin like ubiquilin-1 conjugation. In addition, ER-stress trigger JNK activation and induces endocytosis of AβPP to produce Aβ peptides. Consequently, it governs autophagy and caspase-12 mediated apoptosis as a facultative response against Aβ toxicity. Several other UPS components like CHIP, FBXO2, Parkin, TRAF6, UBE2D2, UBE2J1 and UCH-L1 have reported to clear Aβ that get affected in the diseased state. Like Aβ, Tau protein is marked by down regulated USP14 and elevated proteasome activator Blm10 for proteasomal degradation. Moreover, NUB1 prevents hyper-phosphorylation of Tau by Glycogen synthase kinase-3beta (GSK-3β) to alleviate its toxicity, while MDM2 mediates inhibition of interaction between GSK-3β and p53 that enables p53 to trigger transcriptional activation of protective genes to clear off hyper phosphorylated tau. These UPS components get disrupted in diseased state and caused hyper-phosphorylated tau mediated neurotoxicity due to the neuro-fibrillary tangles. (B) Parkinson's Disease- The Lewy body (α-synuclein and Synphilin-1) pathology is mainly governed by defective E3 ligase Parkin. It regulates mitochondrial membrane proteins (Mfns, Nemo, Drp1, Fis1, and Miro) and is affected in the diseased state that triggers impaired mitophagy. Likewise, defective parkin disrupted endoplasmic reticulum associated degradation (ERAD) by affected Pael-R and CDCrel-1 ubiquitination. It also triggers SUMOylation via interacting with RANBP2 for the clearance of pathogenic target proteins. Similarly, CHIP and TRAF6 are associated with the ubiquitination of LRRK2 and PINK1/DJ1 respectively that get affected in the disease state and cause neurodegeneration. (C) Huntington's Disease- The mutant huntingtin forms inclusion bodies- both soluble and insoluble forms. Here, Proteasome activators PA28, PA200 and PA700 are helpful in reversing the inhibition of proteasome by mHtt and inclusion bodies. However, it governs ER stress mediated unfolded protein response and induces GAPDH associated toxicity through its nuclear transport by SIAH1. Many UPS enzymes act differently i.e. positively and negatively to regulate the clearance of mHtt, including WWP1 and TRAF6 that direct K63 and K6, K27, K29 poly-ubiquitination respectively, and retard proteasomal degradation under diseased state. (D) Amyotrophic Lateral Sclerosis- Inclusion bodies in ALS is populated with mutant SOD1, TDP-43, FUS, Optineurin, Ubiquilin2 and dipeptide repeat proteins that affect the different cellular process in neurons. Impaired proteasome and MHC-I mediated immunity and inflammation triggers the

neuronal death in the diseased state. Moreover, CG5445 and USP14 scavenge the ubiquitin protein and thereby interfere with the synaptic functions in ALS. Furthermore, mitochondrial dysfunction mediated ROS production also triggers inflammation and associated neurotoxicity in ALS. (E) Multiple Sclerosis- Auto-antigens and inflammatory agents are the principal cause for the progression of MS. Where depleted expression of certain E3s, including HUWE1, β-TrCP, CBL-b, ITCH and HRD1 lead to compromised pathogenic T-cell apoptosis and its activation to aggravate the symptoms of MS. However, IFN-β1b treatment enabled the suppression of immune and inflammatory responses by targeting auto-antigens and inflammatory agents who prevented neuronal damage in Multiple sclerosis. Dashed arrow- Regulation; Red arrow- Ubiquitination and degradation; Black arrow- Activation; Up arrow (Green) - Elevated expression; Down arrow (Green) - Reduced expression.

3.2.2. Parkinson's disease: Clearance of α-synuclein in Lewy bodies

Parkinson's disease (PD) is one of the prevalent neurodegenerative disorders characterized by progressive loss of muscular coordination and control with primary α-synuclein deposits in Lewy bodies. The brains of PD patients revealed cellular stress, including oxidative stress, ER stress and mitochondrial dysfunction to be the integral part for impairment in proteostasis and dopaminergic neuronal loss (Fig. 3B). In fact, several lines of evidence have presented an intimate relation between UPS aberrations and PD pathogenesis. For instance, deletion of PSMC1/Rpt2 (ATPase Subunit) and variations in PSMC4/Rpt3 proteasomal subunits have exhibited ubiquitin and α-synuclein positive neurons, thereby suggesting the role of UPS in α-synuclein degradation (Vilchez et al., 2014). Similarly, aberrations in E3 ligase Parkin and ubiquitylase UCH-L1 have been widely reported to cause early-onset familial forms of PD with characteristic symptoms of rigidity, tremor, and bradykinesia (Mabb and Ehlers, 2010). Since, mutant Parkin failed to clear toxic accumulation of their substrates, including, Pael-R and CDCrel-1 and contribute towards ER stress associated neurodegeneration (Dawson and Dawson, 2010). Likewise, Parkin in association with peptidases like PINK1 and DJ-1 is famous for the regulation of mitochondrial dynamics, which otherwise lead to oxidative stress, lipid peroxidation and consequent toxicity in neurons (Cenini et al., 2019). Recent studies have revealed that PINK1 activates Parkin in response to mitochondrial depolarization by phosphorylating ubiquitin at Ser65 that recruits autophagy receptors to clear damaged mitochondria via mitophagy (Chin and Li, 2016). Altogether, these proteins and SIAH-1/2 are known to facilitate proteasomal degradation of parkin's substrate synphillin-1 and thus prevent inclusion formation in association with α-synuclein (Szargel et al., 2008). Another Parkin's substrate, PARIS (ZNF746) reported to suppress PGC-1α expression and consequent neurodegeneration in PD signifying another mechanism of Parkin associated neuro protection in PD (Shin et al., 2011). Interestingly, USP13 mediated Parkin's ubiquitination may facilitate its stability via interaction with Beclin-1 and contribute towards the clearance of α-Synuclein (Liu et al., 2019; Lonskaya et al., 2013). Furthermore, a ubiquitylase UCHL1 is also evident to function like E3 ligase that governed K63-linked ubiquitination of alpha-synuclein and promotes its clearance, while its mutation leads to α-Synuclein accumulation (Cartier et al., 2012). Several other ubiquitin-linked proteins such as HRD1, MDM2 and CHIP governed an important role in maintaining protein homeostasis to prevent neurotoxicity in PD. For instance, HRD1 is associated with degradation of ER linked proteins like Pael-R, while MDM2 regulates GSK3B levels and CHIP activates Parkin to mitigate ER stress, proteotoxicity, and mitochondrial stress in PD (Kaneko, 2016; Tetzlaff et al., 2008). On the other side, mutant α-synuclein also found to curtail proteolytic activity of UPS via direct interaction with its 20S core proteolytic unit and attain filament structure. Moreover, recent study has identified CDK5 mediated phosphorylation dependent proteasomal degradation of E3 ligase GP78 triggering neural death in PD (Wang et al., 2018b). Based on these evidences, there is a direct relation between abnormal synuclein aggregation and dopaminergic neuronal death in PD.

3.2.3. Huntington's disease: Clearance of mHtt and inclusion bodies

Huntington's disease (HD) is a poly glutamine expansion disorder of huntingtin protein causing progressive neurodegeneration. Its pathological hallmark includes the mutant huntingtin (mHtt) aggregates with positive inclusions of the ubiquitin proteasome system. In fact, different studies have reported both aggregated as well as soluble forms of polyubiquitinated mHtt in HD transgenic R6/2 mice model, HD patient material, and post-mortem brains (Chang et al., 2015; Juenemann et al., 2015). In spite of mHtt ubiquitination, proteasomes are incapable of cleaving poly glutamine tract in these proteins thereby leading to its aggregation in neurons (Fig. 3C). For instance, in a study mammalian 26S proteasome was found to cleave mHtt within flanking sequences or after first glutamine of a polyQ peptide, but unable to digest polyQ repeats that obstructed their passage through proteasomes and caused its impairment and aggregation in neurons (Juenemann et al., 2013). Moreover, FRET experiments have defined mHtt's interaction with LMP2 subunit of proteasome, thereby impairing their chymotrypsin and caspase like activities leading to compromised peptide cleavage in HD patients (Holmberg et al., 2004). It has previously observed that only oligomeric or soluble forms of mHtt triggered proteasomal impairment while inclusion bodies failed to do so (Díaz-Hernández et al., 2006). Scientists have reported improvement in mHtt clearance by altering proteasomal activities. For instance, proteasome activators such as PA28, PA200, PA700 and 20S catalytic immune-subunits- LMP2 (PSMB9), LMP7 (PSMB8), and Mecl-1 (PSMB10) facilitated polyQ cleavage and clearance due to IFN-γ stimulation (Heink et al., 2005; Schipper-Krom et al., 2012). Interestingly, optical pulse chase experiments have demonstrated selective compartmentalized degradation of mHtt i.e. faster removal in the cytoplasm of the cell body than in neuronal processes (Zhao et al., 2016). Apart from proteasome alteration, E3 ligases have demonstrated to clear mHtt through ubiquitination. For example, nuclear ubiquitin ligases SAN1P and UHRF-2 have facilitated the degradation of mHtt aggregates (Iwata et al., 2009). Likewise, HRD1, Parkin and CHIP also decreased inclusion bodies through mHtt ubiquitination (Harding and Tong, 2018). Similarly, Bcl-2 along with Parkin and VCP found to decrease the affinity of p62/SQSTM1 to polyubiquitinated proteins and consequent reduction in mHtt aggregates (Zhou et al., 2015). Another study revealed K48 mediated ubiquitination of Htt fragments in UBE3A dependent manner (Bhat et al., 2014). Interestingly, other atypical ubiquitination (K6, K27, K29) by TRAF6 and K63 by WWP1 of mHtt found to inhibit proteasome mediated mHtt degradation and subsequent aggregation in HD (Nath and Lieberman, 2017; Zucchelli et al., 2011). Where as an inflammatory kinase like IKK found to promote mHtt degradation through proteasomal and lysosomal pathways (Thompson et al., 2009). Likewise, some E2s conjugating enzymes, including, hE2-25k, CDC34, and UBE2W are reported to regulate the ubiquitination and solubility of mHtt aggregate in different HD models (Mitra and Finkbeiner, 2008; Wang et al., 2018a). Interestingly, recent fundings about ubiquitination of mHtt containing inclusion bodies suggested it a dynamic process involving the recruitment of (de)ubiquitination proteins continuously to trigger ubiquitin-specific clearance pathways in HD (Juenemann et al., 2018).

3.2.4. Amyotrophic lateral sclerosis: Clearance of SOD1, TDP-43 and inclusion bodies

ALS is a rare motor-neurodegenerative disorder affecting brain and spinal cord that triggers progressive paralysis with troubles in chewing, talking, and walking. Recent studies have reported UPS associated impaired turnover of several proteins, including SOD1, FUS, Optineurin, TDP-43, and translation product of intronic hexanucleotide

Table 2
Mechanistic role of ubiquitin.(ation enzymes for the clearance of toxic protein aggregates in neurodegenerative diseases (Fuchsberger et al., 2017; Joshi et al., 2016; Saito et al., 2014; Rodrigues et al., 2016; Rosen et al., 2010; Liu et al., 2008; Gong et al., 2010; Geetha et al., 2012; Singh et al., 2017; Zhao et al., 2016; Morgan et al., 2007; Hong et al., 2014; Burr et al., 2011; Guglielmotto et al., 2017; Ko et al., 2009; Allen et al., 2005; Zhong et al., 2005; Dawson and Dawson, 2010; Rott et al., 2017; Xu et al., 2015; Murata et al., 2013; Mulherkar et al., 2009; Tsang et al., 2009; Haddad et al., 2013; Fiesel et al., 2014; Geisler et al., 2014; Araki and Nagata, 2011; Fiesel et al., 2014; Durcan et al., 2012; Ristic et al., 2014; Gersch et al., 2017; Alexopoulou et al., 2016; Lu et al., 2013; Miller et al., 2005; Mealer et al., 2014; Bhutani et al., 2012; Bae et al., 2006; Zucchelli et al., 2011; Maheshwari et al., 2014; Howard et al., 2007; Ochaba et al., 2016; de Pril et al., 2007; Wang et al., 2018a; Hyrskyluoto et al., 2014; Uchida et al., 2016; Allen et al., 2005; Miyazaki et al., 2004; Hans et al., 2014; Vaden et al., 2015; Xu et al., 2016; Mandel et al., 2012; Sellebjerg et al., 2012; Honke et al., 2016; Malhotra et al., 2013).

NDDs	UPS Components	Molecular Functions	Mechanism of Action	References
Alzheimer's Disease (Aβ, Tau)	**APC/Cdh1**	Regulate glutamate level to control Aβ linked excitotoxicity in neurons	The ubiquitin E3 ligases such as SCF complex, TRAF6, and HRD1 regulate proteasomal degradation of β-secretase, Presenilin (γ-secretase) and AβPP respectively, thereby controlling the level of Aβ in the brain. Other E3s, including Parkin and CHIP directly clear Aβ in the brain. While downregulated UBE3A inhibit Aβ formation by preventing α-secretase degradation. On contrary Aβ also inhibit few E3s, including Cdh1 and NEDD4-1 thereby affecting membrane receptors, synaptic density and neurotransmission. In addition, multiple E2s, including UBE2A, UBE2D1, UBE2K, and deubiquitinases- UBE2J1 and UCH-L1 are associated with the conjugation and recycling of ubiquitin to regulate turnover of various synaptic proteins. However, in diseased state alterations in these E2s, E3s and DUBs affect gap junctions, membrane receptors, synaptic protein densities, and associated synaptic transmission and neurological behaviors in AD patients.	(Fuchsberger et al., 2017)
	CHIP	Ubiquitinates phosphorylated Tau and also promote Aβ degradation		(Joshi et al., 2016)
	HRD1	ER-associated degradation of misfolded MHC-I heavy chains, Pael-R and AβPP		(Saito et al., 2014)
	NEDD4-1	Aβ triggered ubiquitination/ degradation of AMPARs to regulate synaptic function, strength and dendritic spine density		(Rodrigues et al., 2016)
	Parkin	Reduces Aβ levels and improve long term potentiation and behavioral abnormalities		(Rosen et al., 2010)
	RNF182	Ubiquitinate ATP6V06 protein and disrupt gap junction formation and neurotransmitter release in the diseased state		(Liu et al., 2008)
	SCF Complex (SKP1, CUL1, Fbx2)	Degrade BACE1 enzyme that attenuate amyloidosis and improves synaptic function		(Gong et al., 2010)
	TRAF6	Abrogate Aβ-mediated inhibition of p75 neurotrophin receptor's ubiquitination and restore neuronal cell survival. Also degrade mutated Presenilin		(Geetha et al., 2012)
	UBE3A	Ubiquitination of α-secretase (ADAM10) and its compromised action reduces Aβ level through non-amyloidogenic pathway		(Singh et al., 2017)
	UBE2A	Coordinates amyloid peptide's degradation through proteolysis		(Zhao et al., 2016)
	UBE2D1	Genetically associated with the late onset Alzheimer's susceptible gene APOE		(Morgan et al., 2007)
	UBE2K (E2-25K/Hip2)	Upstream regulator of ASK1 and JNK that manage Aβ neurotoxicity. Also, colocalize with mutant ubiquitin (UBB^{+1}) to inhibit proteasome activity resulting in Presenilin accumulation		(Hong et al., 2014)
	UBE2J1,	Increase cellular levels of mono-ubiquitin and consequent protein turnover by UPS that improves contextual memory, and restores synaptic function in brain		(Burr et al., 2011; Guglielmotto et al., 2017)
	UCH-L1			
Parkinson's Disease (Lewy bodies; α-synuclein, Synphilin1)	**CHIP**	Enhance parkin's activity and ubiquitinates α-synuclein and LRRK2 for degradation	Multiple E3 ligases, including Parkin, TRAF6, SIAH1, UBE3A, CHIP and DJ1 directly or indirectly ubiquitinate and clear α-synuclein from the cell. Among them, Parkin is the key E3 ligase centrally regulating the clearance of alpha synuclein with help of other E3 ligases- NRDP1, XIAP, CHIP, and DJ1, where NRDP1 and XIAP1 stabilize Parkin, while CHIP enhances Parkin's activity. Moreover,	(Ko et al., 2009)
	Gigaxonin	Regulate degradation of microtubule associated protein 1B and intermediator filaments to maintain neuronal health		(Allen et al., 2005)
	NRDP1	Regulate Parkin's stability and activity via regulating its proteasomal degradation		(Zhong et al., 2005)
	Parkin	Ubiquitinate and clear α-synuclein aggregates and various mitochondrial membrane proteins to regulate mitophagy		(Dawson and Dawson, 2010)
	RANBP2	SUMOylate its pathogenic target and trigger protein degradation through Parkin interaction		(Rott et al., 2017)
	SIAH1	Ubiquitinate and degrade alpha-synuclein and synphilin-1 proteins		(Xu et al., 2015)
	TRAF6	Promotes ubiquitination of α-synuclein, DJ1		(Murata et al.,

(*continued on next page*)

Table 2 (*continued*)

		and PINK1	Parkin also regulates the ubiquitination of other E3 ligase- RANBP2 and its SUMOylated mitochondrial membrane proteins to regulate mitophagy. While Gigaxonin ubiquitinates MAP1B and regulate the stability of intermediators filaments. Altogether, these E3 ligases along with a diverse set of conjugating enzymes, including UBE2A, UBE2C, UBE2D, UBE2E, UBE2J, UBE2K, UBE2L, UBE2N, UBE2R and various deubiquitinases like Ataxin3, UCHL1, USP8, USP9X, USP30 regulate membrane trafficking, synaptic transmission and microtubular organization to control neuro-muscular functions in PD	2013)
	UBE3A	Colocalize with alpha synuclein to clear it in a proteasome-dependent manner		(Mulherkar et al., 2009)
	XIAP	Trigger S-nitrosylation of Parkin and thereby affect neuronal survival in PD		(Tsang et al., 2009)
	UBE2A	Associate with parkin and regulate the clearance of dysfunctional mitochondria in mice and human cells		(Haddad et al., 2013)
	UBE2C, UBE2D family members, UBE2-E1 /K /L3 /L6 /N	Contribute towards mitophagy by redundantly charging RING-HECT hybrid ligase Parkin with ubiquitin		(Fiesel et al., 2014; Geisler et al., 2014)
	UBE2K, UBE2J1	Ubiquitinate and selectively degradation misfolded membrane proteins from the ER		(Araki and Nagata, 2011)
	UBE2R1	Negative regulator of Parkin that restrict parkin translocation to mitochondria		(Fiesel et al., 2014)
	Ataxin-3	Coupled with parkin to govern ubiquitination and decomposition of polyubiquitin chain		(Durcan et al., 2012)
	UCH-L1, USP9X	Mutated form lead to impairment in abnormal protein clearance		(Ristic et al., 2014)
	USP30	Associated with PINK1 and Parkin to clear damaged mitochondria in PD		(Gersch et al., 2017)
	USP8	Associated with the clearance of α-synuclein and modify its toxicity in PD		(Alexopoulou et al., 2016)
Huntington's Disease (PolyQ, mHtt)	**BCR Complex (BTB-CUL3 - RBX1)**	Associated with the proteasomal degradation of mHtt in association with NUB1-NEDD8-CUL3 pathway	Diverse E3s act differently to regulate the cytotoxicity of mHtt aggregates. For instance, BCR complex, SCF complex, TRAF6 and CHIP regulate mHtt levels via ubiquitination and proteasomal degradation. Besides, SIAH1 regulate nuclear transport associated toxicity of mHtt. In addition, UBE3A regulate dopamine regulated phosphoprotein (DARPP-32) and directs neurotransmission. Similarly, RHES binds Beclin1 to regulate autophagy process and overcome the cytotoxicity caused by the protein aggregates. Collectively these E3s with the assistance of multiple E2s and DUBs- UBE2A, UBE2D2, UBE2K, UBE2V1, UBE2W, UBE2I, and USP14 effectively remove the cellular aggregates of mHtt to relieve the cytotoxicity in neurons.	(Lu et al., 2013)
	CHIP	suppression of aggregation and toxicity of PolyQ huntingtin		(Miller et al., 2005)
	RHES	Binds autophagy regulator Beclin-1 to activate autophagy, while its co-expression with mHtt blocks Rhes-induced autophagy		(Mealer et al., 2014)
	SCF Complex (Skp1-Cul1-F-box)	Deficient Cul1 and Skp1 are associated with increased mHtt levels in brain		(Bhutani et al., 2012)
	SIAH1	Enhance nuclear translocation of mHtt and cytotoxicity in association with GAPDH		(Bae et al., 2006)
	TRAF6	Associated with ubiquitination and degradation of mHtt and PolyQ proteins		(Zucchelli et al., 2011)
	UBE3A	Its deficiency trigger protein overload and decrease in dopamine regulated phosphoprotein DARPP-32, and play crucial role in HD progression		(Maheshwari et al., 2014)
	UBE2A, UBE2D2, UBE2V1	Ubiquitinates poly-glutamine aggregates and controls the size and number of protein aggregates		(Howard et al., 2007)
	UBE2I	Conjugate SUMO protein to E3 SUMO ligase- PIAS1, which regulates SUMOylation of huntingtin protein		(Ochaba et al., 2016)
	UBE2K (E2-25/Hip2)	Interacts directly with huntingtin and may mediate ubiquitination of neuronal intranuclear inclusions in HD		(de Pril et al., 2007)
	UBE2W	Markedly decreases mHTT aggregate formation and increases the level of soluble monomers		(B. Wang et al., 2018)
	USP14	Effectively reduce cellular aggregates and prevent cellular degeneration by counteracting ER stress-mediated IRE1 activation		(Hyrskyluoto et al., 2014)

(*continued on next page*)

Table 2 (*continued*)

Amyotrophic Lateral Sclerosis (TDP43, mSOD1, TRAP, Dvl1 and other insoluble protein aggregates)	**$CRL2^{VHL}$ Complex (Cul2-Rbx1-EloBC-VHL)**	Role in the ubiquitination and proteasomal degradation of misfolded TDP-43	The multiple protein aggregates, including TDP-43, mSOD1, TRAP, Dvl1 and other insoluble proteins are prominent in ALS. Multiple E3 ligases such as $CRL2^{VHL}$ complex, NEDL1 and Gigaxonin are reported to trigger ubiquitination and proteasomal degradation of TDP-43, mSOD1 and MAP1B proteins. Altogether, these E3s along with E2s- UBE2E3/ UBCM2 and deubiquitinase USP14 regulate ER-associated degradation, intermediary filaments, neuromuscular junctions and synaptic transmission to regulate motor functions in ALS.	(Uchida et al., 2016)
	Gigaxonin	Regulate intermediary filament by ubiquitination and degradation of microtubule associated protein-1B		(Allen et al., 2005)
	NEDL1	Associated with the ubiquitination and degradation of mutant SOD1, translocon-associated protein-delta (TRAP), and Dishevelled-1 (Dvl1)		(Miyazaki et al., 2004)
	UBE2E3/ UBCM2	Enhance the ubiquitination of TDP-43 and controls its solubility and associated neurodegeneration		(Hans et al., 2014)
	USP14	Maintain synaptic ubiquitin level for proper synaptic functions and development of neuromuscular junctions		(Vaden et al., 2015)
Multiple Sclerosis (Soluble oligomers, auto-antigens and auto-inflammatory agents)	**HRD1**	Facilitates T-cell proliferation by the destruction of cyclin-dependent kinase inhibitor $p27^{kip1}$ and thus aids in T-cell mediated immunity	Multiple E3s regulate cellular functions such as HRD1 ubiquitinate p27 (a CDK inhibitor) and elevate T-cell expression. Similarly, HUWE1 and FBXW1A regulate the expression of anti- and pro-apoptotic proteins MCL1 and BAK respectively that govern T-cell survival. Additionally, ITCH and CBLB also regulate T cell antigen surface expression. Collectively these UPS enzymes regulate T-cell expression, neuro-inflammation, and autoimmune responses to protect against myelin damage in MS.	(Xu et al., 2016)
	HUWE1	Regulate the levels of anti-apoptotic protein MCL-1 and pro-apoptotic BAK, which get disrupted and thus enhance the survival of pathogenic T-cells		(Mandel et al., 2012)
	FBXW1A (β-TrCP)			
	ITCH	Maintain cell surface expression of cytotoxic T lymphocyte antigen 4 (CTLA-4) on $CD4^{+}CD25^{high}$ T cells thereby regulating T cell function in MS		(Sellebjerg et al., 2012)
	CBLB			
	USP18	Acts as a negative regulator of type-I interferon signaling and thus a therapeutic response to IFN-β in MS pathology. Also necessary for T helper-17 cell differentiation and autoimmune responses		(Honke et al., 2016; Malhotra et al., 2013)

Ubiquitin E3 Ligase- Red color; **Ubiquitin E2 Conjugating Enzyme- Green color**; **Deubiquitinase Enzyme- Blue color**

Ubiquitin E3 Ligase- Red color; Ubiquitin E2 Conjugating Enzyme- Green color; Deubiquitinase Enzyme- Blue color.

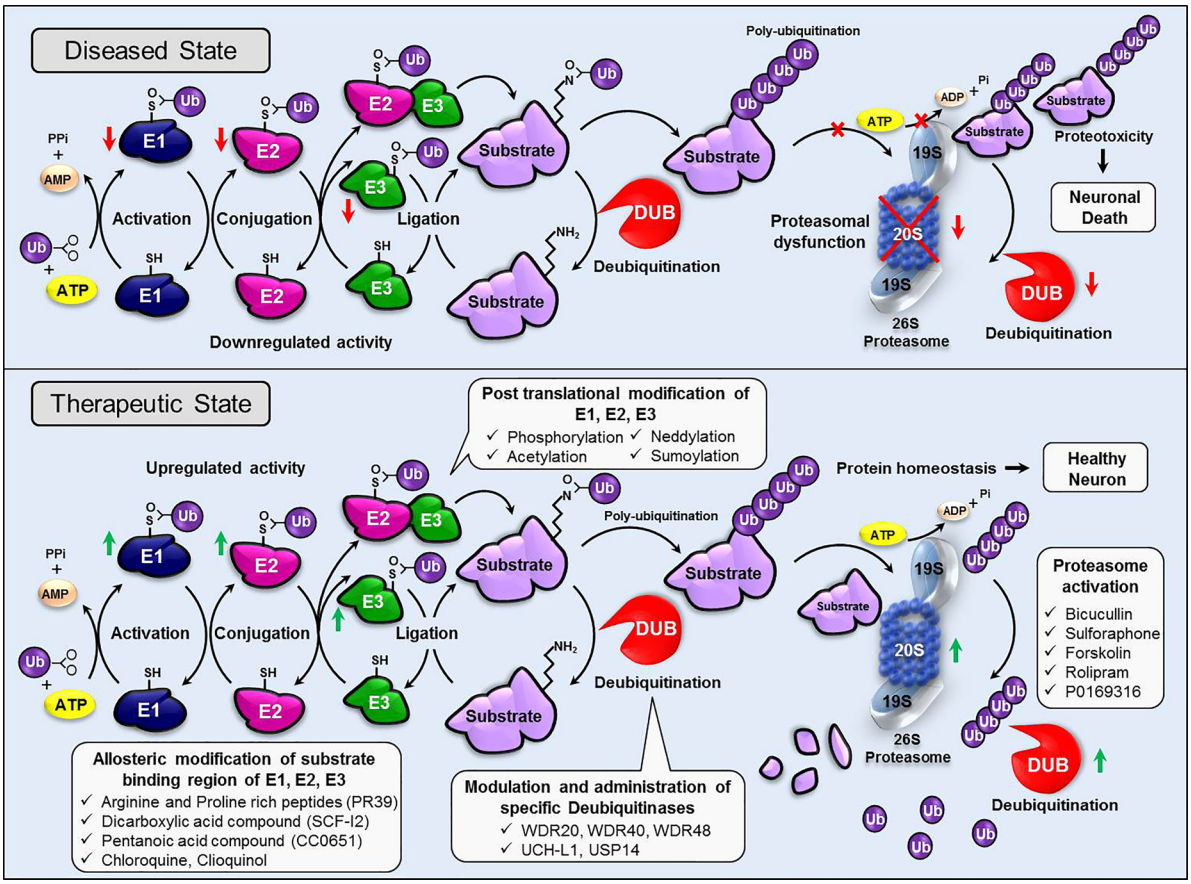


Fig. 4. Therapeutic approaches to alleviate neurodegenerative pathologies: Under diseased conditions the activity of ubiquitinating enzymes (E1s, E2s, and E3s), deubiquitinase enzymes (DUBs), proteasome assembly and functions are compromised leading to proteotoxic stress in the neurons causing its death. The disease pathologies might be reversed upon adapting the following remedial strategies- i) Allosteric modification of the substrate binding region, which could increase or decrease the activity of E1, E2 and E3 enzymes in order to regulate the clearance of toxic, misfolded or unwanted proteins. ii) Post translational modification of E1, E2, and E3 enzymes lead to the protection of these enzymes from their self-ubiquitination and degradation thereby enabling them to be active during diseased conditions for elevated clearance of harmful proteins. iii) Modulation and administration of specific deubiquitinases could aid in overcoming the burden of native or mutant ubiquitin charged enzymes and substrates, facilitating the UPS machinery for the efficient clearance; iv) Proteasomal activation aids in the mitigation of proteotoxic burden of neurons enabling them to maintain their homeostatic state. E1-Ubiquitin E1 activating enzyme; E2-Ubiquitin E2 conjugating enzyme; E3-Ubiquitin E3 ligase; DUB-Deubiquitinase enzyme; Ub-Ubiquitin protein; ATP-Adenosine triphosphate; AMP-Adenosine monophosphate; PPi-Pyrophosphate.

repeats of C9ORF72 gene as the most susceptible cause for ALS (Fig. 3D) (Burk and Pasterkamp, 2019). For instance, conditional knockout of proteasomal subunit PSMC4/Rpt3 disclosed aggregates of TDP-43 and FUS proteins with affected locomotor function and motor neuronal loss, which were absent in autophagy related protein Atg7 knockout study, signifying the involvement of proteasome in ALS development (Tashiro et al., 2012). Moreover, inducible subunits of proteasome including, LMP2, LMP7 and MECL-1 get overexpressed early during disease pathogenesis and generates peptides crucial for major histocompatibility class-I presentation and subsequent immunity in the SOD1 mice model of ALS (Johnston-Carey et al., 2016). Likewise, alteration in E3 ligases- TRAF6 and Mahogunin ring finger 1 (MGRN1) conferred mutant SOD1 linked cytotoxic protein aggregates in the ALS mouse model (Chhangani et al., 2016; Semmler et al., 2019). Conversely, the role of UPS has also been demonstrated in the clearance of ALS pathological proteins. For example, the clearance of mutated or excessive SOD1 in ER is triggered by an E3 ligase NEDL1 with help of translocon protein TRAP, while in the cytosol by E3 ligases- E6-AP and Dorfin (Mishra et al., 2013). Another E3 ligase CHIP along with ubiquitin binding proteins- Bag1 and VCP (p97), facilitated mutant-SOD1 degradation by translocating it from 19S ATPase subunit to core 20S proteasome particle (Choi and Lee, 2010). Moreover, mitochondrial E3 ligase MITOL is found to ubiquitinate mSOD1 and mitofusin2 to regulate mitochondrial dynamics and its associated ROS production (Sugiura et al., 2013). Similarly, TDP-43 is ubiquitinated by Parkin and facilitated its translocation from the nucleus to cytosol and its accumulation in a multi-protein complex with HDAC6 (Hebron et al., 2013). Moreover, a novel E3 ligase ZNF179 is evident to ubiquitinate TDP-43 and ameliorates TDP-43 associated neuropathy in ALS (Lee et al., 2018). Furthermore, a CCNF missense mutation has identified to alter E3 ligase complex ($SCF^{Cyclin-F}$) thereby causing accumulation of abnormally ubiquitinated TDP-43 proteins (Williams et al., 2016). Likewise, HRD1 reported to regulate optineurin degradation and microtubule-dependent aggresome formation (Mao et al., 2017). Additionally, C9ORF72 gene associated mutations are responsible for dipeptide repeats in them leading to their aggregation and immunoreactivity against ubiquitin and p62 in ALS (Freibaum and Taylor, 2017; Schludi et al., 2015). Recently, a p53 family protein (p63) has been reported to augment cellular levels of E3 ligase TRIM63 (MuRF1) that regulates the degradation of muscular proteins during muscular-atrophy in ALS (von Grabowiecki et al., 2016). This way, UPS regulates protein turnover dynamics in ALS to combat neurodegenerative pathologies.

3.2.5. Multiple sclerosis: Clearance of auto-antigens and inflammatory agents

Multiple sclerosis (MS) is a chronic inflammatory progressive neurodegenerative disorder of CNS, triggering demyelination and muscular

Table 3
Potential compounds modulating UPS activity for therapeutic implications in neurodegenerative disorders.

S.No.	Compounds (PubChem CID)	Impact on UPS	Target UPS Component	Molecular Action	Target Pathology	BBB permeability	References
1	Benzamil (108107108,107)	UPS activator	_	Rescue acid sensing ion channel (ASIC) dependent acid toxicity and enhance UPS activity	HD, MS	Yes	(Durham-Lee et al., 2011; Friese et al., 2007; Wong et al., 2008)
2	Betulinic acid (64,971)	Proteasome activator	20S β5 proteasome subunit	Activates chymotrypsin-like activity of proteasome and enhance GABAergic activity, Prevent neurobehavioral activity in Alzheimer's disease	AD, Autism	Yes	(Das et al., 2016; Huang et al., 2007; Navabi et al., 2018)
3	Bicuculline (10,237)	Proteasome activator and UPS activator	Rpt6 (19S Proteasome regulatory subunit) and other UPS component	Stimulate calcium/calmodulin-dependent protein kinase II (CaMKII) mediated proteasomal activity stimulate AMPARs ubiquitination to modulate synaptic connections in neurons	AD, PD	No	(Djakovic et al., 2009; Widagdo et al., 2017)
4	Clioquinol (2788)	Proteasome inhibitor	20S Proteasome	Inhibit 20S proteasome inhibition by clioquinol-copper complex and affect its chymotrypsin-like activity. Disrupt interaction between metals and Aβ peptide in the brain and down regulates mutant Htt level in HD models	AD, HD	Yes	(Regland et al., 2001; Wang et al., 2009)
5	Curcumin (969,516)	_	APC, HSP70, HSP90	Inhibits APC leading to cell cycle arrest and apoptosis in mylenomas. Increase Hsp70/90 activity to prevent amyloid formation and subsequent neuronal death	AD, Tauopathies, HD	Yes	(Lee and Langhans, 2012; Maiti et al., 2014; Tsai et al., 2011)
6	Delta-9-Tetrahydrocannabinol (Δ9-THC) (5,321,846)	DUB Stimulator	USP3	Its antioxidant effects and increased level of USP3 exert direct neuroprotective role in cellular models of Parkinson's disease	PD	Yes	(Carroll et al., 2012; Schou et al., 1977)
7	Doxycycline (54,671,203)	E3 inhibitor	HUWE1	Reshapes α-synuclein oligomers into off-pathway, high-molecular-weight species that do not evolve into fibrils preventing its toxic effects	AD, PD	Yes	(Balducci and Forloni, 2019; González-Lizárraga et al., 2017)
8	Forskolin (47,936)	Proteasome inhibitor	26S proteasome	Enhance proteasomal degradation of protein aggregates (mutant FUS, SOD1, TDP43, Tau) in cAMP/PKA pathway dependent manner	PD, ALS	Yes	(Guo et al., 2016; Lokireddy et al., 2015)
9	IU1 (675,434)	Proteasome activator	USP14	Boost proteasome activity by ameliorating USP14 mediated suppression of protein degradation including tau	AD	Yes	(Kiprowska et al., 2017; Lee et al., 2010; Min et al., 2017)
10	Mitoxantrone (4212)	DUB inhibitor	USP11	Target ubiquitin-specific peptidase 11 and acts as immunosuppressive drug protective in multiple sclerosis	MS	No, but may cross the disrupted BBB in MS	(Cocco and Marrosu, 2014; Li et al., 2012b)
11	NC-043 (CAS#1053172-87-4)	DUB inhibitor	USP30	Antagonizes Parkin-dependent mitophagy in cells	PD	Unknown	(Yue et al., 2014)
12	Nicotine (89,594)	E2, Proteasome, DUB stimulator	UBE2J2, Proteasome subunits, USP16, USP34	Stimulates nicotinic acetylcholine receptors with alterations in UbE2s, proteasomal activity and DUBs-USP16, USP34 reduce disease progression, improved symptoms and/or decreased L-dopa-induced dyskinesias	PD	Yes	(Hawkins et al., 2004; Quik et al., 2015)
13	PD169316 (4712)	Proteasome activator	26S Proteasome	Trigger proteasome for PROTAC and ubiquitin dependent α-synuclein degradation by p38 MAPK inhibition	PD	Unknown	(Leestemaker et al., 2017)
14	PDTC (Pyrrolidine dithiocarbamate) (65,351)	Proteasome inhibitor	Immunoproteasome	Inhibit immunoproteasome and play protective role by acting as NF-κB inhibitor, antioxidant, Akt and Nrf2 pathway inducer that improve spacial learning in AD and motor functions in PD	AD, PD	Yes	(Abd-El-Fattah et al., 2014; Chabicovsky et al., 2010; Kan et al., 2016; Malm et al., 2007)
15	Pimozide (16,362)	DUB inhibitor	USP1	Role in treatment of psychotic disorders and stabilize neuro-muscular transmission and survival in the disease	ALS, Schizophrenia	Yes	(Lee et al., 2016; Patten et al., 2017)
16	Rasagiline derivative M-30 (6,918,822) and Iron chelator VK-28	Proteasome stimulator	–	Improved lactacystin-induced DA neuron loss, proteasomal inhibition, iron accumulation, and microglial activation in Substantia nigra	AD	Yes	(Weinreb et al., 2009; Zhu et al., 2007)
17	Resveratrol (445,154)	E3 and DUB stimulator	MID1, HRD1, USP10	Induces PP2A activity by targeting MID1 to reduce tau phosphorylation. Stimulate Aβ degradation and enhance USP10-mediated ubiquitylation of p53 to induce apoptosis	AD	Yes	(Andrade et al., 2018; Schweiger et al., 2017; Yan et al., 2016)

(*continued on next page*)

Table 3 (*continued*)

S.No.	Compounds (PubChem CID)	Impact on UPS	Target UPS Component	Molecular Action	Target Pathology	BBB permeability	References
18	Rolipram (5092)	Proteasome stimulator	26S Proteasome	Enhance proteasomal degradation of protein aggregates in cAMP/PKA pathway dependent manner	AD	Yes	(Liu et al., 2012; Lokireddy et al., 2015; Myeku et al., 2016)
19	Sildenafil (135,398,744)	Proteasome stimulator	26S proteasome	Act as phosphodiesterases 5 inhibitor and stimulate proteasome to reduce protein aggregate	AD and Proteotoxic diseases	Yes	(Gómez-Vallejo et al., 2016; Myeku and Duff, 2018)
20	Sulforaphane (5350)	Proteasome stimulator	26S Proteasome	Induce Nrf2 mediated proteasomal activity and mHtt degradation and inhibit rotenone-induced locomotor deficiency and dopaminergic neuronal loss. Also, up-regulate p75 neurotrophin receptors to protect against Aβ burden	AD, PD, HD	Yes	(Jazwa et al., 2011; Liu et al., 2014; Zhang et al., 2017; Zhou et al., 2016)
21	Trehalose (7427)	E3 stimulator	CHIP	Bind to expanded polyglutamine and stabilize partially unfolded polyglutamine-containing protein and increase CHIP expression to protect against polyglutamine mediated toxicity	HD	Yes	(Allavena et al., 2019; Casarejos et al., 2014; Fernandez-Estevez et al., 2014)

AD- Alzheimer's disease; PD- Parkinson's disease; HD- Huntington's disease; ALS- Amyotrophic lateral sclerosis; MS- Multiple sclerosis; BBB- Blood Brain Barrier.

atrophy with prominent visual impairments. It primarily affects the nerves in the brain and spinal cord resulting in a range of physical, mental and psychiatric problems. Earlier diagnosis reported human leukocyte antigen as first genetic factor while later several other causes have been identified such as chronic oxidative injury, mitochondrial damage, and limited axonal transport with affected protein homeostasis (Mahad et al., 2015). However, numerous studies have reported impaired activity and composition of UPS, but exact mechanism of UPS associated MS is still remained elusive. For instance, abnormally ubiquitinated axons are reported in normally myelinated white matter in the MS brain (Giordana et al., 2002). Moreover, UPS has also found to trigger inflammatory processes in MS that was abrogated upon interferon-beta-1b treatment (Minagar et al., 2012). Another study evidenced the ubiquitination mediated pathological immunity triggered by Th17 cells via targeting its activator RORγt (retinoic acid-related orphan receptor-γt) and co-activator SRC1 (steroid receptor co-activator 1) for proteasomal degradation (He et al., 2016). Another UPS component USP18, a deubiquitinating enzyme was deficient and unresponsive to therapeutic IFNβ treatment of diseased conditions (Malhotra et al., 2013). Similarly, another ubiquitylase A20 has been implicated in the regulation of inflammation and cell death in multiple sclerosis (Wertz et al., 2015). Interestingly, a recent study has demonstrated ubiquitination independent 26S proteasomal degradation of an auto antigen, myelin basic protein (MBP) in MS (Belogurov et al., 2014). Supporting this notion, a pleiotropic cytokine osteopontin found to be processed by extracellular proteasomes to form chemotactic fragments and subsequent inflammatory responses in MS (Dianzani et al., 2017). Furthermore, impaired elimination of harmful autoimmune cells in MS is attributed to affected UPS components. For instance, pro and anti-apoptotic protein BAK and MCL1 get repressed due to altered expressions of its associated E3 ligases- HUWE1 and β-TrCP in MS patients (Mandel et al., 2012). Likewise, other E3 ligases CBL-b and ITCH associated with T-cell activation were considerably diminished in MS brains (Sellebjerg et al., 2012). Additionally, UPS has a much more complicated role in MS pathogenesis governed by combinatorial effect of both inflammatory and neurodegenerative pathways (Fig. 3E). Therefore, further identification of UPS components is required to elucidate their molecular pathogens to devise novel therapies. Moreover, the specific action of UPS enzymes regulating clearance of toxic metabolites in AD, PD, HD, ALS, and MS has been summarized in Table 2.

4. Therapeutic implications of ubiquitination system for combating neurodegenerative pathologies

Recent advancements in the pharmacological interventions pertaining ubiquitin proteasome system has uncovered prospective treatment avenues for neurodegenerative diseases. The dynamic regulation of UPS components is crucial for achieving toxic metabolite clearance in neurons. For instance, mostly alteration in the activity of many E3 ligases- Parkin, CHIP, TRAF6, HRD1; E2 conjugating enzymes- UBE2D, UBE2J, UBE2K; and DUBs- USP14, USP18, and UCH-L1 have reported to trigger diverse neurological pathologies, including synaptic dysfunction, motor impairment, mitochondrial dysfunction, ER and oxidative stress in NDDs (Zheng et al., 2016). However, there is growing evidence that identified many active forms of (de)ubiquitinases- RNF182, NEDD4-1, UBE3A, SIAH1, UBE2R1 involved in the progression of neurodegenerative pathologies (Bae et al., 2006; Fiesel et al., 2014; Liu et al., 2008; Rodrigues et al., 2016; Singh et al., 2017). Therefore, the following therapeutic strategies can be adopted to alleviate the nuisance caused by UPS dysregulation (Fig. 4).

4.1. Current therapeutic strategies to target UPS mediated pathologies

One of the strategies could be *(i) the allosteric modification of the substrate binding region of ubiquitination enzymes* with the help of small

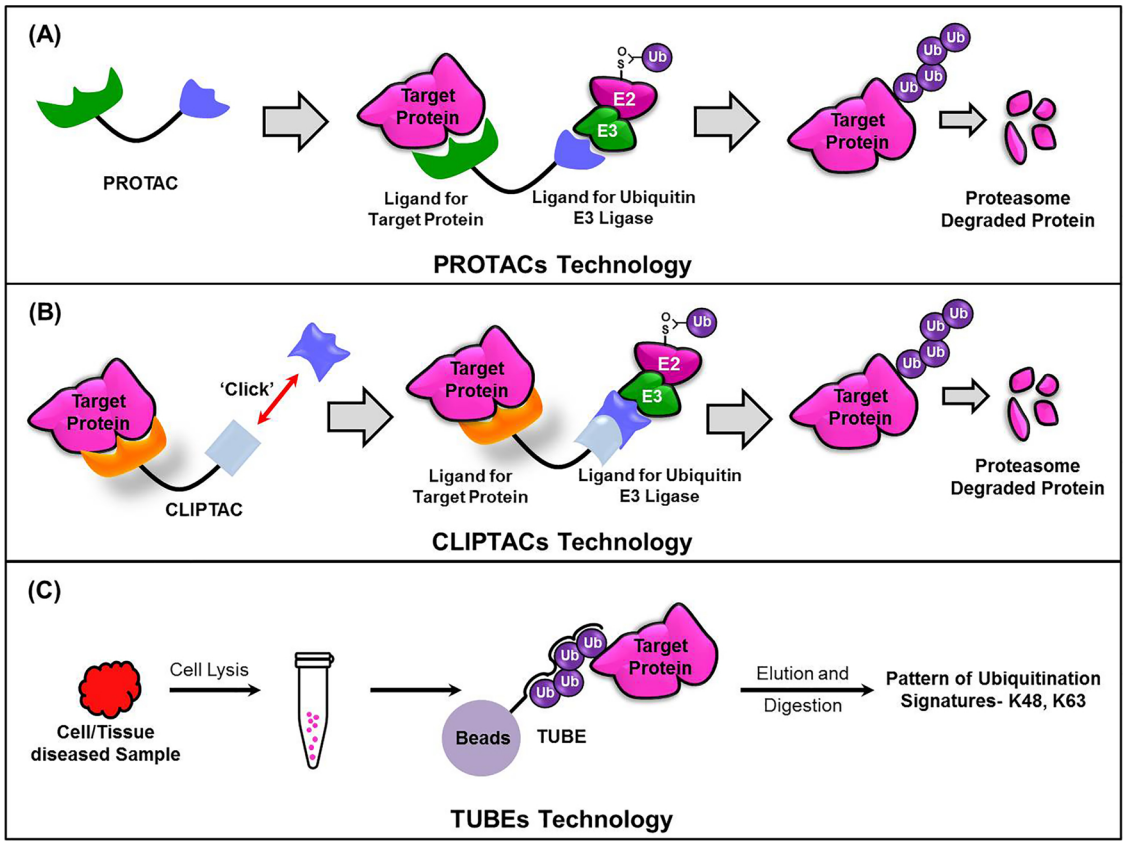


Fig. 5. Recent technologies for targeted protein degradation and ubiquitination research: (A) PROTACs Technology- This technique employs proteolysis targeting chimeras containing ligands for both target protein and E3 ligase to recruit E3 ligase to the target protein for its proteasomal degradation. (B) CLIPTACs Technology- This technique employs two-cell permeable ligands, which undergo a cycloaddition reaction to form a "In-cell click-formed proteolysis targeting chimeras" with ligands for both target protein and E3 ligase and proteasomally degrades the target protein. (C) TUBEs Technology- This technique employs the high-affinity poly-ubiquitin binding matrix, i.e. Tandem Ubiquitin Binding Entities to identify unique patterns of ubiquitination using K48 and K63 selective TUBEs.

molecules, which would enable us to increase or decrease the specificity towards their substrate to regulate protein degradation (Paiva et al., 2018). Such allosteric drugs include arginine and proline rich peptides, chloroquine and its derivatives, clioquinol, dicarboxylic acid compound (SCF-I2), and pentanoic acid compound (CC0651) that showed promising therapy against various diseases and can be studied for their efficacies in NDDs (Ceccarelli et al., 2011; Gizyńska et al., 2019; Khan et al., 2017; Kim et al., 2016; Nalbandian et al., 2015; Orlicky et al., 2010). Similarly, another way could be *(ii) the administration of the deubiquitinases and DUB regulator,* which can enhance the deubiquitination of mutant poly ubiquitin (UBB^{+1}) or normal polyUb chains to release proteasome burden and regulate synaptic functions. For instance, a DUB called UCHL1 is well identified for their roles in synaptic plasticity and prevention against neurodegeneration (Cartier et al., 2009). Another DUB, USP14 has implicated in the regulation of hippocampal short-term synaptic-plasticity (Walters et al., 2014). Recently, a study demonstrated the importance of WD repeat proteins WDR20, WDR40, and WDR48 in regulating the expression and catalytic activity of DUBs for controlling the synaptic plasticity and thus neuronal health (Hodul et al., 2017). Apart from the allosteric drugs, other ways could be *(iii) the post-transnational modifications of the UPS components,* including phosphorylation, acetylation, SUMOylation, and neddylation to modulate their activities. To illustrate, phosphorylation of Parkin (an E3 ligase) by PINK1 boosted its ligase activity, and increases unfolded protein response in stressed neurons (Kane et al., 2014). Likewise, Neddylation, i.e. Nedd8 (an Ubl) mediated modification of Cullin-1 enhanced the activity of a ubiquitin E3 ligase SCF^{skp2} towards the proteasomal degradation of p27 (Duda et al., 2008; Bornstein et al., 2006). Similarly, acetylation, SUMOylation and phosphorylation of MDM-2 augmented p53 ubiquitination (Miyauchi et al., 2002; Nihira et al., 2017). Moreover, *(iv) Selective target based specific ubiquitin-like (UBL) modifications* can aid in the therapeutics of neurodegenerative disorders; Since, Ubl modifications such as ISGlation, FATtylation of pathogenic proteins has shown interesting results in toxic protein clearance. For instance, FAT10 (an Ubl) is identified to bind and modulate the solubility of poly glutamine proteins revealing it as a therapeutic agent for Huntington's disease (Nagashima et al., 2011). Other Ubls like ATG8 and ISG15 have reported to facilitate the autophagic clearance of toxic proteins in association with p62/SQSTM1 binding and IFN-γ induction respectively (Nakashima et al., 2015; Pankiv et al., 2007; Lu et al., 2014, 2017). Likewise, another potential approach could be *(v) the activation of proteasome with suitable drugs,* which may be achieved through any of the following ways. (a) Up-regulation of 19S and 20S complex assembly, (b) eliciting the catalytic activity of 20S core proteasomal unit, and (c) triggering unfolding of aggregated proteins by modulating chaperonic activity of either chaperonins or ATPase subunit of 19S complex (Leestemaker et al., 2017). For example, a study has demonstrated to restrain Aβ levels upon resveratrol treatment by stimulating its proteasome-mediated degradation (Li et al., 2012a). Another study reported a small-molecule bicuculline stimulating calcium/calmodulin-dependent protein kinase II (CaMKII) mediated proteasomal activity to modulate synaptic connections in neurons (Djakovic et al., 2009). Likewise, another molecule IU (1-[1-(4-fluorophenyl)-2, 5-dimethylpyrrol-3-yl]-2-pyrrolidin-1-ylethanone) boosted proteasome activity by ameliorating USP14 mediated suppression of ubiquitinated conjugate's degradation (Lee et al., 2010). In addition, a natural occurring compound Sulforaphane is reported to induce proteasomal activity via triggering a transcriptional factor Nrf2

to govern mHtt degradation (Liu et al., 2014). Likewise, Forskolin, Rolipram and other cyclic AMP (cAMP) inducing compounds have shown enhanced proteasomal degradation of protein aggregates in cAMP/PKA pathway dependent manner (Lokireddy et al., 2015; Myeku et al., 2016). Recently, another small-molecule PD169316 has been found to trigger proteasomal activation and thus PROTAC and ubiquitin dependent α-synuclein degradation via hampering p38 MAPK pathway (Leestemaker et al., 2017). There is growing evidence of potential drugs and leading compound discoveries that have shown promising results in targeting the progression of various neurodegenerative diseases, including AD, PD, HD, ALS, and MS (Table 3). In spite of the advancements in drug discovery, the designing of substrate-specific drugs against diverse (de)ubiquitinases remained a challenging task due to structural similarities in their domains. In order to solve this problem a recent technology of Proteolysis-Targeting Chimera (PROTAC™) has evolved, which has been taken up by big pharmaceutical companies to target different cancers and neurodegenerative disorders.

4.2. Clinical applications of chimera construct to target neurodegenerative diseases

The selective protein degradation achieved by proteolysis targeting chimeras (PROTACs) has revolutionized the therapeutic avenues towards neurodegenerative diseases like AD and PD. The current PROTAC technology employed the use of hetero bifunctional molecules, including a ligand for an intracellular target protein and a recruiting group for a ubiquitin E3 ligase linked by a linker (Lai and Crews, 2017). Target protein degradation is initiated when PROTACs stimulate the target protein and ubiquitin E3 ligase to form a ternary complex. Then, these chimeras facilitate the ubiquitination of target proteins by specific E3 ligases, when ubiquitination machinery comes in close proximity followed by its recognition and degradation by the 26*S* proteasome in a eukaryotic cell (Fig. 5A). The specialties of these PROTACs are their ability to cross blood-brain barrier and their routes of administration, including oral, intravenous, subcutaneous, and intrathecal. Moreover, Arvinas Biopharmaceutical Company has explored the potential of PROTACs technology for targeting neurodegenerative diseases like Alzheimer's and Parkinson's disease. For instance, they demonstrated the 50 % reduction of tau proteins in the hippocampus upon direct injection of tau-directed PROTAC protein degraders. Further, they reported tau- and α-synuclein targeted PROTAC protein degraders to cross the blood-brain barriers *in-vivo,* signifying their therapeutic potential against AD and PD (http://arvinas.com/therapeutic-programs/next-generation-targets/). Another group has developed PROTACs that target polypeptide types and degrade the Tau protein. Moreover, they have demonstrated TH006 (a PROTAC) to regulate the Tau level in the brain of an AD mouse model (Chu et al., 2016). Interestingly, Jiang group has also developed a peptide PROTAC by recruiting Keap1-Cul3 ubiquitin E3 ligase and knockdown Tau by the UPS (Lu et al., 2018). Silva et al., has also demonstrated QC-01-175 effected degradation of aberrant Tau utilizing $CRL4^{CRBN}$ E3-ligase in frontotemporal dementia patient-derived neuronal cell models (Silva et al., 2019). Different pharmaceutical companies, including Genentech, Pfizer and distinct academicians have shown interest in this area and made collaborations with Arvinas. Other industries like Astex pharmaceuticals are working on the CLIPTACs protein degradation technology, i.e. in-cell click-formed proteolysis targeting chimeras (Fig. 5B). This technique employed two-cell permeable ligands that undergo a cycloaddition reaction to form a CLIPTAC, i.e. click-formed PROTAC inside the cell and thus facilitate ubiquitination and subsequent proteasomal degradation of target proteins, For successful application of CLIPTAC, we should ensure the administration of two distinct chemical entities separated in time to prevent PROTAC formation outside the cell. For instance, a CLIPTAC is developed with help of an eight-poly-D-arginine tag to improve the cell permeability of artificial ligand AP21998 and short peptide ALAPYIP to recruit mutant (F36V) FKBP12 and the E3 ubiquitin ligase CRL2 VHL complex for its proteasomal degradation (Schneekloth et al., 2004). Further, a tetrazine tagged thalidomide derivative has been designed that rapidly reacts with a trans-cyclo-octene tagged ligand of the target protein in cells to form a cereblon E3 ligase recruiting PROTAC molecule to target BRD4 and ERK1/2 for proteasomal degradation (Lebraud et al., 2016). Recently, Wang et al., has designed a CRBN-recruiting Brd4-PROTAC through multiple optimization cycles and achieved >90 % tumor regression in the corresponding xenograft model (Zhou et al., 2018). Another structure-guided design of [1,4] oxazepines (QCA570) as potent bromo domain and extra-terminal (BET) degrader is developed to restrict cell growth in acute leukemia cell lines (Qin et al., 2018). Likewise, another company, Captor Therapeutics is working to develop a molecular repair platform, i.e. low molecular compounds capable of reconstituting interactions between ubiquitin E3 ligases and their target proteins to employ UPS for target protein degradation. Similarly, other groups at LifeSensors are working to identify the bio markers for PD and other NDDs based on their TUBEs technology, i.e. pulling of significant amounts of ubiquitinated proteins from patient samples (Fig. 5C). The TUBE (Tandem ubiquitin binding entity) technology is based on the principle of identifying tetra-ubiquitin with a markedly higher affinity than single ubiquitin-associated (UBA) domains, thus enabling efficient purification of poly-ubiquitinated proteins from cell extracts in native conditions (Hjerpe et al., 2009). Therefore, this technology would enable us to identify the precise bio-markers for the complex diseases. For instance, Serna et al., has employed TUBEs technology-based microarray analysis to identify the cellular processes regulated by ubiquitination and its associated pathologies with aberrant ubiquitination levels. To achieve this, they used Adriamycin-stimulated MCF7 cells to differentiate protein ubiquitylation levels between cells that are sensitive or resistant to Adriamycin treatment (Serna et al., 2016). Interestingly, Yoshida et al. also employed this technique along with the anti-Lys-ε-Gly-Gly antibodies to identify specific ubiquitin ligase–substrate pairs. They designed trypsin-resistant (TR)-TUBE for expression in cells and co-expressed TR-TUBE with a ubiquitin ligase that stabilized the ubiquitinated substrates by masking the ubiquitin chains and thereby enabled them to identify ubiquitin ligase-substrate pairs (Yoshida et al., 2015). In this manner, the pharmaceutical companies are exploring the starring role of proteolysis in neurodegeneration, which would enable us to develop the potential future therapies avoiding non-specific side effects with enhanced drug efficacies against neurodegenerative disorders.

5. Conclusions

Although ubiquitination system has been extensively studied for the protein clearance in neurodegenerative proteinopathies over last few decades, but ubiquitination markers, sites and mechanisms remained largely unknown. The dynamic regulation of polyubiquitination system is quite intriguing at different physiological and/or pathologic conditions. Herein, we surveyed the role of ubiquitination in learning/memory processing, signaling and brain homeostasis. Moreover, it studied UPS dysregulation mechanisms in the pathophysiology of several major neurodegenerative diseases and their therapeutic regulation strategies. However, more research is required to formulate substrate-specific drugs using proteolysis-targeting chimera products as future therapy for successful treatment of neurodegenerative pathologies.

Declaration of competing interests

There is no conflict or competing interest declared by the authors.

Acknowledgments

We would like to thanks the senior management of Delhi Technological University for constant support and encouragement.

References

Abd-El-Fattah, M.A., Abdelakader, N.F., Zaki, H.F., 2014. Pyrrolidine dithiocarbamate protects against scopolamine-induced cognitive impairment in rats. Eur. J. Pharmacol. 723, 330–338. https://doi.org/10.1016/j.ejphar.2013.11.008.

Abraham, W.C., Jones, O.D., Glanzman, D.L., 2019. Is plasticity of synapses the mechanism of long-term memory storage? NPJ Sci. Learn. 4, 9. https://doi.org/10.1038/s41539-019-0048-y.

Abumrad, N.A., Moore, D.J., 2011. Parkin reinvents itself to regulate fatty acid metabolism by tagging CD36. J. Clin. Invest. 121, 3389–3392. https://doi.org/10.1172/JCI59219.

Alexopoulou, Z., Lang, J., Perrett, R.M., Elschami, M., Hurry, M.E.D., Kim, H.T., Mazaraki, D., Szabo, A., Kessler, B.M., Goldberg, A.L., Ansorge, O., Fulga, T.A., Tofaris, G.K., 2016. Deubiquitinase Usp8 regulates α-synuclein clearance and modifies its toxicity in Lewy body disease. Proc. Natl. Acad. Sci. U. S. A. 113, E4688–E4697. https://doi.org/10.1073/pnas.1523597113.

Alkalay, I., Yaron, A., Hatzubai, A., Orian, A., Ciechanover, A., Ben-Neriah, Y., 1995. Stimulation-dependent IκBα phosphorylation marks the NF-κb inhibitor for degradation via the ubiquitin-proteasome pathway. Proc. Natl. Acad. Sci. U. S. A. 92, 10599–10603. https://doi.org/10.1073/pnas.92.23.10599.

Allavena, G., Del Bello, B., Tini, P., Volpi, N., Valacchi, G., Miracco, C., Pirtoli, L., Maellaro, E., 2019. Trehalose inhibits cell proliferation and amplifies long-term temozolomide- and radiation-induced cytotoxicity in melanoma cells: a role for autophagy and premature senescence. J. Cell. Physiol. 234, 11708–11721. https://doi.org/10.1002/jcp.27838.

Allen, E., Ding, J., Wang, W., Pramanik, S., Chou, J., Yau, V., Yang, Y., 2005. Gigaxonin-controlled degradation of MAP1B light chain is critical to neuronal survival. Nature 438, 224–228. https://doi.org/10.1038/nature04256.

Andrade, S., Ramalho, M.J., Pereira, M.D.C., Loureiro, J.A., 2018. Resveratrol brain delivery for neurological disorders prevention and treatment. Front. Pharmacol. 9, 1261. https://doi.org/10.3389/fphar.2018.01261.

Ang, X.L., Seeburg, D.P., Sheng, M., Harper, J.W., 2008. Regulation of postsynaptic RapGAP SPAR by polo-like kinase 2 and the SCFβ-TRCP ubiquitin ligase in hippocampal neurons. J. Biol. Chem. 283, 29424–29432. https://doi.org/10.1074/jbc.M802475200.

Araki, K., Nagata, K., 2011. Protein folding and quality control in the ER. Cold Spring Harb. Perspect. Biol. 3, a007526. https://doi.org/10.1101/cshperspect.a007526.

Atkin, G., Hunt, J., Minakawa, E., Sharkey, L., Tipper, N., Tennant, W., Paulson, H.L., 2014. F-box only protein 2 (Fbxo2) regulates amyloid precursor protein levels and processing. J. Biol. Chem. 289, 7038–7048. https://doi.org/10.1074/jbc.M113.515056.

Bae, B.Il, Hara, M.R., Cascio, M.B., Wellington, C.L., Hayden, M.R., Ross, C.A., Ha, H.C., Li, X.J., Snyder, S.H., Sawa, A., 2006. Mutant Huntingtin: nuclear translocation and cytotoxicity mediated by GAPDH. Proc. Natl. Acad. Sci. U. S. A. 103, 3405–3409. https://doi.org/10.1073/pnas.0511316103.

Balducci, C., Forloni, G., 2019. Doxycycline for Alzheimer's disease: fighting β-amyloid oligomers and neuroinflammation. Front. Pharmacol. 10, 738. https://doi.org/10.3389/fphar.2019.00738.

Banerjee, S., Neveu, P., Kosik, K.S., 2009. A coordinated local translational control point at the synapse involving relief from silencing and MOV10 degradation. Neuron 64, 871–884. https://doi.org/10.1016/j.neuron.2009.11.023.

Belogurov, A., Kudriaeva, A., Kuzina, E., Smirnov, I., Bobik, T., Ponomarenko, N., Kravtsova-Ivantsiv, Y., Ciechanover, A., Gabibov, A., 2014. Multiple sclerosis autoantigen myelin basic protein escapes control by ubiquitination during proteasomal degradation. J. Biol. Chem. 289, 17758–17766. https://doi.org/10.1074/jbc.M113.544247.

Benvegnù, S., Mateo, M.I., Palomer, E., Jurado-Arjona, J., Dotti, C.G., 2017a. Aging triggers cytoplasmic depletion and nuclear translocation of the E3 ligase mahogunin: a function for ubiquitin in neuronal survival. Mol. Cell 66, 358–372.e7. https://doi.org/10.1016/j.molcel.2017.04.005.

Benvegnù, S., Wahle, T., Dotti, C.G., 2017b. E3 ligase mahogunin (MGRN1) influences amyloid precursor protein maturation and secretion. Oncotarget 8, 89439–89450. https://doi.org/10.18632/oncotarget.20143.

Bhat, K.P., Yan, S., Wang, C.E., Li, S., Li, X.J., 2014. Differential ubiquitination and degradation of huntingtin fragments modulated by ubiquitin-protein ligase E3A. Proc. Natl. Acad. Sci. U. S. A. 111, 5706–5711. https://doi.org/10.1073/pnas.1402215111.

Bhutani, S., Das, A., Maheshwari, M., Lakhotia, S.C., Jana, N.R., 2012. Dysregulation of core components of SCF complex in poly-glutamine disorders. Cell Death Dis. 3, e428. https://doi.org/10.1038/cddis.2012.166.

Bingol, B., Sheng, M., 2011. Deconstruction for reconstruction: the role of proteolysis in neural plasticity and disease. Neuron 69, 22–32. https://doi.org/10.1016/j.neuron.2010.11.006.

Bingol, B., Wang, C.F., Arnott, D., Cheng, D., Peng, J., Sheng, M., 2010. Autophosphorylated CaMKIIα acts as a scaffold to recruit proteasomes to dendritic spines. Cell 140, 567–578. https://doi.org/10.1016/j.cell.2010.01.024.

Bornstein, G., Ganoth, D., Hershko, A., 2006. Regulation of neddylation and deneddylation of cullin1 in SCF Skp2 ubiquitin ligase by F-box protein and substrate. Proc. Natl. Acad. Sci. U. S. A. 103, 11515–11520. https://doi.org/10.1073/pnas.0603921103.

Boselli, M., Lee, B.H., Robert, J., Prado, M.A., Min, S.W., Cheng, C., Catarina Silva, M., Seong, C., Elsasser, S., Hatle, K.M., Gahman, T.C., Gygi, S.P., Haggarty, S.J., Gan, L., King, R.W., Finley, D., 2017. An inhibitor of the proteasomal deubiquitinating enzyme USP14 induces tau elimination in cultured neurons. J. Biol. Chem. 292, 19209–19225. https://doi.org/10.1074/jbc.M117.815126.

Boyer, N.P., Monkiewicz, C., Menon, S., Moy, S.S., Gupton, S.L., 2018. Mammalian TRIM67 functions in brain development and behavior. eNeuro 5https://doi.org/10.1523/ENEURO.0186-18.2018. ENEURO.0186-18.2018.

Burk, K., Pasterkamp, R.J., 2019. Disrupted neuronal trafficking in amyotrophic lateral sclerosis. Acta Neuropathol. 137, 859–877. https://doi.org/10.1007/s00401-019-01964-7.

Burr, M.L., Cano, F., Svobodova, S., Boyle, L.H., Boname, J.M., Lehner, P.J., 2011. HRD1 and UBE2J1 target misfolded MHC class I heavy chains for endoplasmic reticulum-associated degradation. Proc. Natl. Acad. Sci. U. S. A. 8, 2034–2039. https://doi.org/10.1073/pnas.1016229108.

Bustamante, H.A., González, A.E., Cerda-Troncoso, C., Shaughnessy, R., Otth, C., Soza, A., Burgos, P.V., 2018. Interplay between the autophagy-lysosomal pathway and the ubiquitin-proteasome system: a target for therapeutic development in Alzheimer's disease. Front. Cell. Neurosci. 12, 126. https://doi.org/10.3389/fncel.2018.00126.

Callis, J., 2014. The ubiquitination machinery of the ubiquitin system. Arab. B 12, e0174. https://doi.org/10.1199/tab.0174.

Camera, D., Coleman, H.A., Parkington, H.C., Jenkins, T.A., Pow, D.V., Boase, N., Kumar, S., Poronnik, P., 2016. Learning, memory and long-term potentiation are altered in Nedd4 heterozygous mice. Behav. Brain Res. 303, 176–181. https://doi.org/10.1016/j.bbr.2016.01.054.

Caroni, P., Donato, F., Muller, D., 2012. Structural plasticity upon learning: regulation and functions. Nat. Rev. Neurosci. 13, 478–490. https://doi.org/10.1038/nrn3258.

Carroll, C.B., Zeissler, M.L., Hanemann, C.O., Zajicek, J.P., 2012. Δ9-tetrahydrocannabinol (Δ9-THC) exerts a direct neuroprotective effect in a human cell culture model of Parkinson's disease. Neuropathol. Appl. Neurobiol. 38, 535–547. https://doi.org/10.1111/j.1365-2990.2011.01248.x.

Cartier, A.E., Djakovic, S.N., Salehi, A., Wilson, S.M., Masliah, E., Patrick, G.N., 2009. Regulation of synaptic structure by ubiquitin C-terminal hydrolase L1. J. Neurosci. 29, 7857–7868. https://doi.org/10.1523/JNEUROSCI.1817-09.2009.

Cartier, A.E., Ubhi, K., Spencer, B., Vazquez-Roque, R.A., Kosberg, K.A., Fourgeaud, L., Kanayson, P., Patrick, C., Rockenstein, E., Patrick, G.N., Masliah, E., 2012. Differential effects of UCHL1 modulation on alpha-synuclein in PD-like models of alpha-synucleinopathy. PLoS One 7, e34713. https://doi.org/10.1371/journal.pone.0034713.

Casarejos, M.J., Perucho, J., López-Sendón, J.L., García De Yébenes, J., Bettencourt, C., Gómez, A., Ruiz, C., Heutink, P., Rizzu, P., Mena, M.A., 2014. Trehalose improves human fibroblast deficits in a new CHIP-mutation related ataxia. PLoS One 9, e106931. https://doi.org/10.1371/journal.pone.0106931.

Castegna, A., Aksenov, M., Aksenova, M., Thongboonkerd, V., Klein, J.B., Pierce, W.M., Booze, R., Markesbery, W.R., Butterfield, D.A., 2002. Proteomic identification of oxidatively modified proteins in Alzheimer's disease brain. Part I: Creatine kinase BB, glutamine synthase, and ubiquitin carboxy-terminal hydrolase L-1. Free Radic. Biol. Med. 33, 562–571. https://doi.org/10.1016/S0891-5849(02)00914-0.

Ceccarelli, D.F., Tang, X., Pelletier, B., Orlicky, S., Xie, W., Plantevin, V., Neculai, D., Chou, Y.C., Ogunjimi, A., Al-Hakim, A., Varelas, X., Koszela, J., Wasney, G.A., Vedadi, M., Dhe-Paganon, S., Cox, S., Xu, S., Lopez-Girona, A., Mercurio, F., Wrana, J., Durocher, D., Meloche, S., Webb, D.R., Tyers, M., Sicheri, F., 2011. An allosteric inhibitor of the human Cdc34 ubiquitin-conjugating enzyme. Cell 145, 1075–1087. https://doi.org/10.1016/j.cell.2011.05.039.

Cenini, G., Lloret, A., Cascella, R., 2019. Oxidative stress in neurodegenerative diseases: from a mitochondrial point of view. Oxid. Med. Cell. Longev. 2019, 2105607. https://doi.org/10.1155/2019/2105607.

Chabicovsky, M., Prieschl-Grassauer, E., Seipelt, J., Muster, T., Szolar, O.H.J., Hebar, A., Doblhoff-Dier, O., 2010. Pre-clinical safety evaluation of pyrrolidine dithiocarbamate. Basic Clin. Pharmacol. Toxicol. 107, 758–767. https://doi.org/10.1111/j.1742-7843.2010.00573.x.

Chang, R., Liu, X., Li, S., Li, X.J., 2015. Transgenic animal models for study of the pathogenesis of Huntington's disease and therapy. Drug Des. Devel. Ther. 9, 2179–2188. https://doi.org/10.2147/DDDT.S58470.

Changjiang, L., Cheng, L., Haitao, W., Peijie, H., Zhang, Y., Yang, Y., Chen, J., Chen, M., 2018. Activation of the KEAP1-NRF2-ARE signaling pathway reduces oxidative stress in Hep2 cells. Mol. Med. Rep. 18, 2541–2550. https://doi.org/10.3892/mmr.2018.9288.

Chhangani, D., Endo, F., Amanullah, A., Upadhyay, A., Watanabe, S., Mishra, R., Yamanaka, K., Mishra, A., 2016. Mahogunin ring finger 1 confers cytoprotection against mutant SOD1 aggresomes and is defective in an ALS mouse model. Neurobiol. Dis. 86, 16–28. https://doi.org/10.1016/j.nbd.2015.11.017.

Chin, L.S., Li, L., 2016. Ubiquitin phosphorylation in Parkinson's disease: implications for pathogenesis and treatment. Transl. Neurodegener. 5, 1. https://doi.org/10.1186/s40035-015-0049-6.

Chin, L.S., Vavalle, J.P., Li, A., 2002. Staring, a novel E3 ubiquitin-protein ligase that targets syntaxin 1 for degradation. J. Biol. Chem. 277, 35071–35079. https://doi.org/10.1074/jbc.M203300200.

Choi, J.S., Lee, D.H., 2010. CHIP promotes the degradation of mutant SOD1 by reducing its interaction with VCP and S6/S6' subunits of 26S proteasome. Anim. Cells Syst. (Seoul) 14, 1–10. https://doi.org/10.1080/19768351003765145.

Choi, J., Levey, A.I., Weintraub, S.T., Rees, H.D., Gearing, M., Chin, L.S., Li, L., 2004. Oxidative modifications and down-regulation of ubiquitin carboxyl-terminal hydrolase L1 associated with idiopathic Parkinson's and Alzheimer's diseases. J. Biol. Chem. 279, 13256–13264. https://doi.org/10.1074/jbc.M314124200.

Chu, T.T., Gao, N., Li, Q.Q., Chen, P.G., Yang, X.F., Chen, Y.X., Zhao, Y.F., Li, Y.M., 2016. Specific knockdown of endogenous tau protein by peptide-directed ubiquitin-proteasome degradation. Cell Chem. Biol. 23, 453–461. https://doi.org/10.1016/j.chembiol.2016.02.016.

Chung, K.K.K., Thomas, B., Li, X., Pletnikova, O., Troncoso, J.C., Marsh, L., Dawson, V.L., Dawson, T.M., 2004. S-nitrosylation of parkin regulates ubiquitination and compromises parkin's protective function. Science 304, 1328–1331. https://doi.org/10.

1126/science.1093891.
Cocco, E., Marrosu, M.G., 2014. The current role of mitoxantrone in the treatment of multiple sclerosis. Expert Rev. Neurother. 14, 607–616. https://doi.org/10.1586/14737175.2014.915742.
Colledge, M., Snyder, E.M., Crozier, R.A., Soderling, J.A., Jin, Y., Langeberg, L.K., Lu, H., Bear, M.F., Scott, J.D., 2003. Ubiquitination regulates PSD-95 degradation and AMPA receptor surface expression. Neuron 40, 595–607. https://doi.org/10.1016/S0896-6273(03)00687-1.
Collins, P., Mitxitorena, I., Carmody, R., 2016. The ubiquitination of NF-κB subunits in the control of transcription. Cells 5, 23. https://doi.org/10.3390/cells5020023.
D'Souza, M.S., 2015. Glutamatergic transmission in drug reward: implications for drug addiction. Front. Neurosci. 9, 404. https://doi.org/10.3389/fnins.2015.00404.
Das, J., Samadder, A., Das, S., Paul, A., Khuda-Bukhsh, A.R., 2016. Nanopharmaceutical approach for enhanced anti-cancer activity of betulinic acid in lung-cancer treatment via activation of PARP: interaction with DNA as a target anti-cancer potential of nano-betulinic acid in lung cancer. J. Pharmacopuncture 19, 37–44. https://doi.org/10.3831/KPI.2016.19.005.
Dawson, T.M., Dawson, V.L., 2010. The role of parkin in familial and sporadic Parkinson's disease. Mov. Disord. 25, S32–S39. https://doi.org/10.1002/mds.22798.
de Pril, R., Fischer, D.F., Roos, R.A.C., van Leeuwen, F.W., 2007. Ubiquitin-conjugating enzyme E2-25K increases aggregate formation and cell death in polyglutamine diseases. Mol. Cell. Neurosci. 34, 10–19. https://doi.org/10.1016/j.mcn.2006.09.006.
Deture, M.A., Dickson, D.W., 2019. The neuropathological diagnosis of Alzheimer's disease. Mol. Neurodegener. 14, 32. https://doi.org/10.1186/s13024-019-0333-5.
DiAntonio, A., Haghighi, A.P., Portman, S.L., Lee, J.D., Amaranto, A.M., Goodman, C.S., 2001. Ubiquitination-dependent mechanisms regulate synaptic growth and function. Nature 412, 449–452. https://doi.org/10.1038/35086595.
Dianzani, C., Bellavista, E., Liepe, J., Verderio, C., Martucci, M., Santoro, A., Chiocchetti, A., Gigliotti, C.L., Boggio, E., Ferrara, B., Riganti, L., Keller, C., Janek, K., Niewienda, A., Fenoglio, C., Sorosina, M., Cantello, R., Kloetzel, P.M., Stumpf, M.P.H., Paul, F., Ruprecht, K., Galimberti, D., Boneschi, F.M., Comi, C., Dianzani, U., Mishto, M., 2017. Extracellular proteasome-osteopontin circuit regulates cell migration with implications in multiple sclerosis. Sci. Rep. 7, 43718. https://doi.org/10.1038/srep43718.
Díaz-Hernández, M., Valera, A.G., Morán, M.A., Gómez-Ramos, P., Alvarez-Castelao, B., Castaño, J.G., Hernández, F., Lucas, J.J., 2006. Inhibition of 26S proteasome activity by huntingtin filaments but not inclusion bodies isolated from mouse and human brain. J. Neurochem. 98, 1585–1596. https://doi.org/10.1111/j.1471-4159.2006.03968.x.
Djakovic, S.N., Schwarz, L.A., Barylko, B., DeMartino, G.N., Patrick, G.N., 2009. Regulation of the proteasome by neuronal activity and calcium/calmodulin-dependent protein kinase II. J. Biol. Chem. 284, 26655–26665. https://doi.org/10.1074/jbc.M109.021956.
Douglas, P.M., Dillin, A., 2010. Protein homeostasis and aging in neurodegeneration. J. Cell Biol. 190, 719–729. https://doi.org/10.1083/jcb.201005144.
Duda, D.M., Borg, L.A., Scott, D.C., Hunt, H.W., Hammel, M., Schulman, B.A., 2008. Structural insights into NEDD8 activation of cullin-RING ligases: conformational control of conjugation. Cell 134, 995–1006. https://doi.org/10.1016/j.cell.2008.07.022.
Durcan, T.M., Kontogiannea, M., Bedard, N., Wing, S.S., Fon, E.A., 2012. Ataxin-3 deubiquitination is coupled to parkin ubiquitination via E2 ubiquitin-conjugating enzyme. J. Biol. Chem. 287, 531–541. https://doi.org/10.1074/jbc.M111.288449.
Durham-Lee, J.C., Mokkapati, V.U.L., Johnson, K.M., Nesic, O., 2011. Amiloride improves locomotor recovery after spinal cord injury. J. Neurotrauma 28, 1319–1326. https://doi.org/10.1089/neu.2011.1921.
Esteban, J.A., Shi, S.H., Wilson, C., Nuriya, M., Huganir, R.L., Malinow, R., 2003. PKA phosphorylation of AMPA receptor subunits controls synaptic trafficking underlying plasticity. Nat. Neurosci. 6, 136–143. https://doi.org/10.1038/nn997.
Feduccia, A.A., Chatterjee, S., Bartlett, S.E., 2012. Neuronal nicotinic acetylcholine receptors: neuroplastic changes underlying alcohol and nicotine addictions. Front. Mol. Neurosci. 5, 83. https://doi.org/10.3389/fnmol.2012.00083.
Fernandez-Estevez, M.A., Casarejos, M.J., Sendon, J.L., Caldentey, J.G., Ruiz, C., Gomez, A., Perucho, J., De Yebenes, J.G., Mena, M.A., 2014. Trehalose reverses cell malfunction in fibroblasts from normal and Huntington's disease patients caused by proteosome inhibition. PLoS One 9, e90202. https://doi.org/10.1371/journal.pone.0090202.
Ferrara, N.C., Jarome, T.J., Cullen, P.K., Orsi, S.A., Kwapis, J.L., Trask, S., Pullins, S.E., Helmstetter, F.J., 2019. GluR2 endocytosis-dependent protein degradation in the amygdala mediates memory updating. Sci. Rep. 9, 5180. https://doi.org/10.1038/s41598-019-41526-1.
Fiesel, F.C., Moussaud-Lamodière, E.L., Ando, M., Springer, W., 2014. A specific subset of E2 ubiquitin-conjugating enzymes regulate Parkin activation and mitophagy differently. J. Cell Sci. 127, 3488–3504. https://doi.org/10.1242/jcs.147520.
Fonseca, A.C., Oliveira, C.R., Pereira, C.F., Cardoso, S.M., 2014. Loss of proteostasis induced by amyloid beta peptide in brain endothelial cells. Biochim. Biophys. Acta Mol. Cell Res. 1843, 1150–1161. https://doi.org/10.1016/j.bbamcr.2014.02.016.
Freibaum, B.D., Taylor, J.P., 2017. The role of dipeptide repeats in C9ORF72-related ALS-FTD. Front. Mol. Neurosci. 10, 35. https://doi.org/10.3389/fnmol.2017.00035.
Friese, M.A., Craner, M.J., Etzensperger, R., Vergo, S., Wemmie, J.A., Welsh, M.J., Vincent, A., Fugger, L., 2007. Acid-sensing ion channel-1 contributes to axonal degeneration in autoimmune inflammation of the central nervous system. Nat. Med. 13, 1483–1489. https://doi.org/10.1038/nm1668.
Fu, A.K.Y., Hung, K.W., Fu, W.Y., Shen, C., Chen, Y., Xia, J., Lai, K.O., Ip, N.Y., 2011. APCCdh1 mediates EphA4-dependent downregulation of AMPA receptors in homeostatic plasticity. Nat. Neurosci. 14, 181–191. https://doi.org/10.1038/nn.2715.
Fu, H., Hardy, J., Duff, K.E., 2018. Selective vulnerability in neurodegenerative diseases. Nat. Neurosci. 21, 1350–1358. https://doi.org/10.1038/s41593-018-0221-2.
Fuchsberger, T., Lloret, A., Viña, J., 2017. New functions of APC/C ubiquitin ligase in the nervous system and its role in Alzheimer's disease. Int. J. Mol. Sci. 18https://doi.org/10.3390/ijms18051057. pii: E1057.
Gadhave, K., Bolshette, N., Ahire, A., Pardeshi, R., Thakur, K., Trandafir, C., Istrate, A., Ahmed, S., Lahkar, M., Muresanu, D.F., Balea, M., 2016. The ubiquitin proteasomal system: a potential target for the management of Alzheimer's disease. J. Cell. Mol. Med. 20, 1392–1407. https://doi.org/10.1111/jcmm.12817.
Geetha, T., Zheng, C., McGregor, W.C., Douglas White, B., Diaz-Meco, M.T., Moscat, J., Babu, J.R., 2012. TRAF6 and p62 inhibit amyloid β-induced neuronal death through p75 neurotrophin receptor. Neurochem. Int. 61, 1289–1293. https://doi.org/10.1016/j.neuint.2012.09.005.
Geisler, S., Vollmer, S., Golombek, S., Kahle, P.J., 2014. The ubiquitin-conjugating enzymes UBE2N, UBE2L3 and UBE2D2/3 are essential for Parkin-dependent mitophagy. J. Cell Sci. 127, 3280–3293. https://doi.org/10.1242/jcs.146035.
Gerakis, Y., Hetz, C., 2018. Emerging roles of ER stress in the etiology and pathogenesis of Alzheimer's disease. FEBS J. 285, 995–1011. https://doi.org/10.1111/febs.14332.
Gersch, M., Gladkova, C., Schubert, A.F., Michel, M.A., Maslen, S., Komander, D., 2017. Mechanism and regulation of the Lys6-selective deubiquitinase USP30. Nat. Struct. Mol. Biol. 24, 920–930. https://doi.org/10.1038/nsmb.3475.
Gilbert, T.M., Zürcher, N.R., Catanese, M.C., Tseng, C.E.J., Di Biase, M.A., Lyall, A.E., Hightower, B.G., Parmar, A.J., Bhanot, A., Wu, C.J., Hibert, M.L., Kim, M., Mahmood, U., Stufflebeam, S.M., Schroeder, F.A., Wang, C., Roffman, J.L., Holt, D.J., Greve, D.N., Pasternak, O., Kubicki, M., Wey, H.Y., Hooker, J.M., 2019. Neuroepigenetic signatures of age and sex in the living human brain. Nat. Commun. 10, 2945. https://doi.org/10.1038/s41467-019-11031-0.
Giordana, M.T., Richiardi, P., Trevisan, E., Boghi, A., Palmucci, L., 2002. Abnormal ubiquitination of axons in normally myelinated white matter in multiple sclerosis brain. Neuropathol. Appl. Neurobiol. 28, 35–41. https://doi.org/10.1046/j.1365-2990.2002.00372.x.
Gizyńska, M., Witkowska, J., Karpowicz, P., Rostankowski, R., Chocron, E.S., Pickering, A.M., Osmulski, P., Gaczynska, M., Jankowska, E., 2019. Proline- and arginine-rich peptides as flexible allosteric modulators of human proteasome activity. J. Med. Chem. 62, 359–370. https://doi.org/10.1021/acs.jmedchem.8b01025.
Gõmez-Vallejo, V., Ugarte, A., García-Barroso, C., Cuadrado-Tejedor, M., Szczupak, B., Dopeso-Reyes, I.G., Lanciego, J.L., García-Osta, A., Llop, J., Oyarzabal, J., Franco, R., 2016. Pharmacokinetic investigation of sildenafil using positron emission tomography and determination of its effect on cerebrospinal fluid cGMP levels. J. Neurochem. 136, 403–415. https://doi.org/10.1111/jnc.13454.
Gong, B., Cao, Z., Zheng, P., Vitolo, O.V., Liu, S., Staniszewski, A., Moolman, D., Zhang, H., Shelanski, M., Arancio, O., 2006. Ubiquitin hydrolase Uch-L1 rescues β-Amyloid-Induced decreases in synaptic function and contextual memory. Cell 126, 775–788. https://doi.org/10.1016/j.cell.2006.06.046.
Gong, B., Chen, F., Pan, Y., Arrieta-Cruz, I., Yoshida, Y., Haroutunian, V., Pasinetti, G.M., 2010. SCFFbx2-E3-ligase-mediated degradation of BACE1 attenuates Alzheimer's disease amyloidosis and improves synaptic function. Aging Cell 9, 1018–1031. https://doi.org/10.1111/j.1474-9726.2010.00632.x.
González-Lizárraga, F., Socías, S.B., Ávila, C.L., Torres-Bugeau, C.M., Barbosa, L.R.S., Binolfi, A., Sepúlveda-Díaz, J.E., Del-Bel, E., Fernandez, C.O., Papy-Garcia, D., Itri, R., Raisman-Vozari, R., Chehín, R.N., 2017. Repurposing doxycycline for synucleinopathies: remodelling of α-synuclein oligomers towards non-toxic parallel beta-sheet structured species. Sci. Rep. 7, 41755. https://doi.org/10.1038/srep41755.
Guglielmotto, M., Monteleone, D., Vasciaveo, V., Repetto, I.E., Manassero, G., Tabaton, M., Tamagno, E., 2017. The decrease of Uch-L1 activity is a common mechanism responsible for Aβ 42 accumulation in Alzheimer's and vascular disease. Front. Aging Neurosci. 9, 320. https://doi.org/10.3389/fnagi.2017.00320.
Guntupalli, S., Jang, S.E., Zhu, T., Huganir, R.L., Widagdo, J., Anggono, V., 2017. GluA1 subunit ubiquitination mediates amyloid-β-induced loss of surface α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors. J. Biol. Chem. 292, 8186–8194. https://doi.org/10.1074/jbc.M116.774554.
Guo, Y.X., He, L.Y., Zhang, M., Wang, F., Liu, F., Peng, W.X., 2016. 1,25-Dihydroxyvitamin D3 regulates expression of LRP1 and RAGE in vitro and in vivo, enhancing Aβ1-40 brain-to-blood efflux and peripheral uptake transport. Neuroscience 8, 20. https://doi.org/10.1016/j.neuroscience.2016.01.041.
Haddad, D.M., Vilain, S., Vos, M., Esposito, G., Matta, S., Kalscheuer, V.M., Craessaerts, K., Leyssen, M., Nascimento, R.M.P., Vianna-Morgante, A.M., DeStrooper, B., VanEsch, H., Morais, V.A., Verstreken, P., 2013. Mutations in the intellectual disability gene Ube2a cause neuronal dysfunction and impair parkin-dependent mitophagy. Mol. Cell 50, 831–843. https://doi.org/10.1016/j.molcel.2013.04.012.
Hans, F., Fiesel, F.C., Strong, J.C., Jackel, S., Rasse, T.M., Geisler, S., Springer, W., Schulz, J.B., Voigt, A., Kahle, P.J., 2014. UBE2E ubiquitin-conjugating enzymes and ubiquitin isopeptidase y regulate TDP-43 protein ubiquitination. J. Biol. Chem. 289, 19164–19179. https://doi.org/10.1074/jbc.M114.561704.
Harding, R.J., Tong, Y.F., 2018. Proteostasis in Huntington's disease: disease mechanisms and therapeutic opportunities. Acta Pharmacol. Sin. 39, 754–769. https://doi.org/10.1038/aps.2018.11.
Hawkins, B.T., Abbruscato, T.J., Egleton, R.D., Brown, R.C., Huber, J.D., Campos, C.R., Davis, T.P., 2004. Nicotine increases in vivo blood-brain barrier permeability and alters cerebral microvascular tight junction protein distribution. Brain Res. 1027, 48–58. https://doi.org/10.1016/j.brainres.2004.08.043.
He, Z., Wang, F., Ma, J., Sen, S., Zhang, J., Gwack, Y., Zhou, Y., Sun, Z., 2016. Ubiquitination of RORγt at lysine 446 limits Th17 differentiation by controlling coactivator recruitment. J. Immunol. 197, 1148–1158. https://doi.org/10.4049/jimmunol.1600548.
Hebron, M.L., Lonskaya, I., Sharpe, K., Weerasinghe, P.P.K., Algarzae, N.K., Shekoyan,

A.R., Moussa, C.E.H., 2013. Parkin ubiquitinates tar-DNA binding protein-43 (TDP-43) and promotes its cytosolic accumulation via interaction with histone deacetylase 6 (HDAC6). J. Biol. Chem. 288, 4103–4115. https://doi.org/10.1074/jbc.M112.419945.

Heink, S., Ludwig, D., Kloetzel, P.M., Krüger, E., 2005. IFN-γ-induced immune adaptation of the proteasome system is an accelerated and transient response. Proc. Natl. Acad. Sci. U. S. A. 102, 9241–9246. https://doi.org/10.1073/pnas.0501711102.

Hershko, A., 2005. The ubiquitin system for protein degradation and some of its roles in the control of the cell division cycle. Cell Death Differ. 44, 5932–5943. https://doi.org/10.1038/sj.cdd.4401702.

Hjerpe, R., Aillet, F., Lopitz-Otsoa, F., Lang, V., England, P., Rodriguez, M.S., 2009. Efficient protection and isolation of ubiquitylated proteins using tandem ubiquitin-binding entities. EMBO Rep. 10, 1250–1258. https://doi.org/10.1038/embor.2009.192.

Hodul, M., Dahlberg, C.L., Juo, P., 2017. Function of the deubiquitinating enzyme USP46 in the nervous system and its regulation by WD40-Repeat proteins. Front. Synaptic Neurosci. 9, 16. https://doi.org/10.3389/fnsyn.2017.00016.

Holmberg, C.I., Staniszewski, K.E., Mensah, K.N., Matouschek, A., Morimoto, R.I., 2004. Inefficient degradation of truncated polyglutamine proteins by the proteasome. EMBO J. 23, 4307–4318. https://doi.org/10.1038/sj.emboj.7600426.

Hong, L., Huang, H.C., Jiang, Z.F., 2014. Relationship between amyloid-beta and the ubiquitin-proteasome system in Alzheimer's disease. Neurol. Res. 36, 276–282. https://doi.org/10.1179/1743132813Y.0000000288.

Honke, N., Shaabani, N., Zhang, D.E., Hardt, C., Lang, K.S., 2016. Multiple functions of USP18. Cell Death Dis. 7, e2444. https://doi.org/10.1038/cddis.2016.326.

Howard, R.A., Sharma, P., Hajjar, C., Caldwell, K.A., Caldwell, G.A., du Breuil, R., Moore, R., Boyd, L., 2007. Ubiquitin conjugating enzymes participate in polyglutamine protein aggregation. BMC Cell Biol. 8, 32. https://doi.org/10.1186/1471-2121-8-32.

Hsieh, M.C., Ho, Y.C., Lai, C.Y., Chou, D., Chen, G.Den, Lin, T.Bin, Peng, H.Y., 2018. Spinal TNF-α impedes Fbxo45-dependent Munc13-1 ubiquitination to mediate neuropathic allodynia in rats. Cell Death Dis. 9, 811. https://doi.org/10.1038/s41419-018-0859-4.

Huang, L., Ho, P., Chen, C.H., 2007. Activation and inhibition of the proteasome by betulinic acid and its derivatives. FEBS Lett. 581, 4955–4959. https://doi.org/10.1016/j.febslet.2007.09.031.

Hung, A.Y., Sung, C.C., Brito, I.L., Sheng, M., 2010. Degradation of postsynaptic scaffold GKAP and regulation of dendritic spine morphology by the TRIM3 ubiquitin ligase in rat hippocampal neurons. PLoS One 5, e9842. https://doi.org/10.1371/journal.pone.0009842.

Hyrskyluoto, A., Bruelle, C., Lundh, S.H., Do, H.T., Kivinen, J., Rappou, E., Reijonen, S., Waltimo, T., Petersén, Å., Lindholm, D., Korhonen, L., 2014. Ubiquitin-specific protease-14 reduces cellular aggregates and protects against mutant huntingtin-induced cell degeneration: involvement of the proteasome and ER stress-activated kinase IRE1α Hum. Mol. Genet. 23, 5928–5939. https://doi.org/10.1093/hmg/ddu317.

Iwata, A., Nagashima, Y., Matsumoto, L., Suzuki, T., Yamanaka, T., Date, H., Deoka, K., Nukina, N., Tsuji, S., 2009. Intranuclear degradation of polyglutamine aggregates by the ubiquitin-proteasome system. J. Biol. Chem. 284, 9796–9803. https://doi.org/10.1074/jbc.M809739200.

Jacoupy, M., Hamon-Keromen, E., Ordureau, A., Erpapazoglou, Z., Coge, F., Corvol, J.-C., Nosjean, O., Mannoury la Cour, C., Millan, M.J., Boutin, J.A., Harper, J.W., Brice, A., Guedin, D., Gautier, C.A., Corti, O., 2019. The PINK1 kinase-driven ubiquitin ligase Parkin promotes mitochondrial protein import through the presequence pathway in living cells. Sci. Rep. 9, 11829. https://doi.org/10.1038/s41598-019-47352-9.

Jansen, A.H.P., Reits, E.A.J., Hol, E.M., 2014. The ubiquitin proteasome system in glia and its role in neurodegenerative diseases. Front. Mol. Neurosci. 7, 73. https://doi.org/10.3389/fnmol.2014.00073.

Jarome, T.J., Werner, C.T., Kwapis, J.L., Helmstetter, F.J., 2011. Activity dependent protein degradation is critical for the formation and stability of fear memory in the amygdala. PLoS One 6, e24349. https://doi.org/10.1371/journal.pone.0024349.

Jazwa, A., Rojo, A.I., Innamorato, N.G., Hesse, M., Fernández-Ruiz, J., Cuadrado, A., 2011. Pharmacological targeting of the transcription factor NRf2 at the basal ganglia provides disease modifying therapy for experimental Parkinsonism. Antioxid. Redox Signal. 14, 2347–2360. https://doi.org/10.1089/ars.2010.3731.

Jedynak, J., Hearing, M., Ingebretson, A., Ebner, S.R., Kelly, M., Fischer, R.A., Kourrich, S., Thomas, M.J., 2016. Cocaine and amphetamine induce overlapping but distinct patterns of AMPAR plasticity in nucleus accumbens medium spiny neurons. Neuropsychopharmacology 41, 464–476. https://doi.org/10.1038/npp.2015.168.

Ji, C.H., Kwon, Y.T., 2017. Crosstalk and interplay between the ubiquitin-proteasome system and autophagy. Mol. Cells 40, 441–449. https://doi.org/10.14348/molcells.2017.0115.

Jing, X., Sui, W.H., Wang, S., Xu, X.F., Yuan, R.R., Chen, X.R., Ma, H.X., Zhu, Y.X., Sun, J.K., Yi, F., Chen, Z.Y., Wang, Y., 2017. HDAC7 ubiquitination by the E3 Ligase CBX4 is involved in contextual fear conditioning memory formation. J. Neurosci. 37, 3848–3863. https://doi.org/10.1523/JNEUROSCI.2773-16.2017.

Johnston-Carey, H.K., Pomatto, L.C.D., Davies, K.J.A., 2016. The Immunoproteasome in oxidative stress, aging, and disease. Crit. Rev. Biochem. Mol. Biol. 51, 268–281. https://doi.org/10.3109/10409238.2016.1172554.

Joshi, V., Amanullah, A., Upadhyay, A., Mishra, R., Kumar, A., Mishra, A., 2016. A decade of boon or burden: what has the chip ever done for cellular protein quality control mechanism implicated in neurodegeneration and aging? Front. Mol. Neurosci. 9, 93. https://doi.org/10.3389/fnmol.2016.00093.

Juenemann, K., Schipper-Krom, S., Wiemhoefer, A., Kloss, A., Sanz, A.S., Reits, E.A.J., 2013. Expanded polyglutamine-containing N-terminal huntingtin fragments are entirely degraded by mammalian proteasomes. J. Biol. Chem. 288, 27068–27084. https://doi.org/10.1074/jbc.M113.486076.

Juenemann, K., Wiemhoefer, A., Reits, E.A., 2015. Detection of ubiquitinated huntingtin species in intracellular aggregates. Front. Mol. Neurosci. 8, 1. https://doi.org/10.3389/fnmol.2015.00001.

Juenemann, K., Jansen, A.H.P., Van Riel, L., Merkx, R., Mulder, M.P.C., An, H., Statsyuk, A., Kirstein, J., Ovaa, H., Reits, E.A., 2018. Dynamic recruitment of ubiquitin to mutant huntingtin inclusion bodies. Sci. Rep. 8, 1405. https://doi.org/10.1038/s41598-018-19538-0.

Kan, M.H., Yang, T., Fu, H.Q., Fan, L., Wu, Y., Terrando, N., Wang, T.L., 2016. Pyrrolidine dithiocarbamate prevents neuroinflammation and cognitive dysfunction after endotoxemia in rats. Front. Aging Neurosci. 8, 175. https://doi.org/10.3389/fnagi.2016.00175.

Kanarek, N., Ben-Neriah, Y., 2012. Regulation of NF-κB by ubiquitination and degradation of the IκBs. Immunol. Rev. 246, 77–94. https://doi.org/10.1111/j.1600-065X.2012.01098.x.

Kane, L.A., Lazarou, M., Fogel, A.I., Li, Y., Yamano, K., Sarraf, S.A., Banerjee, S., Youle, R.J., 2014. PINK1 phosphorylates ubiquitin to activate parkin E3 ubiquitin ligase activity. J. Cell Biol. 205, 143–153. https://doi.org/10.1083/jcb.201402104.

Kaneko, M., 2016. Physiological roles of ubiquitin ligases related to the endoplasmic reticulum. Yakugaku Zasshi 136, 805–809. https://doi.org/10.1248/yakushi.15-00292-2.

Kato, A., Rouach, N., Nicoll, R.A., Bredt, D.S., 2005. Activity-dependent NMDA receptor degradation mediated by retrotranslocation and ubiquitination. Proc. Natl. Acad. Sci. U. S. A. 102, 5600–5605. https://doi.org/10.1073/pnas.0501769102.

Keller, J.N., Hanni, K.B., Markesbery, W.R., 2000a. Impaired proteasome function in Alzheimer's disease. J. Neurochem. 75, 436–439. https://doi.org/10.1046/j.1471-4159.2000.0750436.x.

Keller, J.N., Hanni, K.B., Markesbery, W.R., 2000b. Possible involvement of proteasome inhibition in aging: implications for oxidative stress. Mech. Ageing Dev. 113, 61–70. https://doi.org/10.1016/S0047-6374(99)00101-3.

Keller, J.N., Huang, F.F., Markesbery, W.R., 2000c. Decreased levels of proteasome activity and proteasome expression in aging spinal cord. Neuroscience 98, 149–156. https://doi.org/10.1016/S0306-4522(00)00067-1.

Khan, N.A., Khan, I., Abid, S.M.A., Zaib, S., Ibrar, A., Andleeb, H., Hameed, S., Iqbal, J., 2017. Quinolinic carboxylic acid derivatives as potential multi-target compounds for neurodegeneration: monoamine oxidase and cholinesterase inhibition. Med. Chem. (Los Angeles) 14, 74–85. https://doi.org/10.2174/1573406413666170525125231.

Kim, C.H., Leblanc, P., Kim, K.S., 2016. 4-amino-7-chloroquinoline derivatives for treating Parkinsons disease: implications for drug discovery. Expert Opin. Drug Discov. 11, 337–341. https://doi.org/10.1517/17460441.2016.1154529.

Kiprowska, M.J., Stepanova, A., Todaro, D.R., Galkin, A., Haas, A., Wilson, S.M., Figueiredo-Pereira, M.E., 2017. Neurotoxic mechanisms by which the USP14 inhibitor IU1 depletes ubiquitinated proteins and Tau in rat cerebral cortical neurons: relevance to Alzheimer's disease. Biochim. Biophys. Acta Mol. Basis Dis. 1863, 1157–1170. https://doi.org/10.1016/j.bbadis.2017.03.017.

Ko, H.S., Bailey, R., Smith, W.W., Liu, Z., Shin, J.H., Lee, Y.Il, Zhang, Y.J., Jiang, H., Ross, C.A., Moore, D.J., Patterson, C., Petrucelli, L., Dawson, T.M., Dawson, V.L., 2009. CHIP regulates leucine-rich repeat kinase-2 ubiquitination, degradation, and toxicity. Proc. Natl. Acad. Sci. U. S. A. 106, 2897–2902. https://doi.org/10.1073/pnas.0810123106.

Kwon, Y.T., Ciechanover, A., 2017. The ubiquitin code in the ubiquitin-proteasome system and autophagy. Trends Biochem. Sci. 42, 873–886. https://doi.org/10.1016/j.tibs.2017.09.002.

Lai, A.C., Crews, C.M., 2017. Induced protein degradation: an emerging drug discovery paradigm. Nat. Rev. Drug Discov. 16, 101–114. https://doi.org/10.1038/nrd.2016.211.

Launay, N., Ruiz, M., Fourcade, S., Schlüter, A., Guilera, C., Ferrer, I., Knecht, E., Pujol, A., 2013. Oxidative stress regulates the ubiquitin-proteasome system and immunoproteasome functioning in a mouse model of X-adrenoleukodystrophy. Brain 136, 891–904. https://doi.org/10.1093/brain/aws370.

Lebraud, H., Wright, D.J., Johnson, C.N., Heightman, T.D., 2016. Protein degradation by in-cell self-assembly of proteolysis targeting chimeras. ACS Cent. Sci. 2, 927–934. https://doi.org/10.1021/acscentsci.6b00280.

Lee, S.J., Langhans, S.A., 2012. Anaphase-promoting complex/cyclosome protein Cdc27 is a target for curcumin-induced cell cycle arrest and apoptosis. BMC Cancer 12, 44. https://doi.org/10.1186/1471-2407-12-44.

Lee, B.H., Lee, M.J., Park, S., Oh, D.C., Elsasser, S., Chen, P.C., Gartner, C., Dimova, N., Hanna, J., Gygi, S.P., Wilson, S.M., King, R.W., Finley, D., 2010. Enhancement of proteasome activity by a small-molecule inhibitor of USP14. Nature 467, 179–184. https://doi.org/10.1038/nature09299.

Lee, J.K., Chang, N., Yoon, Y., Yang, H., Cho, H., Kim, E., Shin, Y., Kang, W., Oh, Y.T., Mun, G.I., Joo, K.M., Nam, D.H., Lee, J., 2016. USP1 targeting impedes GBM growth by inhibiting stem cell maintenance and radioresistance. Neuro Oncol. 18, 37–47. https://doi.org/10.1093/neuonc/nov091.

Lee, Y.C., Huang, W.C., Lin, J.H., Kao, T.J., Lin, Hui Ching, Lee, K.H., Lin, Hsin Chuan, Shen, C.K.J., Chang, W.C., Huang, C.C., 2018. Znf179 E3 ligase-mediated TDP-43 polyubiquitination is involved in TDP-43- ubiquitinated inclusions (UBI) (+)-related neurodegenerative pathology. J. Biomed. Sci. 25, 76. https://doi.org/10.1186/s12929-018-0479-4.

Leestemaker, Y., de Jong, A., Witting, K.F., Penning, R., Schuurman, K., Rodenko, B., Zaal, E.A., van de Kooij, B., Laufer, S., Heck, A.J.R., Borst, J., Scheper, W., Berkers, C.R., Ovaa, H., 2017. Proteasome activation by small molecules. Cell Chem. Biol. 24, 725–736.e7. https://doi.org/10.1016/j.chembiol.2017.05.010.

Lefaki, M., Papaevgeniou, N., Chondrogianni, N., 2017. Redox regulation of proteasome function. Redox Biol. 13, 452–458. https://doi.org/10.1016/j.redox.2017.07.005.

Li, F., Gong, Q., Dong, H., Shi, J., 2012a. Resveratrol, a neuroprotective supplement for Alzheimer's disease. Curr. Pharm. Des. 18, 27–33. https://doi.org/10.2174/138161212798919075.

Li, J.M., Yang, Y., Zhu, P., Zheng, F., Gong, F.L., Mei, Y.W., 2012b. Mitoxantrone exerts both cytotoxic and immunoregulatory effects on activated microglial cells. Immunopharmacol. Immunotoxicol. 34, 36–41. https://doi.org/10.3109/08923973.2011.572890.

Lin, A.W., Man, H.Y., 2013. Ubiquitination of neurotransmitter receptors and post-synaptic scaffolding proteins. Neural Plast. 2013, 432057. https://doi.org/10.1155/2013/432057.

Liu, S., Chen, Z.J., 2011. Expanding role of ubiquitination in NF-κB signaling. Cell Res. 21, 6–21. https://doi.org/10.1038/cr.2010.170.

Liu, Q.Y., Lei, J.X., Sikorska, M., Liu, R., 2008. A novel brain-enriched E3 ubiquitin ligase RNF182 is up regulated in the brains of Alzheimer's patients and targets ATP6V0C for degradation. Mol. Neurodegener. 3, 4. https://doi.org/10.1186/1750-1326-3-4.

Liu, S., Yu, C., Yang, F., Paganini-Hill, A., Fisher, M.J., 2012. Phosphodiesterase inhibitor modulation of brain microvascular endothelial cell barrier properties. J. Neurol. Sci. 320, 45–51. https://doi.org/10.1016/j.jns.2012.06.005.

Liu, Y., Hettinger, C.L., Zhang, D., Rezvani, K., Wang, X., Wang, H., 2014. Sulforaphane enhances proteasomal and autophagic activities in mice and is a potential therapeutic reagent for Huntington's disease. J. Neurochem. 129, 539–547. https://doi.org/10.1111/jnc.12647.

Liu, H., Li, W., Rose, M.E., Hickey, R.W., Chen, J., Uechi, G.T., Balasubramani, M., Day, B.W., Patel, K.V., Graham, S.H., 2015. The point mutation UCH-L1 C152A protects primary neurons against cyclopentenone prostaglandin-induced cytotoxicity: implications for post-ischemic neuronal injury. Cell Death Dis. 6, e1966. https://doi.org/10.1038/cddis.2015.323.

Liu, X., Hebron, M., Shi, W., Lonskaya, I., Moussa, C.E.H., 2019. Ubiquitin specific protease-13 independently regulates parkin ubiquitination and alpha-synuclein clearance in alpha-synucleinopathies. Hum. Mol. Genet. 28, 548–560. https://doi.org/10.1093/hmg/ddy365.

Lokireddy, S., Kukushkin, N.V., Goldberg, A.L., 2015. cAMP-induced phosphorylation of 26S proteasomes on Rpn6/PSMD11 enhances their activity and the degradation of misfolded proteins. Proc. Natl. Acad. Sci. U. S. A. 112, E7176–E7185. https://doi.org/10.1073/pnas.1522332112.

Lonskaya, I., Desforges, N.M., Hebron, M.L., Moussa, C.E.H., 2013. Ubiquitination increases parkin activity to promote autophagic a-synuclein clearance. PLoS One 8, e83914. https://doi.org/10.1371/journal.pone.0083914.

Lu, B., Al-Ramahi, I., Valencia, A., Wang, Q., Berenshteyn, F., Yang, H., Gallego-Flores, T., Ichcho, S., Lacoste, A., Hild, M., Difiglia, M., Botas, J., Palacino, J., 2013. Identification of NUB1 as a suppressor of mutant Huntingtin toxicity via enhanced protein clearance. Nat. Neurosci. 16, 562–570. https://doi.org/10.1038/nn.3367.

Lu, K., Psakhye, I., Jentsch, S., 2014. A new class of ubiquitin-Atg8 receptors involved in selective autophagy and polyQ protein clearance. Autophagy 10 (12), 2381–2382. https://doi.org/10.4161/15548627.2014.981919.

Lu, K., den Brave, F., Jentsch, S., 2017. Receptor oligomerization guides pathway choice between proteasomal and autophagic degradation. Nat. Cell Biol. 19 (6), 732–739. https://doi.org/10.1038/ncb3531.

Lu, M., Liu, T., Jiao, Q., Ji, J., Tao, M., Liu, Y., You, Q., Jiang, Z., 2018. Discovery of a Keap1-dependent peptide PROTAC to knockdown Tau by ubiquitination-proteasome degradation pathway. Eur. J. Med. Chem. 146, 251–259. https://doi.org/10.1016/j.ejmech.2018.01.063.

Lüscher, C., Malenka, R.C., 2011. Drug-evoked synaptic plasticity in addiction: from molecular changes to circuit remodeling. Neuron 69, 650–663. https://doi.org/10.1016/j.neuron.2011.01.017.

Lussier, M.P., Herring, B.E., Nasu-Nishimura, Y., Neutzner, A., Karbowski, M., Youle, R.J., Nicoll, R.A., Roche, K.W., 2012. Ubiquitin ligase RNF167 regulates AMPA receptor-mediated synaptic transmission. Proc. Natl. Acad. Sci. U. S. A. 109, 19426–19431. https://doi.org/10.1073/pnas.1217477109.

Mabb, A.M., Ehlers, M.D., 2010. Ubiquitination in postsynaptic function and plasticity. Annu. Rev. Cell Dev. Biol. 26, 179–210. https://doi.org/10.1146/annurev-cellbio-100109-104129.

Mabb, A.M., Je, H.S., Wall, M.J., Robinson, C.G., Larsen, R.S., Qiang, Y., Corrêa, S.A.L., Ehlers, M.D., 2014. Triad3A regulates synaptic strength by ubiquitination of arc. Neuron 82, 1299–1316. https://doi.org/10.1016/j.neuron.2014.05.016.

Mahad, D.H., Trapp, B.D., Lassmann, H., 2015. Pathological mechanisms in progressive multiple sclerosis. Lancet Neurol. 14, 183–193. https://doi.org/10.1016/S1474-4422(14)70256-X.

Maheshwari, M., Shekhar, S., Singh, B.K. uma., Jamal, I., Vatsa, N., Kumar, V., Sharma, A., Jana, N.R. anja., 2014. Deficiency of Ube3a in Huntington's disease mice brain increases aggregate load and accelerates disease pathology. Hum. Mol. Genet. 23, 6235–6245. https://doi.org/10.1093/hmg/ddu343.

Maiti, P., Manna, J., Veleri, S., Frautschy, S., 2014. Molecular chaperone dysfunction in neurodegenerative diseases and effects of curcumin. Biomed Res. Int. 2014, 495091. https://doi.org/10.1155/2014/495091.

Malhotra, S., Morcillo-Suárez, C., Nurtdinov, R., Rio, J., Sarro, E., Moreno, M., Castilló, J., Navarro, A., Montalban, X., Comabella, M., 2013. Roles of the ubiquitin peptidase USP18 in multiple sclerosis and the response to interferon-β treatment. Eur. J. Neurol. 20, 1390–1397. https://doi.org/10.1111/ene.12193.

Malm, T.M., Iivonen, H., Goldsteins, G., Keksa-Goldsteine, V., Ahtoniemi, T., Kanninen, K., Salminen, A., Auriola, S., Van Groen, T., Tanila, H., Koistinaho, J., 2007. Pyrrolidine dithiocarbamate activates Akt and improves spatial learning in APP/PS1 mice without affecting β-amyloid burden. J. Neurosci. 27, 3712–3721. https://doi.org/10.1523/JNEUROSCI.0059-07.2007.

Mandel, I., Paperna, T., Miller, A., 2012. Aberrant expression of the apoptosis-related proteins BAK and MCL1 in T cells in multiple sclerosis. J. Neuroimmunol. 244, 51–56. https://doi.org/10.1016/j.jneuroim.2011.12.026.

Mao, J., Xia, Q., Liu, C., Ying, Z., Wang, H., Wang, G., 2017. A critical role of Hrd1 in the regulation of optineurin degradation and aggresome formation. Hum. Mol. Genet. 26, 1877–1889. https://doi.org/10.1093/hmg/ddx096.

Massaly, N., Dahan, L., Baudonnat, M., Hovnanian, C., Rekik, K., Solinas, M., David, V., Pech, S., Zajac, J.M., Roullet, P., Mouledous, L., Frances, B., 2013. Involvement of protein degradation by the ubiquitin proteasome system in opiate addictive behaviors. Neuropsychopharmacology 38, 596–604. https://doi.org/10.1038/npp.2012.217.

Massaly, N., Francès, B., Moulédous, L., 2015. Roles of the ubiquitin proteasome system in the effects of drugs of abuse. Front. Mol. Neurosci. 7, 99. https://doi.org/10.3389/fnmol.2014.00099.

Mealer, R.G., Murray, A.J., Shahani, N., Subramaniam, S., Snyder, S.H., 2014. Rhes, a striatal-selective protein implicated in huntington disease, binds Beclin-1 and activates autophagy. J. Biol. Chem. 289, 3547–3554. https://doi.org/10.1074/jbc.M113.536912.

Medina, J.H., Viola, H., 2018. ERK1/2: a key cellular component for the formation, retrieval, reconsolidation and persistence of memory. Front. Mol. Neurosci. 11, 361. https://doi.org/10.3389/fnmol.2018.00361.

Mhyre, T.R., Boyd, J.T., Hamill, R.W., Maguire-Zeiss, K.A., 2012. Parkinson's disease. Subcell. Biochem. 65, 389–455. https://doi.org/10.1007/978-94-007-5416-4_16.

Miller, V.M., Nelson, R.F., Gouvion, C.M., Williams, A., Rodriguez-Lebron, E., Harper, S.Q., Davidson, B.L., Rebagliati, M.R., Paulson, H.L., 2005. CHIP suppresses polyglutamine aggregation and toxicity in vitro and in vivo. J. Neurosci. 25, 9152–9161. https://doi.org/10.1523/JNEUROSCI.3001-05.2005.

Min, J.W., Lü, L., Freeling, J.L., Martin, D.S., Wang, H., 2017. USP14 inhibitor attenuates cerebral ischemia/reperfusion-induced neuronal injury in mice. J. Neurochem. 140, 826–833. https://doi.org/10.1111/jnc.13941.

Minagar, A., Ma, W., Zhang, X., Wang, X., Zhang, K., Steven Alexander, J., Gonzalez-Toledo, E., Albitar, M., 2012. Plasma ubiquitin-proteasome system profile in patients with multiple sclerosis: correlation with clinical features, neuroimaging, and treatment with interferon-beta-1b. Neurol. Res. 34, 611–618. https://doi.org/10.1179/1743132812Y.0000000055.

Mishra, A., Maheshwari, M., Chhangani, D., Fujimori-Tonou, N., Endo, F., Joshi, A.P., Jana, N.R., Yamanaka, K., 2013. E6-AP association promotes SOD1 aggresomes degradation and suppresses toxicity. Neurobiol. Aging 34, 1310.e11–1310.e23. https://doi.org/10.1016/j.neurobiolaging.2012.08.016.

Mitra, S., Finkbeiner, S., 2008. The ubiquitin-proteasome pathway in Huntington's disease. Sci. World J. 8, 421–433. https://doi.org/10.1100/tsw.2008.60.

Miyauchi, Y., Yogosawa, S., Honda, R., Nishida, T., Yasuda, H., 2002. Sumoylation of Mdm2 by protein inhibitor of activated STAT (PIAS) and RanBP2 enzymes. J. Biol. Chem. 277, 50131–50136. https://doi.org/10.1074/jbc.M208319200.

Miyazaki, K., Fujita, T., Ozaki, T., Kato, C., Kurose, Y., Sakamoto, M., Kato, S., Goto, T., Itoyama, Y., Aoki, M., Nakagawara, A., 2004. NEDL1, a novel ubiquitin-protein isopeptide ligase for dishevelled-1, targets mutant superoxide dismutase-1. J. Biol. Chem. 279, 11327–11335. https://doi.org/10.1074/jbc.M312389200.

Morgan, A.R., Turic, D., Jehu, L., Hamilton, G., Hollingworth, P., Moskvina, V., Jones, L., Lovestone, S., Brayne, C., Rubinsztein, D.C., Lawlor, B., Gill, M., O'Donovan, M.C., Owen, M.J., Williams, J., 2007. Association studies of 23 positional/functional candidate genes on chromosome 10 in late-onset Alzheimer's disease. Am. J. Med. Genet. B Neuropsychiatr. Genet. 144, 762–770. https://doi.org/10.1002/ajmg.b.30509.

Mukaetova-Ladinska, E.B., Hurt, J., Jakes, R., Xuereb, J., Honer, W.G., Wischik, C.M., 2000. α-Synuclein inclusions in Alzheimer and Lewy body diseases. J. Neuropathol. Exp. Neurol. 59, 408–417. https://doi.org/10.1093/jnen/59.5.408.

Mulherkar, S.A., Sharma, J., Jana, N.R., 2009. The ubiquitin ligase E6-AP promotes degradation of α-synuclein. J. Neurochem. 110, 1955–1964. https://doi.org/10.1111/j.1471-4159.2009.06293.x.

Murata, H., Sakaguchi, M., Kataoka, K., Huh, N.H., 2013. SARM1 and TRAF6 bind to and stabilize PINK1 on depolarized mitochondria. Mol. Biol. Cell 24, 2772–2784. https://doi.org/10.1091/mbc.E13-01-0016.

Musiek, E.S., Holtzman, D.M., 2016. Mechanisms linking circadian clocks, sleep, and neurodegeneration. Science (80-.) 354, 1004–1008. https://doi.org/10.1126/science.aah4968.

Myeku, N., Duff, K.E., 2018. Targeting the 26S proteasome to protect against proteotoxic diseases. Trends Mol. Med. 24, 18–29. https://doi.org/10.1016/j.molmed.2017.11.006.

Myeku, N., Clelland, C.L., Emrani, S., Kukushkin, N.V., Yu, W.H., Goldberg, A.L., Duff, K.E., 2016. Tau-driven 26S proteasome impairment and cognitive dysfunction can be prevented early in disease by activating cAMP-PKA signaling. Nat. Med. 22, 46–53. https://doi.org/10.1038/nm.4011.

Nagamine, S., Kabuta, T., Furuta, A., Yamamoto, K., Takahashi, A., Wada, K., 2010. Deficiency of ubiquitin carboxy-terminal hydrolase-L1 (UCH-L1) leads to vulnerability to lipid peroxidation. Neurochem. Int. 57, 102–110. https://doi.org/10.1016/j.neuint.2010.04.015.

Nagashima, Y., Kowa, H., Tsuji, S., Iwata, A., 2011. FAT10 protein binds to polyglutamine proteins and modulates their solubility. J. Biol. Chem. 286, 29594–29600. https://doi.org/10.1074/jbc.M111.261032.

Naik, E., Webster, J.D., DeVoss, J., Liu, J., Suriben, R., Dixit, V.M., 2014. Regulation of proximal T cell receptor signaling and tolerance induction by deubiquitinase Usp9X. J. Exp. Med. 211, 1947–1955. https://doi.org/10.1084/jem.20140860.

Nakashima, H., Nguyen, T., Goins, W.F., Chiocca, E.A., 2015. Interferon-stimulated gene 15 (ISG15) and ISG15-linked proteins can associate with members of the selective autophagic process, histone deacetylase 6 (HDAC6) and SQSTM1/p62. J. Biol. Chem. 290, 1485–1495. https://doi.org/10.1074/jbc.M114.593871.

Nalbandian, A., Llewellyn, K.J., Nguyen, C., Yazdi, P.G., Kimonis, V.E., 2015. Rapamycin and chloroquine: the in vitro and in vivo effects of autophagy-modifying drugs show promising results in Valosin containing protein multisystem proteinopathy. PLoS One 10, e0122888. https://doi.org/10.1371/journal.pone.0122888.

Nath, S.R., Lieberman, A.P., 2017. The ubiquitination, disaggregation and proteasomal

degradation machineries in polyglutamine disease. Front. Mol. Neurosci. 10, 78. https://doi.org/10.3389/fnmol.2017.00078.
Navabi, S.P., Sarkaki, A., Mansouri, E., Badavi, M., Ghadiri, A., Farbood, Y., 2018. The effects of betulinic acid on neurobehavioral activity, electrophysiology and histological changes in an animal model of the Alzheimer's disease. Behav. Brain Res. 337, 99–106. https://doi.org/10.1016/j.bbr.2017.10.002.
Nemes, Z., Devreese, B., Steinert, P.M., Van Beeumen, J., Fésüs, L., 2004. Cross-linking of ubiquitin, HSP27, parkin, and α-synuclein by γ-glutamyl-ε-lysine bonds in Alzheimer's neurofibrillary tangles. FASEB J. 18, 1135–1137. https://doi.org/10.1096/fj.04-1493fje.
Nihira, N.T., Ogura, K., Shimizu, K., North, B.J., Zhang, J., Gao, D., Inuzuka, H., Wei, W., 2017. Acetylation-dependent regulation of MDM2 E3 ligase activity dictates its oncogenic function. Sci. Signal. 10, eaai8026. https://doi.org/10.1126/scisignal.aai8026.
Ochaba, J., Monteys, A.M., O'Rourke, J.G., Reidling, J.C., Steffan, J.S., Davidson, B.L., Thompson, L.M., 2016. PIAS1 regulates mutant huntingtin accumulation and Huntington's disease-associated phenotypes in vivo. Neuron 90, 507–520. https://doi.org/10.1016/j.neuron.2016.03.016.
Orlicky, S., Tang, X., Neduva, V., Elowe, N., Brown, E.D., Sicheri, F., Tyers, M., 2010. An allosteric inhibitor of substrate recognition by the SCF Cdc4 ubiquitin ligase. Nat. Biotechnol. 28, 733–737. https://doi.org/10.1038/nbt.1646.
Paiva, S.L., Da Silva, S.R., De Araujo, E.D., Gunning, P.T., 2018. Regulating the master regulator: controlling ubiquitination by thinking outside the active site. J. Med. Chem. 61, 405–421. https://doi.org/10.1021/acs.jmedchem.6b01346.
Pankiv, S., Clausen, T.H., Lamark, T., Brech, A., Bruun, J.A., Outzen, H., Øvervatn, A., Bjørkøy, G., Johansen, T., 2007. p62/SQSTM1 binds directly to Atg8/LC3 to facilitate degradation of ubiquitinated protein aggregates by autophagy*[S]. J. Biol. Chem. 282, 24131–24145. https://doi.org/10.1074/jbc.M702824200.
Patten, S.A., Aggad, D., Martinez, J., Tremblay, E., Petrillo, J., Armstrong, G.A., La Fontaine, A., Maios, C., Liao, M., Ciura, S., Wen, X.Y., Rafuse, V., Ichida, J., Zinman, L., Julien, J.P., Kabashi, E., Robitaille, R., Korngut, L., Parker, J.A., Drapeau, P., 2017. Neuroleptics as therapeutic compounds stabilizing neuromuscular transmission in amyotrophic lateral sclerosis. JCI Insight 2https://doi.org/10.1172/jci.insight.97152. pii: 97152.
Peterson, L.K., Fujinami, R.S., 2007. Inflammation, demyelination, neurodegeneration and neuroprotection in the pathogenesis of multiple sclerosis. J. Neuroimmunol. 184, 37–44. https://doi.org/10.1016/j.jneuroim.2006.11.015.
Popa-Wagner, A., Mitran, S., Sivanesan, S., Chang, E., Buga, A.M., 2013. ROS and brain diseases: the good, the bad, and the ugly. Oxid. Med. Cell. Longev. 2013, 963520. https://doi.org/10.1155/2013/963520.
Proctor, C.J., Gray, D.A., 2010. GSK3 and p53 — is there a link in Alzheimer's disease? Mol. Neurodegener. 5, 7. https://doi.org/10.1186/1750-1326-5-7.
Qin, C., Hu, Y., Zhou, B., Fernandez-Salas, E., Yang, C.Y., Liu, L., McEachern, D., Przybranowski, S., Wang, M., Stuckey, J., Meagher, J., Bai, L., Chen, Z., Lin, M., Yang, J., Ziazadeh, D.N., Xu, F., Hu, J., Xiang, W., Huang, L., Li, S., Wen, B., Sun, D., Wang, S., 2018. Discovery of QCA570 as an exceptionally potent and efficacious proteolysis targeting chimera (PROTAC) degrader of the bromodomain and extra-terminal (BET) proteins capable of inducing complete and durable tumor regression. J. Med. Chem. 61, 6685–6704. https://doi.org/10.1021/acs.jmedchem.8b00506.
Quik, M., Bordia, T., Zhang, D., Perez, X.A., 2015. Nicotine and nicotinic receptor drugs: potential for parkinson's disease and drug-induced movement disorders. Int. Rev. Neurobiol. 124, 247–271. https://doi.org/10.1016/bs.irn.2015.07.005.
Ragagnin, A.M.G., Shadfar, S., Vidal, M., Jamali, M.S., Atkin, J.D., 2019. Motor neuron susceptibility in ALS/FTD. Front. Neurosci. 13, 532. https://doi.org/10.3389/fnins.2019.00532.
Rajakumar, D., Alexander, M., Oommen, A., 2013. Oxidative stress, NF-κB and the ubiquitin proteasomal pathway in the pathology of calpainopathy. Neurochem. Res. 38, 2009–2018. https://doi.org/10.1007/s11064-013-1107-z.
Reddy, P.H., Oliver, D.M., 2019. Amyloid Beta and phosphorylated tau-induced defective autophagy and mitophagy in Alzheimer's disease. Cells 8, 488. https://doi.org/10.3390/cells8050488.
Regland, B., Lehmann, W., Abedini, I., Blennow, K., Jonsson, M., Karlsson, I., Sjögren, M., Wallin, A., Xilinas, M., Gottfries, C.G., 2001. Treatment of Alzheimer's disease with clioquinol. Dement. Geriatr. Cogn. Disord. 12, 408–414. https://doi.org/10.1159/000051288.
Ren, Z.Y., Liu, M.M., Xue, Y.X., Ding, Z.B., Xue, L.F., Zhai, S.Di, Lu, L., 2013. A critical role for protein degradation in the nucleus accumbens core in cocaine reward memory. Neuropsychopharmacology 38, 778–790. https://doi.org/10.1038/npp.2012.243.
Rial, D., Castro, A.A., Machado, N., Garção, P., Gonçalves, F.Q., Silva, H.B., Tomé, Â.R., Köfalvi, A., Corti, O., Raisman-Vozari, R., Cunha, R.A., Prediger, R.D., 2014. Behavioral phenotyping of Parkin-deficient mice: looking for early preclinical features of Parkinson's disease. PLoS One 9, e114216. https://doi.org/10.1371/journal.pone.0114216.
Richet, E., Pooler, A.M., Rodriguez, T., Novoselov, S.S., Schmidtke, G., Groettrup, M., Hanger, D.P., Cheetham, M.E., van der Spuy, J., 2012. NUB1 modulation of GSK3β reduces tau aggregation. Hum. Mol. Genet. 21, 5254–5267. https://doi.org/10.1093/hmg/dds376.
Ristic, G., Tsou, W.L., Todi, S.V., 2014. An optimal ubiquitin-proteasome pathway in the nervous system: the role of deubiquitinating enzymes. Front. Mol. Neurosci. 7, 72. https://doi.org/10.3389/fnmol.2014.00072.
Rodrigues, E.M., Scudder, S.L., Goo, M.S., Patrick, G.N., 2016. Aβ-induced synaptic alterations require the E3 ubiquitin ligase Nedd4-1. J. Neurosci. 36, 1590–1595. https://doi.org/10.1523/JNEUROSCI.2964-15.2016.
Rose, I., 2005. Ubiquitin at fox chase. Cell Death Differ. 102, 11575–11577. https://doi.org/10.1038/sj.cdd.4401710.
Rosen, K.M., Moussa, C.E.H., Lee, H.K., Kumar, P., Kitada, T., Qin, G., Fu, Q., Querfurth, H.W., 2010. Parkin reverses intracellular β-amyloid accumulation and its negative effects on proteasome function. J. Neurosci. Res. 88, 167–178. https://doi.org/10.1002/jnr.22178.
Rott, R., Szargel, R., Shani, V., Hamza, H., Savyon, M., Elghani, F.A., Bandopadhyay, R., Engelender, S., 2017. SUMOylation and ubiquitination reciprocally regulate α-synuclein degradation and pathological aggregation. Proc. Natl. Acad. Sci. U. S. A. 114, 13176–13181. https://doi.org/10.1073/pnas.1704351114.
Saito, R., Kaneko, M., Kitamura, Y., Takata, K., Kawada, K., Okuma, Y., Nomura, Y., 2014. Effects of oxidative stress on the solubility of HRD1, a ubiquitin ligase implicated in Alzheimer's disease. PLoS One 9, e94576. https://doi.org/10.1371/journal.pone.0094576.
Sari, Y., 2011. Huntington's disease: from mutant huntingtin protein to neurotrophic factor therapy. Int. J. Biomed. Sci. 7, 89–100.
Schipper-Krom, S., Juenemann, K., Reits, E.A.J., 2012. The ubiquitin-proteasome system in Huntington's disease: are proteasomes impaired, initiators of disease, or coming to the rescue? Biochem. Res. Int. 2012, 837015. https://doi.org/10.1155/2012/837015.
Schludi, M.H., May, S., Grässer, F.A., Rentzsch, K., Kremmer, E., Küpper, C., Klopstock, T., Ceballos-Baumann, A., Danek, A., Diehl-Schmid, J., Fassbender, K., Förstl, H., Kornhuber, J., Otto, M., Dieterich, M., Feuerecker, R., Giese, A., Klünemann, H., Kurz, A., Levin, J., Lorenzl, S., Meyer, T., Nübling, G., Roeber, S., 2015. Distribution of dipeptide repeat proteins in cellular models and C9orf72 mutation cases suggests link to transcriptional silencing. Acta Neuropathol. 130, 537–555. https://doi.org/10.1007/s00401-015-1450-z.
Schneekloth, J.S., Fonseca, F.N., Koldobskiy, M., Mandal, A., Deshaies, R., Sakamoto, K., Crews, C.M., 2004. Chemical genetic control of protein levels: selective in vivo targeted degradation. J. Am. Chem. Soc. 126, 3748–3754. https://doi.org/10.1021/ja039025z.
Schou, J., Prockop, L.D., Dahlström, G., Rohde, C., 1977. Penetration of delta-9-tetrahydrocannabinol and 11-OH-delta-9-tetrahydrocannabinol through the Blood-brain Barrier. Acta Pharmacol. Toxicol. (Copenh) 41, 33–38. https://doi.org/10.1111/j.1600-0773.1977.tb02120.x.
Schreiber, J., Végh, M.J., Dawitz, J., Kroon, T., Loos, M., Labonté, D., Li, K.W., Van Nierop, P., Van Diepen, M.T., De Zeeuw, C.I., Kneussel, M., Meredith, R.M., Smit, A.B., Van Kesteren, R.E., 2015. Ubiquitin ligase TRIM3 controls hippocampal plasticity and learning by regulating synaptic γ-actin levels. J. Cell Biol. 211, 569–586. https://doi.org/10.1083/jcb.201506048.
Schweiger, S., Matthes, F., Posey, K., Kickstein, E., Weber, S., Hettich, M.M., Pfurtscheller, S., Ehninger, D., Schneider, R., Krauß, S., 2017. Resveratrol induces dephosphorylation of Tau by interfering with the MID1-PP2A complex. Sci. Rep. 7, 13753. https://doi.org/10.1038/s41598-017-12974-4.
Sellebjerg, F., Krakauer, M., Khademi, M., Olsson, T., Sørensen, P.S., 2012. FOXP3, CBLB and ITCH gene expression and cytotoxic T lymphocyte antigen 4 expression on CD4 + CD25 high T cells in multiple sclerosis. Clin. Exp. Immunol. 170, 149–155. https://doi.org/10.1111/j.1365-2249.2012.04654.x.
Semmler, S., Gagné, M., Garg, P., Pickles, S.R., Baudouin, C., Hamon-Keromen, E., Destroismaisons, L., Khalfallah, Y., Chaineau, M., Caron, E., Bayne, A.N., Trempe, J.-F., Cashman, N.R., Star, A.T., Haqqani, A.S., Durcan, T.M., Meiering, E.M., Robertson, J., Grandvaux, N., Plotkin, S.S., McBride, H.M., Velde, C.V., 2019. The E3 ubiquitin ligase TRAF6 is a novel interacting protein of amyotrophic lateral sclerosis-linked misfolded SOD1. bioRxiv, 780460. https://doi.org/10.1101/780460.
Serna, S., Xolalpa, W., Lang, V., Aillet, F., England, P., Reichardt, N., Rodriguez, M.S., 2016. Efficient monitoring of protein ubiquitylation levels using TUBEs-based microarrays. FEBS Lett. 590, 2748–2756. https://doi.org/10.1002/1873-3468.12289.
Shin, J.H., Ko, H.S., Kang, H., Lee, Yunjong, Lee, Yun Il, Pletinkova, O., Troconso, J.C., Dawson, V.L., Dawson, T.M., 2011. PARIS (ZNF746) repression of PGC-1α contributes to neurodegeneration in Parkinson's disease. Cell 144, 689–702. https://doi.org/10.1016/j.cell.2011.02.010.
Silva, M.C., Ferguson, F.M., Cai, Q., Donovan, K.A., Nandi, G., Patnaik, D., Zhang, T., Huang, H.T., Lucente, D.E., Dickerson, B.C., Mitchison, T.J., Fischer, E.S., Gray, N.S., Haggarty, S.J., 2019. Targeted degradation of aberrant tau in frontotemporal dementia patient-derived neuronal cell models. eLife 8, e45457. https://doi.org/10.7554/eLife.45457.
Singh, B.K., Vatsa, N., Kumar, V., Shekhar, S., Sharma, A., Jana, N.R., 2017. Ube3a deficiency inhibits amyloid plaque formation in APPswe/PS1δE9 mouse model of Alzheimer's disease. Hum. Mol. Genet. 26, 4042–4054. https://doi.org/10.1093/hmg/ddx295.
Sivandzade, F., Prasad, S., Bhalerao, A., Cucullo, L., 2019. NRF2 and NF-κB interplay in cerebrovascular and neurodegenerative disorders: molecular mechanisms and possible therapeutic approaches. Redox Biol. 21, 101059. https://doi.org/10.1016/j.redox.2018.11.017.
Song, T., Liang, S., Liu, J., Zhang, T., Yin, Y., Geng, C., Gao, S., Feng, Y., Xu, H., Guo, D., Roberts, A., Gu, Y., Cang, Y., 2018. CRL4 antagonizes SCFFbxo7-mediated turnover of cereblon and BK channel to regulate learning and memory. PLoS Genet. 14, e1007165. https://doi.org/10.1371/journal.pgen.1007165.
Sugiura, A., Nagashima, S., Tokuyama, T., Amo, T., Matsuki, Y., Ishido, S., Kudo, Y., McBride, H.M., Fukuda, T., Matsushita, N., Inatome, R., Yanagi, S., 2013. MITOL regulates endoplasmic reticulum-mitochondria contacts via mitofusin2. Mol. Cell 51, 20–34. https://doi.org/10.1016/j.molcel.2013.04.023.
Sultana, R., Boyd-Kimball, D., Poon, H.F., Cai, J., Pierce, W.M., Klein, J.B., Merchant, M., Markesbery, W.R., Butterfield, D.A., 2006. Redox proteomics identification of oxidized proteins in Alzheimer's disease hippocampus and cerebellum: an approach to understand pathological and biochemical alterations in AD. Neurobiol. Aging 27, 1564–1576. https://doi.org/10.1016/j.neurobiolaging.2005.09.021.
Sun, J., Zhu, G., Liu, Y., Standley, S., Ji, A., Tunuguntla, R., Wang, Y., Claus, C., Luo, Y.,

Baudry, M., Bi, X., 2015. UBE3A regulates synaptic plasticity and learning and memory by controlling SK2 channel endocytosis. Cell Rep. 12, 449–461. https://doi.org/10.1016/j.celrep.2015.06.023.
Sweeney, P., Park, H., Baumann, M., Dunlop, J., Frydman, J., Kopito, R., McCampbell, A., Leblanc, G., Venkateswaran, A., Nurmi, A., Hodgson, R., 2017. Protein misfolding in neurodegenerative diseases: implications and strategies. Transl. Neurodegener. 6, 6. https://doi.org/10.1186/s40035-017-0077-5.
Szargel, R., Rott, R., Engelender, S., 2008. Synphilin-1 isoforms in Parkinson's disease: regulation by phosphorylation and ubiquitylation. Cell. Mol. Life Sci. 65, 80–88. https://doi.org/10.1007/s00018-007-7343-0.
Tashiro, Y., Urushitani, M., Inoue, H., Koike, M., Uchiyama, Y., Komatsu, M., Tanaka, K., Yamazaki, M., Abe, M., Misawa, H., Sakimura, K., Ito, H., Takahashia, R., 2012. Motor neuron-specific disruption of proteasomes, but not autophagy, replicates amyotrophic lateral sclerosis. J. Biol. Chem. 287, 42984–42994. https://doi.org/10.1074/jbc.M112.417600.
Tetzlaff, J.E., Putcha, P., Outeiro, T.F., Ivanov, A., Berezovska, O., Hyman, B.T., McLean, P.J., 2008. CHIP targets toxic α-synuclein oligomers for degradation. J. Biol. Chem. 283, 17962–17968. https://doi.org/10.1074/jbc.M802283200.
Thompson, L.M., Aiken, C.T., Kaltenbach, L.S., Agrawal, N., Illes, K., Khoshnan, A., Martinez-Vincente, M., Arrasate, M., O'Rourke, J.G., Khashwji, H., Lukacsovich, T., Zhu, Y.Z., Lau, A.L., Massey, A., Hayden, M.R., Zeitlin, S.O., Finkbeiner, S., Green, K.N., LaFerla, F.M., Bates, G., Huang, L., Patterson, P.H., Lo, D.C., Cuervo, A.M., Marsh, J.L., Steffan, J.S., 2009. IKK phosphorylates Huntingtin and targets it for degradation by the proteasome and lysosome. J. Cell Biol. 187, 1083–1099. https://doi.org/10.1083/jcb.200909067.
Tramutola, A., Di Domenico, F., Barone, E., Perluigi, M., Butterfield, D.A., 2016. It is all about (U)biquitin: role of altered ubiquitin-proteasome system and UCHL1 in alzheimer disease. Oxid. Med. Cell. Longev. 2016, 2756068. https://doi.org/10.1155/2016/2756068.
Tramutola, A., Triani, F., Di Domenico, F., Barone, E., Cai, J., Klein, J.B., Perluigi, M., Butterfield, D.A., 2018. Poly-ubiquitin profile in Alzheimer disease brain. Neurobiol. Dis. 118, 129–141. https://doi.org/10.1016/j.nbd.2018.07.006.
Tsai, Y.M., Chien, C.F., Lin, L.C., Tsai, T.H., 2011. Curcumin and its nano-formulation: the kinetics of tissue distribution and blood-brain barrier penetration. Int. J. Pharm. 416, 331–338. https://doi.org/10.1016/j.ijpharm.2011.06.030.
Tsang, A.H.K., Lee, Y.I.L., Ko, H.S., Savitt, J.M., Pletnikova, O., Troncoso, J.C., Dawson, V.L., Dawson, T.M., Chung, K.K.K., 2009. S-Nitrosylation of XIAP compromises neuronal survival in Parkinson's disease. Proc. Natl. Acad. Sci. U. S. A. 106, 4900–4905. https://doi.org/10.1073/pnas.0810595106.
Twohig, D., Nielsen, H.M., 2019. α-Synuclein in the pathophysiology of Alzheimer's disease. Mol. Neurodegener. 14, 23. https://doi.org/10.1186/s13024-019-0320-x.
Uchida, T., Tamaki, Y., Ayaki, T., Shodai, A., Kaji, S., Morimura, T., Banno, Y., Nishitsuji, K., Sakashita, N., Maki, T., Yamashita, H., Ito, H., Takahashi, R., Urushitani, M., 2016. CUL2-mediated clearance of misfolded TDP-43 is paradoxically affected by VHL in oligodendrocytes in ALS. Sci. Rep. 6, 19118. https://doi.org/10.1038/srep19118.
Umeda, T., Tomiyama, T., Sakama, N., Tanaka, S., Lambert, M.P., Klein, W.L., Mori, H., 2011. Intraneuronal amyloid β oligomers cause cell death via endoplasmic reticulum stress, endosomal/lysosomal leakage, and mitochondrial dysfunction in vivo. J. Neurosci. Res. 89, 1031–1042. https://doi.org/10.1002/jnr.22640.
United Nations, 2019. World Population Prospects 2019]. United Nations. Dep. Econ. Soc. Aff. World Popul. Prospect. 2019. Report ST/. . https://population.un.org/wpp/Publications/Files/W.
Upadhya, S.C., Smith, T.K., Hegde, A.N., 2004. Ubiquitin-proteasome-mediated CREB repressor degradation during induction of long-term facilitation. J. Neurochem. 91, 210–219. https://doi.org/10.1111/j.1471-4159.2004.02707.x.
Uranga, R.M., Katz, S., Salvador, G.A., 2013. Enhanced phosphatidylinositol 3-kinase (PI3K)/Akt signaling has pleiotropic targets in hippocampal neurons exposed to iron-induced oxidative stress. J. Biol. Chem. 288, 19773–19784. https://doi.org/10.1074/jbc.M113.457622.
Vaden, J.H., Bhattacharyya, B.J., Chen, P.C., Watson, J.A., Marshall, A.G., Phillips, S.E., Wilson, J.A., King, G.D., Miller, R.J., Wilson, S.M., 2015. Ubiquitin-specific protease 14 regulates c-Jun N-terminal kinase signaling at the neuromuscular junction. Mol. Neurodegener. 10, 3. https://doi.org/10.1186/1750-1326-10-3.
Vilchez, D., Saez, I., Dillin, A., 2014. The role of protein clearance mechanisms in organismal ageing and age-related diseases. Nat. Commun. 5, 5659. https://doi.org/10.1038/ncomms6659.
von Grabowiecki, Y., Abreu, P., Blanchard, O., Palamiuc, L., Benosman, S., Mériaux, S., Devignot, V., Gross, I., Mellitzer, G., Gonzalez de Aguilar, J.L., Gaiddon, C., 2016. Transcriptional activator TAp63 is upregulated in muscular atrophy during ALS and induces the pro-atrophic ubiquitin ligase Trim63. eLife 5, e10528. https://doi.org/10.7554/eLife.10528.
Walters, B.J., Hallengren, J.J., Theile, C.S., Ploegh, H.L., Wilson, S.M., Dobrunz, L.E., 2014. A catalytic independent function of the deubiquitinating enzyme USP14 regulates hippocampal synaptic short-term plasticity and vesicle number. J. Physiol. 592, 571–586. https://doi.org/10.1113/jphysiol.2013.266015.
Wang, Y., Branicky, R., Stepanyan, Z., Carroll, M., Guimond, M.P., Hihi, A., Hayes, S., McBride, K., Hekimi, S., 2009. The anti-neurodegeneration drug clioquinol inhibits the aging-associated protein CLK-1. J. Biol. Chem. 284, 314–323. https://doi.org/10.1074/jbc.M807579200.
Wang, B., Zeng, L., Merillat, S.A., Fischer, S., Ochaba, J., Thompson, L.M., Barmada, S.J., Scaglione, K.M., Paulson, H.L., 2018a. The ubiquitin conjugating enzyme Ube2W regulates solubility of the Huntington's disease protein, huntingtin. Neurobiol. Dis. 109, 127–136. https://doi.org/10.1016/j.nbd.2017.10.002.
Wang, Q., Jiao, F., Zhang, P., Yan, J., Zhang, Z., He, F., Zhang, Q., Lv, Z., Peng, X., Cai, H., Tian, B., 2018b. CDK5-mediated phosphorylation-dependent ubiquitination and degradation of E3 ubiquitin ligases GP78 accelerates neuronal death in Parkinson's disease. Mol. Neurobiol. 55, 3709–3717. https://doi.org/10.1007/s12035-017-0579-2.
Weinreb, O., Mandel, S., Bar-Am, O., Yogev-Falach, M., Avramovich-Tirosh, Y., Amit, T., Youdim, M.B.H., 2009. Multifunctional neuroprotective derivatives of Rasagiline as anti-Alzheimer's disease drugs. Neurotherapeutics 6, 163–174. https://doi.org/10.1016/j.nurt.2008.10.030.
Wertz, I.E., Newton, K., Seshasayee, D., Kusam, S., Lam, C., Zhang, J., Popovych, N., Helgason, E., Schoeffler, A., Jeet, S., Ramamoorthi, N., Kategaya, L., Newman, R.J., Horikawa, K., Dugger, D., Sandoval, W., Mukund, S., Zindal, A., Martin, F., Quan, C., Tom, J., Fairbrother, W.J., Townsend, M., Warming, S., Devoss, J., Liu, J., Dueber, E., Caplazi, P., Lee, W.P., Goodnow, C.C., Balazs, M., Yu, K., Kolumam, G., Dixit, V.M., 2015. Phosphorylation and linear ubiquitin direct A20 inhibition of inflammation. Nature 528, 370–375. https://doi.org/10.1038/nature16165.
Widagdo, J., Chai, Y.J., Ridder, M.C., Chau, Y.Q., Johnson, R.C., Sah, P., Huganir, R.L., Anggono, V., 2015. Activity-dependent ubiquitination of GluA1 and GluA2 regulates AMPA receptor intracellular sorting and degradation. Cell Rep. 10, 783–795. https://doi.org/10.1016/j.celrep.2015.01.015.
Widagdo, J., Guntupalli, S., Jang, S.E., Anggono, V., 2017. Regulation of AMPA receptor trafficking by protein ubiquitination. Front. Mol. Neurosci. 10, 347. https://doi.org/10.3389/fnmol.2017.00347.
Williams, K.L., Topp, S., Yang, S., Smith, B., Fifita, J.A., Warraich, S.T., Zhang, K.Y., Farrawell, N., Vance, C., Hu, X., Chesi, A., Leblond, C.S., Lee, A., Rayner, S.L., Sundaramoorthy, V., Dobson-Stone, C., Molloy, M.P., Van Blitterswijk, M., Dickson, D.W., Petersen, R.C., Graff-Radford, N.R., Boeve, B.F., Murray, M.E., Pottier, C., Don, E., Winnick, C., McCann, E.P., Hogan, A., Daoud, H., Levert, A., Dion, P.A., Mitsui, J., Ishiura, H., Takahashi, Y., Goto, J., Kost, J., Gellera, C., Gkazi, A.S., Miller, J., Stockton, J., Brooks, W.S., Boundy, K., Polak, M., Muñoz-Blanco, J.L., Esteban-Pérez, J., Rábano, A., Hardiman, O., Morrison, K.E., Ticozzi, N., Silani, V., De Belleroche, J., Glass, J.D., Kwok, J.B.J., Guillemin, G.J., Chung, R.S., Tsuji, S., Brown, R.H., García-Redondo, A., Rademakers, R., Landers, J.E., Gitler, A.D., Rouleau, G.A., Cole, N.J., Yerbury, J.J., Atkin, J.D., Shaw, C.E., Nicholson, G.A., Blair, I.P., 2016. CCNF mutations in amyotrophic lateral sclerosis and frontotemporal dementia. Nat. Commun. 7, 11253. https://doi.org/10.1038/ncomms11253.
Winkle, C.C., Olsen, R.H.J., Kim, H., Moy, S.S., Song, J., Gupton, S.L., 2016. Trim9 deletion alters the morphogenesis of developing and adult-born hippocampal neurons and impairs spatial learning and memory. J. Neurosci. 36, 4940–4958. https://doi.org/10.1523/JNEUROSCI.3876-15.2016.
Wong, H.K., Bauer, P.O., Kurosawa, M., Goswami, A., Washizu, C., Machida, Y., Tosaki, A., Yamada, M., Knöpfel, T., Nakamura, T., Nukina, N., 2008. Blocking acid-sensing ion channel 1 alleviates Huntington's disease pathology via an ubiquitin-proteasome system-dependent mechanism. Hum. Mol. Genet. 17, 3223–3235. https://doi.org/10.1093/hmg/ddn218.
Xu, P., Duong, D.M., Seyfried, N.T., Cheng, D., Xie, Y., Robert, J., Rush, J., Hochstrasser, M., Finley, D., Peng, J., 2009. Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation. Cell 137, 133–145. https://doi.org/10.1016/j.cell.2009.01.041.
Xu, J., Zhang, X.Z., Zhang, Y.J., Li, Xiu Ming, Cai, Z.L., Li, Xiao Min, 2015. Silencing of SIAH1 in SH-SY5Y affects α-synuclein degradation pathway. Int. J. Clin. Exp. Pathol. 8, 12885–12892.
Xu, Y., Zhao, F., Qiu, Q., Chen, K., Wei, J., Kong, Q., Gao, B., Melo-Cardenas, J., Zhang, B., Zhang, Jinping, Song, J., Zhang, D.D., Zhang, Jianing, Fan, Y., Li, H., Fang, D., 2016. The ER membrane-anchored ubiquitin ligase Hrd1 is a positive regulator of T-cell immunity. Nat. Commun. 7, 12073. https://doi.org/10.1038/ncomms12073.
Yan, C., Xu, W., Huang, Y., Li, M., Shen, Y., You, H., Liang, X., 2016. HRD1-mediated IGF-1R ubiquitination contributes to renal protection of resveratrol in db/db mice. Mol. Endocrinol. 30, 600–613. https://doi.org/10.1210/me.2015-1277.
Yao, I., Takagi, H., Ageta, H., Kahyo, T., Sato, S., Hatanaka, K., Fukuda, Y., Chiba, T., Morone, N., Yuasa, S., Inokuchi, K., Ohtsuka, T., MacGregor, G.R., Tanaka, K., Setou, M., 2007. SCRAPPER-dependent ubiquitination of active zone protein RIM1 regulates synaptic vesicle release. Cell 130, 943–957. https://doi.org/10.1016/j.cell.2007.06.052.
Yao, I., Takao, K., Miyakawa, T., Ito, S., Setou, M., 2011. Synaptic E3 ligase SCRAPPER in contextual fear conditioning: extensive behavioral phenotyping of Scrapper heterozygote and overexpressing mutant mice. PLoS One 6, e17317. https://doi.org/10.1371/journal.pone.0017317.
Yoon, S.O., Park, D.J., Ryu, J.C., Ozer, H.G., Tep, C., Shin, Y.J., Lim, T.H., Pastorino, L., Kunwar, A.J., Walton, J.C., Nagahara, A.H., Lu, K.P., Nelson, R.J., Tuszynski, M.H., Huang, K., 2012. JNK3 perpetuates metabolic stress induced by aβ peptides. Neuron 75, 824–837. https://doi.org/10.1016/j.neuron.2012.06.024.
Yoshida, Y., Saeki, Y., Murakami, A., Kawawaki, J., Tsuchiya, H., Yoshihara, H., Shindo, M., Tanaka, K., 2015. A comprehensive method for detecting ubiquitinated substrates using TR-TUBE. Proc. Natl. Acad. Sci. U. S. A. 112, 4630–4635. https://doi.org/10.1073/pnas.1422313112.
Yue, W., Chen, Z., Liu, H., Yan, Chen, Chen, M., Feng, D., Yan, Chaojun, Wu, H., Du, L., Wang, Y., Liu, J., Huang, X., Xia, L., Liu, L., Wang, X., Jin, H., Wang, J., Song, Z., Hao, X., Chen, Q., 2014. A small natural molecule promotes mitochondrial fusion through inhibition of the deubiquitinase USP30. Cell Res. 24, 482–496. https://doi.org/10.1038/cr.2014.20.
Zhang, M., Deng, Y., Luo, Y., Zhang, S., Zou, H., Cai, F., Wada, K., Song, W., 2012. Control of BACE1 degradaton and APP processing by ubiquitin carboxyl-terminal hydrolase L1. J. Neurochem. 120, 1129–1138. https://doi.org/10.1111/j.1471-4159.2011.07644.x.
Zhang, M., Cai, F., Zhang, Shuting, Zhang, Si, Song, W., 2014. Overexpression of ubiquitin carboxyl-terminal hydrolase L1 (UCHL1) delays Alzheimer's progression in vivo. Sci. Rep. 4, 7298. https://doi.org/10.1038/srep07298.

Zhang, J., Zhang, R., Zhan, Z., Li, X., Zhou, F., Xing, A., Jiang, C., Chen, Y., An, L., 2017. Beneficial effects of sulforaphane treatment in Alzheimer's disease may be mediated through reduced HDAC1/3 and increased P75NTR expression. Front. Aging Neurosci. 9, 121. https://doi.org/10.3389/fnagi.2017.00121.

Zhao, T., Hong, Y., Li, S., Li, X.J., 2016. Compartment-dependent degradation of mutant huntingtin accounts for its preferential accumulation in neuronal processes. J. Neurosci. 36, 8317–8328. https://doi.org/10.1523/JNEUROSCI.0806-16.2016.

Zheng, Z., Keifer, J., 2009. PKA has a critical role in synaptic delivery of GluR1- and GluR4-containing AMPARs during initial stages of acquisition of in vitro classical conditioning. J. Neurophysiol. 101, 2539–2549. https://doi.org/10.1152/jn.91282.2008.

Zheng, Q., Huang, T., Zhang, L., Zhou, Y., Luo, H., Xu, H., Wang, X., 2016. Dysregulation of ubiquitin-proteasome system in neurodegenerative diseases. Front. Aging Neurosci. 8, 303. https://doi.org/10.3389/fnagi.2016.00303.

Zhong, L., Tan, Y., Zhou, A., Yu, Q., Zhou, J., 2005. Ring finger ubiquitin-protein isopeptide ligase Nrdp1/FLRF regulates parkin stability and activity. J. Biol. Chem. 280, 9425–9430. https://doi.org/10.1074/jbc.M408955200.

Zhou, L., Wang, H., Ren, H., Hu, Q., Ying, Z., Wang, G., 2015. Bcl-2 decreases the affinity of SQSTM1/p62 to poly-ubiquitin chains and suppresses the aggregation of misfolded protein in neurodegenerative disease. Mol. Neurobiol. 52, 1180–1189. https://doi.org/10.1007/s12035-014-8908-1.

Zhou, Q., Chen, B., Wang, X., Wu, L., Yang, Y., Cheng, X., Hu, Z., Cai, X., Yang, J., Sun, X., Lu, W., Yan, H., Chen, J., Ye, J., Shen, J., Cao, P., 2016. Sulforaphane protects against rotenone-induced neurotoxicity in vivo: involvement of the mTOR, Nrf2, and autophagy pathways. Sci. Rep. 6, 32206. https://doi.org/10.1038/srep32206.

Zhou, B., Hu, J., Xu, F., Chen, Z., Bai, L., Fernandez-Salas, E., Lin, M., Liu, L., Yang, C.Y., Zhao, Y., McEachern, D., Przybranowski, S., Wen, B., Sun, D., Wang, S., 2018. Discovery of a small-molecule degrader of Bromodomain and extra-terminal (BET) proteins with picomolar cellular potencies and capable of achieving tumor regression. J. Med. Chem. 61, 462–481. https://doi.org/10.1021/acs.jmedchem.6b01816.

Zhu, W., Xie, W., Pan, T., Xu, P., Fridkin, M., Zheng, H., Jankovic, J., Youdim, M.B.H., Le, W., 2007. Prevention and restoration of lactacystin-induced nigrostriatal dopamine neuron degeneration by novel brain-permeable iron chelators. FASEB J. 21, 3835–3844. https://doi.org/10.1096/fj.07-8386com.

Zucchelli, S., Marcuzzi, F., Codrich, M., Agostoni, E., Vilotti, S., Biagioli, M., Pinto, M., Carnemolla, A., Santoro, C., Gustincich, S., Persichetti, F., 2011. Tumor necrosis factor receptor-associated factor 6 (TRAF6) associates with Huntingtin protein and promotes its atypical ubiquitination to enhance aggregate formation. J. Biol. Chem. 286, 25108–25117. https://doi.org/10.1074/jbc.M110.187591.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/342927571

Women Empowerment in India: A Critical Analysis

Article · June 2020

CITATIONS
0

READS
7

2 authors:



Seema Singh
Delhi Technological University
49 PUBLICATIONS **64** CITATIONS
SEE PROFILE



Antra Singh
Delhi Technological University
4 PUBLICATIONS **4** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

University & SME's collaboration View project

Enhancing Employability Skill of Girls and Young Women View project

All content following this page was uploaded by Seema Singh on 14 July 2020.

The user has requested enhancement of the downloaded file.

ISSN:2320-0693
Vol-19-Issue-44-June-2020

Women Empowerment in India: A Critical Analysis

Prof. Seema Singh* and Dr. Antra Singh,**

** Professor of Economics and ** ICSSR Post Doctoral Fellow,*

Delhi Technological University, Delhi- 110 042, INDIA

Abstract

Being a traditional patriarchal society, women have been given a secondary status which is reflected in the economic, social and political spheres. However, women equality and empowerment has always remained a priority area and has been taken utmost care by stake holders. The paper critically investigates the Indian status among other countries and tries to find out preparedness to achieve Sustainable development Goal -5 of the United Nations. The paper develops argument on the basis of secondary sources as review of existing literature published in journal, books, reports of various, NGOs, Government and international organisations and websites. The paper critically examines women empowerment in India, various models and dimensions. The paper discusses constitutional safe guards as well as plans and programmes by the government and their implementation, indicators of women empowerment. However, the country ranks low while comparing with other countries. There is need of reassessing and modifying programmes to achieve SDG-5 by 2030.

Keywords: Women Empowerment; Gender Equality; Government; Social; Rights

This is a preliminary study by the University Women Association of Delhi for the project entitled, "Enhancing Employability Skills for Empowerment of Girls and Young Women" funded by Graduate Women International, Geneva.

1. Introduction

Traditionally being a patriarchal society, women have secondary role to play in every household in India. However, issues related to welfare of women have always been a priority among policy makers since independence, though the path has witnessed repositioning with time and as per the requirement from the Fifth Five Year Plan (1974-78) onwards, there has been a remarkable shift in the approach to women's issues from welfare to development. India has also endorsed various global efforts as the the Mexico Plan of Action (1975), the Nairobi Forward Looking Strategies (1985), Convention on Elimination of All Forms of Discrimination Against Women

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

(CEDAW) in 1993, the Beijing Declaration as well as the Platform for Action (1995) and the Outcome Document adopted by the UNGA Session on Gender Equality and Development and Peace for the 21st century. In this background, the paper discusses status of women empowerment in India and its status at the international spectrum with regard to achievement of SDG Goal 5 of the United Nations by 2030. The plan of the study is as follows. The next section deals with the review of literature and discusses the concept of women empowerment with supportive models alongwith dimensions. Section 3 provides the research objectives and methodology of the study. Section 4 analyses various legislations and government programmes for women empowerment, variegated initiatives taken for women empowerment by different N.G.O.s and corporate sector, international comparison regarding status of women empowerment followed by Section 5 which provides suggestions to meet the challenges and path forward to move ahead.

2. Review of Literature

Empowerment refers to a situation where the powerless gain greater control over resources and ideologies. It has been associated with terms such as autonomy, power, status and agency. The Indian constitution has very clearly given an equal level playing ground to women and has directed authorities to frame rules and regulations to safe guard the right. However, feminist scholars during 1970s as a way to challenge patriarchy, as a radical approach concerned with transforming power relations in favour of women's rights and gender equality It in 1980s (Batliwala, 1993, 2007) and as an individual process of self-transformation during 1990s (Batliwala, 1993; Kabeer, 1994; Rowlands, 1997; Sen, 1997).They highlight the complex reciprocal relationship between women's self-understanding (Kabeer, 1994) and capacity for self-expression (Sen, 1997), as well as women's access to and control over material resources.

The decade of nineties may be considered as an impertinent phase for women empowerment throughout the world as most of the world conferences highlighted towards seizing the opportunities. It accentuated issues of reproductive and sexual rights, violence against women and gender inequality. Even among the eight the Millennium Development Goals (MDG) of UN developed in 2000 was targeting to be achieved by 2015, the Goal three was to promote gender equality and empower women. Women empowerment is the defined as the change in the context of a women's life, which enables her increased capacity for leading a fulfilling human life. It gets reflected both in external qualities (health, mobility, education and awareness, status in the

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

family, participation in decision making) and internal qualities (self awareness and self confidence) Human Development in South Asia (2000) (Mathew, 2003). United Nations Development Fund for Women (UNIFEM) has been re-branded as the UN Entity for Gender Equality and the empowerment of Women.

2.1. Models of Women Empowerment

Three Dimensional Model

Women empowerment is considered as a process that takes place over time, making women agents who formulate choices, control resources and make strategic life choices (Lee-Rife, 2010). She emphasised on the strategic life choices (choice of livelihood, marital decision and child bearing decision) which are considered as first-order choices and how these impact second-order choice (inculcating values in children, daily household decision making and managing well being of the family) that are less consequential.

Figure.1 Three Dimensional Model by Kabeer



(Source: Kabeer, 1999)

The researcher emphasised having access to resources (material, human and social in form) is an obligatory pre-condition of empowerment as resources increase the ability to exercise choice (Kabeer, 1999). Kabeer emphasised that merely having access to resources is not sufficient women must also have the ability to identify and utilise the resources (Malhotra, Schuler and Boender, 2002). As per Kabeer, empowerment is both an end and a means to end for example one status is attaining education another status is acquiring gainful employment and a further step to it could be control over decision making related to self or family is the outcome of

Copyright © 2020Author

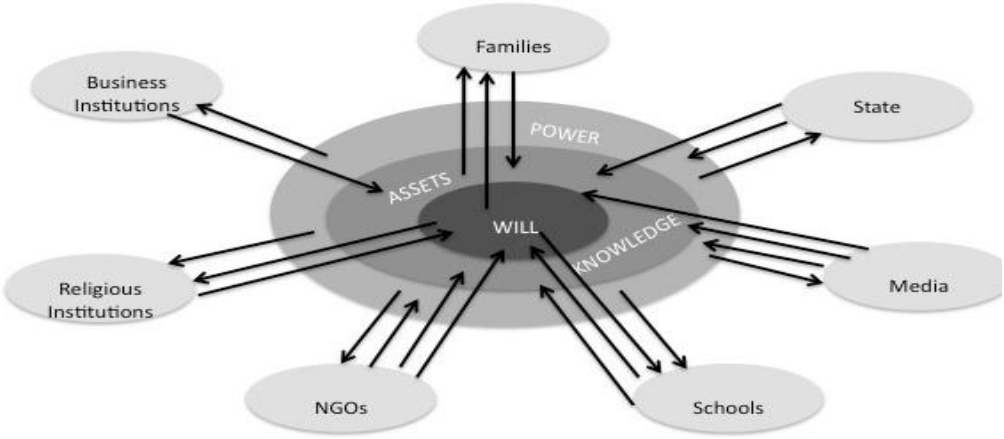
ISSN:2320-0693
Vol-19-Issue-44-June-2020

empowerment. Thus attaining education can be an end or it can also be a means to end but the final outcome of empowerment is having the capabilities of achievement.

The Commission on Women and Development's Conceptual Framework

The Commission on Women and Development, an International Aid Agency, the Government of Belgium designed an empowerment framework and methodology influenced by Kabeer, Longwez, Rowlands and Leon and the DAWN women's movement. The commission was of the view that empowerment occurred at two levels i.e. the individual level and the collective level. But the earlier model primarily highlighted the individual aspects of empowerment and identified the four core concepts of empowerment being assets, knowledge, will and capacity. This model splits resources into assets (material resources that give one greater economic power) and knowledge (provides people with techniques, training, literacy and critical analysis skills). Will is defined as psychological strength or power within to make one's own choices followed by capacity which is described as having the opportunities to use resources (assets, knowledge and will), make decisions and take on responsibility.

Figure 2. Conceptual Framework by the Commission on Women and Development



(Source: Commission of Women and Development, 2007)

The figure illustrates empowerment process as a catalyst for social change depicting will, assets and knowledge in concentric circles surrounded by power (power with others and power within). It also considers the institutional context comprising of family, government, media, educational institutions, NGOs, religious institutions and business institutions. The arrows pointing in and outward illustrates the potential to affect the components and not all arrows are same in weight.

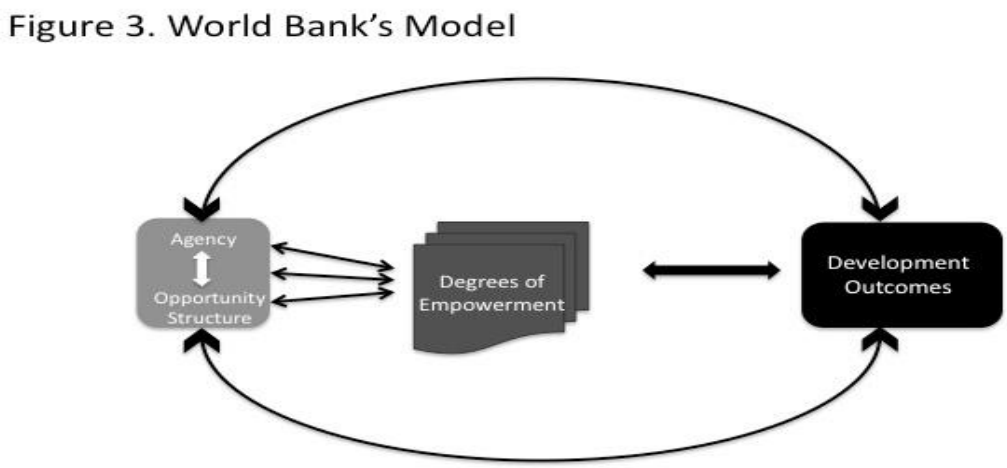
Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

World Bank's Agency and Opportunity Structure Model – Ruth Alsop and Nina Heinsohn

The capacity to make effective choices is dependent on two major factors contributing to empowerment and they are agency and opportunity structure. In this model assets are the indicators of agency and are characterised as psychological, informational, organisational, material, social, financial or human in nature. The other pre-condition in this model is the opportunity structure defined as the informal and formal institutions in which the individuals and the groups consider and includes laws, regulatory frameworks and norms governing behaviour, arguing that the opportunity structure determines who has access to assets and whether they can use their assets. The author hypothesizes that agency and opportunity structure are associated with the degrees (measured by the existence of choice, the use of choice and the achievement of choice) of empowerment of individuals and groups.

Figure 3. World Bank Model



(Source: Alsop and Heinsohn, 2005)

As discussed, researchers viewed women empowerment as a process with numerous inputs and factors that influence the process and finally, the outcome of these dimensions needs to be measured to quantify women empowerment. Numerous indicators are required to capture different dimensions of women's agency and empowerment like sexual and reproductive health, household decision-making, economic activities, participation in community decision making bodies and national politics and gender norms and women's motivations for their choices and

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

actions (Hanmer, 2015). These dimensions are mostly identical with minor alterations in sub divisions which depend on the focus of study and can be elaborated accordingly.

2.2. Dimensions of Women Empowerment

The Human Development Index was introduced for the first time by United Nations Development Program (UNDP) in 1990 to measure the socio economic progress of a nation. The 1995 report emphasised on empowering women as it was required for the overall human development. To study the statistics related to human development and the share of women, two indices were introduced namely Gender Development Index (GDI) that takes into account inequality in the achievements between women and men and Gender Empowerment Measure (GEM) which measures the extent to which a women has influence in decision-making in politics, in professional life and in organisations.

The Women Empowerment Matrix (Wieringa, 1994) was contrived to map out the general gender issues at national level and to outline the contours of dimensions of empowerment and to demonstrate the interlinkages of gender issues related to specific issues like education, health and women labour. The women empowerment matrix is a tool to aid the researcher in getting an overview of the various levels of women subordination presented along with various spheres in which women subordination is acted out. The matrix can be used for raising awareness, planning, policy analysis and others.

Table 1. Women Empowerment Matrix

Particulars	**Physical**	**Sociocultural**	**Religious**	**Political**	**Legal**	**Economic**
Individual						
Household						
Community						
State						
Region						
Global						

(Source: Wieringa, 1994)

The matrix provides a structure and holistic nature of specific issues to be dealt in case of women empowerment. Despite covering major issues still it lacks focus on class and ethinicity alongwith it is useful for qualitative usage and fails to provide historical depth.

The process of empowerment has five dimensions and they are Cognitive, Psychological, Economic, Political and Physical (Mokta, 2014). The different dimensions are explained below:

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

Table 2. Process of Empowerment and its dimensions

Dimensions	**Explanation**	**Parameters**
Cognitive	Women having an understanding of the conditions and causes of their subordination at the micro and macro levels. It involves making choices that may go against cultural expectations and norms	Raising self-esteem and self-confidence of women. • Elimination of discrimination and all forms of violence against women and girl child. •Building and strengthening partnership with civil society particularly women's organisations. • Enforcement of constitutional and legal provisions and safeguarding rights of women. • Building a positive image of women in the society and recognising their contributions in social, economic and political sphere. • Developing ability among women to think critically. • Fostering decision-making and collective action. • Enabling women to make informed choices. • Ensuring women's participation in all walks of life. • Providing information, knowledge, skills for self-employment
Psychological	Belief that women can act at personal and societal levels to improve their individual realities and the society in which they live	
Economic	Women have access to, and control over, productive resources, thus ensuring some degree of financial autonomy. However she notes that changes in the economic balance of power do not necessarily alter traditional gender roles or norms	
Political	Women have the capability to analyse, organise and mobilise for social change;	
Physical	Element of gaining control over one's body and sexuality and the ability to protect oneself against sexual violence to the empowerment process	

(Source: Mokta, 2014)

3. Proposed Model of Study and Methodology

The models discussed in the second section have dealt with factors which influences women empowerment. *Wieringa, 1993 and* Mokta 2014 have also discussed various dimension of Women empowerment. The model proposed herein discusses women empowerment on three facets and they are economic empowerment, social empowerment and political empowerment. These facets of women empowerment depends on legislative protections, affirmative policies for training, control over resources and various protective and promotional programmes of the government, NGOs and the corporate sector.

The study also aims to identify the gaps between present state of women empowerment in India and what is expected from stakeholders to achieve the Sustainable Development Goal – 5 on gender equality by 2030. Relevant secondary sources like scholarly articles, report by various ministries functional under Government of India, reports by corporate and non government organisations and international organisations have been used for review, analysis and discussion.

4. Analysis

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

From the Fifth Five Year Plan (1974-78) onwards there has been a marked shift in the approach to women's issues from welfare to development. In recent years, the empowerment of women has been recognized as the central issue in determining the status of women. The National Commission for Women was set up by an Act of Parliament in 1990 to safeguard the rights and legal entitlements of women. Several initiatives have been taken by the Government of India to achieve women empowerment. In course of discussing women empowerment, Tomar and Singh (2018) have given detailed list of constitutional safe guards. Relevant of them are given below:

4.1. Constitutional Articles, Government Plans and Programmes for Women Empowerment

Indian planners and policy makers have taken up the issue of women empowerment quite seriously. This section deals with the government's effort towards women empowerment.

4.1.1. Constitutional Articles

Women empowerment has been given required attention by the constitution committee headed by Baba Saheb Bhim Rao Ambedkar while framing the Indian Constitution. Some of the articles specifically dedicated to women have been discussed below:

4.1.1.1. Economic Empowerment

Provisions in the Indian Constitution

i. Equality of opportunity for all citizen in matters relating to employment or appointment to any office under the state (Article 16)

ii. The state to direct its policy towards securing for men and women equally the right for an adequate means of livelihood (Article 39(a)); and equal pay for equal work for both men and women [Article 39(d)]

iii. The state to make provision for securing just and humane conditions of work and for maternity relief (Article 42)

4.1.1.2. Political Empowerment

i. The 73rd Constitutional Amendment Acts passed in 1992 by the parliament of India ensures that not less than one-third (including the number of seats reserved for women belonging to Scheduled Caste and Scheduled Tribes) of the total number of seats to be filled by direct

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

election in every Panchayat to be reserved for women and such seats to be allotted by rotation to different constituencies in a Panchayat [Article 243D(3)]

ii. Not less than one-third of the total number of offices of the Chairpersons in the Panchayats at each level to be reserved for women (Article 243 D(4)]

iii. Not less than one -third (including the number of seats reserved for women belonging to the scheduled Caste and the Scheduled Tribes) of the total number of seats to be filled by direct election in every Municipality to be reserved for women and such seats to be allotted by rotation to different constituencies in a Municipality (Article 243T(3))

4.1.1.3. Social Empowerment

i. Equality before Law for women (Article 14)

ii. No discrimination against any citizen on grounds of sex beside religion, race, caste etc. (Article- 15)

iii. The state to make special provision in favour of women and children (Article 15(3))

iv. To promote justice, on the basis of equal opportunity and to provide free legal aid by suitable legislation or scheme or in any other way to ensure that opportunities for securing justice to all irrespective of economic inequality or other disability (Article 39A)

v. The state to promote with special care the educational and economic interests of the weaker sections of the people and to protect them from social injustice and all forms of exploitation (Article 46)

vi. To renounce practices derogatory to the dignity of women [Article 51(A)(e)]

4.1.2. Government Plans and Programmes

Since the inception of gender equality in the Constitution of India, it has been one of the priority areas and with special focus as SDG-5 government has been actively working to achieve gender equality through women empowerment and few initiatives have been discussed as below:

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

4.1.2.1. Economic Empowerment

i. The Ministry of Women and Child development has been administering 'Support to Training and Employment Programme for Women (STEP) Scheme' since 1986-87 as a 'Central Sector Scheme'. The scheme aims to provide employability skills and competency to women in the age group of 16 years and above to enable women to become self-employed/ entrepreneurs. The grant under the Scheme is given to an institution/ organisation including NGOs directly. The assistance will be available in any sector including but not limited to the Agriculture, Horticulture, Food Processing, Handlooms, Tailoring, Stitching, Embroidery, Zari etc, Handicrafts, Computer & IT enable services along with soft skills and skills for the work place such as spoken English, Gems & Jewellery, Travel & Tourism, Hospitality (STEP, 2009)

ii. The Government of India being concerned about the difficulties faced by working women, introduced a scheme in 1972-73 of grant-in- aid for construction of new/ expansion of existing buildings for providing hostel facilities to working women in cities, smaller towns and also in rural areas where employment opportunities for women exist. Based on an evaluation, the existing scheme has been revised to promote availability of safe and conveniently located accommodation for working women who need to live away from their families due to professional commitments (Scheme for Working Women Hostel, 1972-73). Rajiv Gandhi National Creche Scheme for the Children of Working Mothers was recast by government of India in 2006 and the aim of the programme was to provide day-care facilities for children (6 months to 6 years) of working mothers in the community, to improve nutrition and health status of children, to promote physical, cognitive, social and emotional development of children and further to educate and empower parents /caregivers for better childcare.

iii. The Ministry of Women and Child Development launched Mahila E-Haat a bilingual portal on 7th March, 2016. This is a unique direct online marketing platform leveraging technology for supporting women entrepreneurs/Self Help Groups/ Non Government Organisations for showcasing the products and services which are made, manufactured or undertaken by them. It is an initiative for meeting aspirations and needs of women. This was done keeping in mind that technology is a critical component for business efficiency and to make it available to the majority of Indian women entrepreneurs, Self Help Groups and Non Government Organisations. Mahila E-haat received the SKOCH GOLD Award on 09th Sep, 2016 and was also awarded 'SKOCH Order-of-Merit' Award, as it was adjudged as one of the "Top 100 Projects in India"

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

for the year 2016. To increase visibility major PSUs, IRCTC, Nationalised Banks like SBI have provided a link to Mahila E -haat on its website (https://wcd.nic.in/sites/default/files/Mahila%20E-haat%2024012017.pdf).

iv. The Finance Minister in his budget speech (2017-18) announced setting up of Mahila Shakti Kendra at village level in 14 lakh ICDS Anganwadi Centres with an allocation of 500 crore. The purpose of this scheme is to provide "one stop convergent support services for empowering rural women with opportunities for skill development, employment, digital literacy, health and nutrition". Ministry of Women and Child Development has proposed a sub-scheme namely Mahila Shakti Kendra under the Umbrella Scheme of Mission for Protection and Empowerment for Women, this scheme is envisaged to provide an interface for rural women to approach the government for availing their entitlements and for empowering them through training and capacity building. The scheme provides convergent support is being proposed for equal access to healthcare, quality education, career and vocational guidance, employment, health and safety, social security and digital literacy at Gram Panchayats level in selected districts/ blocks across the country (https://wcd.nic.in/sites/default/files/Mahila%20Shakti%20Kendra%20Scheme.pdf).National Mission for Empowerment of women Scheme is effectively merged with MSK wherein the National level and State level structures will continue and facilitate the implementation of MSK at the district, block and Gram panchayat level duly covering the Anganwadi Centres.

4.1.2.2. Social Empowerment

i. Beti Bachao Beti Padhao (BBBP) Scheme was launched in January 2015 in Panipat, Haryana by our Hon'ble Prime Minister to address the issue of decline in Child Sex ratio and related issues of empowerment of girls and women over a life cycle continuum. The scheme was initially launched in 100 districts in 2014-15, was expanded to 61 additional districts in 2015-16 and then it was expanded to all 640 districts of the country through a nation-wide mass media campaign, focused intervention and multi-sectoral action in select districts. Goal of the scheme is to celebrate the girl child and enable her education. (Beti Bachao Beti Padhao Scheme Guidelines, 2019) The campaign aims at ensuring that girls are born, nurtured and educated without discrimination to become empowered citizens of this country with equal rights. The government introduced many innovative activities under this scheme like Digital Guddi-Gudda Board, Udaan-Sapneya Di Duniya De Rubaru, My Aim My Target Campaign, LAkshya se

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

Rubaru, Pahal-Ek Kadam Nari Samman ki Aur, Ghar ki Pehchan Beti ke Naam, Bal Cabinet, Introduce Pink Cards and many more.Panic Button on Mobile Phones was an initiative by The Ministry of Women and Child Development since June 2014 had taken up the issue of installation of physical panic button on mobile phones.

ii. One Stop Centres (OSCs) are intended to support women affected by violence, in private and public spaces, within the family, community and at the workplace. Women facing physical, sexual, emotional, psychological and economic abuse, irrespective of age, class, caste, education status, marital status, race and culture will be facilitated with support and redressal. Aggrieved women facing any kind of violence due to attempted sexual harassment, sexual assault, domestic violence, trafficking, honour related crimes, acid attacks or witch-hunting who have reached out or been referred to the OSC will be provided with specialized services. Under this Scheme, in the first phase, one OSC envisaged to be established in each State/Union Territory to facilitate access to an integrated range of services including medical, legal, and psychological support. Further, 150 additional Centres are taken up in second phase during2016-17 in addition to 36 Centres in the first phase. The OSCs will be integrated with 181 and other existing helplines (One Stop Centre Scheme 2015). Women affected by violence and in need of redressal services could be referred to OSC through these help lines.

iii. The Scheme of Universalisation of Women Helpline is intended to provide 24 hours immediate and emergency response to women affected by violence through referral (linking with appropriate authority such as police, One Stop Centre, hospital) and information about women related government schemes programs across the country through a single uniform number. Women Helpline (WHL) will be integrated with One Stop Centre Scheme (OSC) under which one OSC shall be established in every State/UT to provide integrated support and assistance to women affected by violence, both in private and public spaces under one roof. Women affected by violence and in need of redressal services will be referred to OSC through WHL. Under the Scheme, the States/Union Territories will utilise or augment their existing women helplines through a dedicated single national number. Department of Telecommunication, Government of India has allocated short code 181 to all States/Union Territories (Universalisation of Women Helpline, 2015)

iv. UJJAWALA scheme has been effective from 2016 Ministry has formulated a Central Scheme "Comprehensive Scheme for Prevention of Trafficking for Rescue, Rehabilitation and Re-

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

Integration of Victims of Trafficking for Commercial Sexual Exploitation-Ujjawala". The new scheme has been conceived primarily for the purpose of preventing trafficking on the one hand and rescue and rehabilitation of victims on the other. The beneficiaries of the scheme are women and children who are vulnerable to trafficking for commercial sexual exploitation (Ujjawala, 2016).The implementing agencies can be the Social Welfare/Women and Child Welfare Department of State Government, Women's Development Corporations, Women's Development Centres, Urban Local Bodies, reputed Public/Private Trust or Voluntary Organizations.

v. Swadhar –A Scheme for Women in Difficult Circumstances was launched by the Department of Women and Child Development in 2001- 02. The scheme envisions a supportive institutional framework for women victims of difficult circumstances so that they could lead their life with dignity and conviction (Swadhar Greh, 2015). It envisages that shelter, food, clothing, and health as well as economic and social security are assured for such women. It also envisions that the special needs of these women are properly taken care of and under no circumstances they should be left unattended or abandoned which could lead to their exploitation and desolation.

vi. The Ministry of Women and Child Development, celebrates International Women's Day every year on 8th March as it is the nodal Ministry for the issues related to women. Taking this as an opportune moment to acknowledge Women's achievements, it has been decided by the Ministry in the year 2015, to confer Nari Shakti Puruskars to eminent women and institutions in recognition of their service towards the cause of women empowerment. In the last decade, there has been concerted effort by the Government to recognise and encourage women as reflected through a National Policy for Empowerment of Women in 2001. The issues related to women has gained utmost importance and focussed attention. "Nari Shakti Puruskars" shall showcase the Government's commitment towards women with the aim of strengthening their legitimate place in the society (http://narishaktipuraskar.wcd.gov.in/aboutus). It will also provide an opportunity to the current generation to understand the contribution of women in building of society and the nation.

vii. Following the tragedy (heinous rape of para-medical student on a moving bus in Delhi) of December 2012, the Government has set up a dedicated fund – Nirbhaya Fund – which can be utilized for projects specifically designed to improve the safety and security of women. It is a

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

non-lapsable corpus fund, being administered by Department of Economic Affairs, Ministry of Finance.

As per the guidelines issued by Ministry of Finance dated 25.03.2015, the Ministry of Women and Child Development (MWCD) is the nodal Ministry to appraise/recommend proposals and schemes to be funded under Nirbhaya Fund. MWCD further has the responsibility to review and monitor the progress of sanctioned schemes in conjunction with the line Ministries/Departments. An Empowered Committee was set up on 26.10.2015 to appraise and recommend proposals to be funded under Nirbhaya Fund. Government also initiated a portal to facilitate victims/ complainants to report cyber crime complaints online. Complaints on cyber crimes against women and children that are reported on this portal are dealt by law enforcement agencies/ police based on the information available in the complaints.

Women are susceptible to violence both in public and private spaces, including domestic violence, sexual assault, rape, stalking and others. A gender responsive police service requires specific training, increased presence of female personnel within the police force and community outreach to integrate gender issues into policies, protocols and operational procedures Government of India, 2016)In recent years, there has been enactment of various legislations in the Parliament which address issues related to Gender Based Violence like the Criminal Law Amendment Act, 2013, the Sexual Harassment of Women at Workplace (Prevention, Prohibition and Redressal) Act, 2013,the Protection of Women From Domestic Violence Act, 2005 and provided a platform to women facing violence to take the recourse to law. According to the latest National Crime Records Bureau data, during the year 2018, 3,78,277 incidences of crime against women (both under Indian Penal Code and other laws) were reported as against the 3,38,954 cases reported during 2016.

Table 3. Crime against women (IPC + SLL)

Particulars	**2016**	**2017**	**2018**
Total States	322949	345989	363817
Total Union Territories	16005	13860	14460
Total All India	338954	359849	378277

(Source: National Crime Records Bureau report of 2018, 2017 and 2016)

As observed by the Working Group on Women's Agency and Empowerment, these numbers have to be viewed keeping in mind that not all crimes against women are reported. The actual numbers may give even greater cause for concern. Role of Police is pivotal in safety and

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

security of citizens in general and women in particular. To increase the visibility of women in the police force, Home Ministry has carried forward the initiative to give 33% reservation to women in police force by implementing it in UTs and propagating in the States (12th Five Year Plan Report of the Working Group on Women's Agency and Empowerment, 2011). There has been an increasing emphasis on gender sensitivity of police force through training programmes, performance appraisal, women police stations to tackle crime against women.

Past research concluded that women empowerment can bring about demographic change lead to positive reproductive health outcomes (Mason, 1995). Empowering women economically is associated with higher levels of physical mobility, economic security, autonomous decision-making, exposure to violence, legal and political awareness and participation (Schuler and Hashemi, 1994). Women empowerment has also been linked with self-confidence, financial confidence, attitude towards gender norms, autonomous decision-making, household communication, partner relationships, participation in social groups and collective action (Mabsout, 2011; Hindin, 2000; Yount et al., 2019; Basu and Koolwal, 2005; Upadhyay et al. 2014).

4.2. NGOs Initiative for Women Empowerment

Women empowerments especially in rural areas have been primary focus for Non-Government Organisations (NGOs) beside the Government. The Ministry of Women and Child Development collaborates with many NGOs for the implementation of its various schemes and for conducting research studies. The Government of India implements many of its schemes with the active participation and support of the NGOs as they play an important role in accelerating the pace of implementation at the ground level. Table 4 lists out presence of performing NGOs in states (with 77.55% of NGOs working for women empowerment) and it states the highest presence in Delhi and the least in Goa, Mizoram and Tripura and one of the reasons could be based on the statistics of crime rate in these states.

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

Table 4. State-wise List of NGOs

State	**Performing NGOs**	**% age of NGOs working for Women Empowerment***
Andhra Pradesh	35	68.6
Arunachal Pradesh	4	100
Assam	38	81.6
Bihar	18	100
Chandigarh	7	57.1
Chattisgarh	7	100
Delhi	146	54.8
Goa	1	-
Gujarat	16	93.8
Haryana	17	58.8
Himachal Pradesh	8	100
Jammu and Kashmir	2	50
Jharkhand	11	90.9
Karnataka	63	74.7
Kerala	12	75
Madhya Pradesh	12	75
Maharashtra	82	90.2
Manipur	37	89.1
Meghalaya	2	-
Mizoram	1	100
Nagaland	2	100
Orissa	72	87.5
Punjab	5	60
Rajasthan	25	92
Tamil Nadu	53	87.7
Telangana	20	75
Uttar Pradesh	61	86.9
Uttarakhand	15	73.3
Tripura	1	-
West Bengal	37	73
Total	**811**	**77.5**

* the area of work included women development or/and empowerment

(Source: List of Performing and Non-Performing NGOs issued by Ministry of Women and Child Development, Govt. Of India)

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

Some works conducted by NGOs in this area are as follows:

i. Apna Ghar Ashram support destitute and disabled women get a safe and caring home. More than 50% of their residents have been rehabilitated. 458 men and 1,004 women currently live in the Ashrams spread across 17 major cities across India. The Ashram was also awarded the State Award by the State Government as the Best Services Organization.

ii. Started in 1973, Alpnalaya works for women empowerment in India by helping them to overcome the social, political, and economic barriers. They provide sewing classes and train women who otherwise have not been given the choice or opportunity to become financially independent. 209 women have economically benefited from Self Help groups organized by Apnalaya. In 2017, this NGO for women was awarded the Champion Level- GuideStar India Platinum for the highest levels of transparency and public accountability.

iii. Since its inception in 2005, Maitri has worked with over 45,000 individuals on the issues of social and health problems and public health concerns through education, community outreach, networking and legal advocacy. The organization has been addressing the issue of violence against women, running a victim support program and providing support to abandoned elderly widows in Vrindavan.

iv. Action Aid India since 1972 shelters the poor and the excluded. Gauravi is the first crisis center set up in India with the help of the Department of Health in 2014 to fight for Indian women's rights and saving them from domestic violence. Currently, there are 22 Gauravi centers across the country and over 40,000 cases have been registered in the centers.

v. Ashadeep Mission was initiated to develop a better quality of life (medical treatment and therapy) for people with mental disabilities. The Navachetana home run by Ashadeep rescues women with mental health issues and provides them with shelter, food, and medical care. They also help these women reunite with their families if possible after treatment. So far, 355 women have been rescued and rehabilitated under Ashadeep mission.

vi. Seven Sisters Development Assistance works to create a just and equitable society in Northeast India devoid of poverty. The organization has two programs, leadership programs for the educated youth of North East India and empowering poor rural women through livelihood activities via Self Help Groups (SHGs). Since it was established in 2011, SeSTA has mobilized nearly 8,000 women to form 700 Self Help Groups (SHGs).

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

vii. Foundation for Mother and Child Health works to bring good health and nutritional provisions to vulnerable mothers and children in poor communities. Their programs encourage preventive health, balanced nutrition, and child development practices in underprivileged communities. So far, the programs initiated by FMCH impact close to 927 pregnant and lactating mothers annually. Their services include regular growth monitoring, availability of accurate information and support from the detection of pregnancy till the child completes two years.

viii. Majlis Manch's legal center through its program 'Rahat', provides socio-legal support to victims of domestic violence and sexual abuse. The all-women team of lawyers and social workers meet with victims and explain the process and procedures involved in a court case. They are also told about all the legal, public, and private schemes that they have access to.

4.3. Corporate Social Responsibility (CSR) for Women Empowerment

As per the new Companies Bill passed by the lower house of Indian parliament in December 2012, Public Sector Units (PSUs) and private companies having a turnover of over 10 billion or a net worth of 5 billion or recorded net profit of 50 million and above are required to spend 2 per cent of their average net profits in the preceding three financial years towards CSR. A few of the CSR initiatives by companies for women empowerment is listed in table 5.

Table 5. Companies CSR Initiatives for Women Empowerment

Companies	**Focus Area**	**Prescribed CSR of Co. In FY 2017-18 (amount spent) in INR**	**Locations**	**Implementing Partner**
ITC	Capacity Building of Women (gainful employment opportunities)	290.47 Cr (14.76 Cr)	Andhra Pradesh, Telangana, Karnataka, Tamil Nadu, Bihar, Madhya Pradesh, Uttar Pradesh, Assam, Rajasthan, West Bengal, Delhi, Maharashtra, Odisha, Punjab, Uttarakhand, Himachal Pradesh	BAIF Development Research Foundation, Pratham Education Foundation, Ramakrishna Mission, Bandhan Konnagar, SEWA Bharat, Foundation for Ecological Security
IndusInd Bank Ltd.	Livelihood generation of rescued trafficked women (rehabilitation and skill development of women)	70.17 Cr (11.63 Cr)	Bihar (Gaya, Nalanda and Nawada)	Justice and Care
Godrej Consumer	Project Salon-i (trains young girls	18.83 Cr (8.86 Cr)	Gujarat, Maharashtra, Uttar Pradesh, Bihar, Chattisgarh,	Dhriiti, Ambuja Cement Foundation, DDJF, Don

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

Products Ltd	and women in beauty and hair care)		Assam, Nagaland, Manipur, Mizoram, Delhi	Bosco Tech Society, Father Agnel Ashram, Labournet, Saath, Pratham, NSHM, Save the Children India, Tara Livelihood Academy, Unnati
Inter Globe Aviation Limited	Project 6E Shakti (skill enhancement by providing agriculture and non-agriculture based training)	42.03 Cr (4.67 Cr)	Maharashtra, Assam, Meghalaya, Uttar Pradesh, Chattisgarh, Rajasthan	AFARM Grameen Sahara, Bandhan Konnagar, Divinity Serves
Hindustan Zinc Ltd	Sakhi (provide platforms for women to engage in full and effective participation in the public life of their villages)	189.65 Cr (3.57 Cr)	Rajasthan, Uttarakhand	Saheli Samiti & Manjari Foundation
Bayer Crop Science Limited	Discovering Hands-Training Center for Visually Impaired Women (training centre for visually impaired women to become medical tactile examiners in early breast Cancer Detection)	10.09 Cr (1.88 Cr)	Delhi	Indian Cancer Society, National Association for the Blind, Rajiv Gandhi Cancer & Research Centre, Disha Foundation
Sterlite Technologies Ltd	Vocational Training for Women (conducts vocational training on nursing, computer operations, data entry, tailoring, fashion designing and basic beauty culture)	3.63 Cr (1.72 Cr)	Maharashtra	Jeewan Jyoti Women Empowerment Institution
CEAT Ltd	Swayam (training underprivileged women in driving skills to enhance livelihoods across the transport sector)	10.71 Cr (1.45 Cr)	Maharashtra, Rajasthan, Delhi, Madhya Pradesh, Tamil Nadu	RPG Foundation
Sasken Technologies Ltd	m-bachat (platform for financial independence of women)	4.51 Cr (1.03 Cr)	Gujarat	SEWA
Apollo Tyres Ltd	Livelihood Enhancement Project for Women (supporting women farmers by training them on modern techniques and ways to increase agricultural productivity)	21.57 Cr (0.77 Cr)	Gujarat	Apollo Tyres Foundation

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

(Source: Compiled from the websites of the companies)

Big corporate houses consider women empowerment as a vital aspect of CSR and it has been observed that they usually carry these interventions in the surrounding areas of the main operational regions either by the in-house team or outsourced professional CSR team and it is sometimes carried out in collaboration with NGOs.

The 55th Round of the National Sample Survey (July 1999-June 2000), which was the first-ever nation-wide survey on informal sector non-agricultural enterprises, showed that the total number of informal workers in non-agricultural enterprises was 79.7 million. Out of these, 30 million were home based (Sudarshan et al, 2001) and a vast majority of these workers are women. Organisations like Dastkar has provided a platform to artisans (especially in rural areas) engaged in hand embroidery and other crafts to receive fair price (eliminating the middle men) for their products. It provides fair wage and fair rate for the products of the local artisans.

4.4. SDG Goal 5- Achieve Gender Equality: Status of India

Under SDG Goal 5 to achieve gender equality by 2030 there are several targets to be achieved aand they are as follows:

Target 1.End all forms of discrimination against all women and girls everywhere

Target 2. Eliminate all forms of violence against all women and girls in the public and private spheres, including trafficking and sexual and other types of exploitation

Target 3. Eliminate all harmful practices, such as child, early and forced marriage and female genital mutilation

Target 4.Recognize and value unpaid care and domestic work through the provision of public services, infrastructure and social protection policies and the promotion of shared responsibility within the household and the family as nationally appropriate

Target 5. Ensure women's full and effective participation and equal opportunities for leadership at all levels of decision-making in political, economic and public life

Target 6. Ensure universal access to sexual and reproductive health and reproductive rights as agreed in accordance with the Programme of Action of the International Conference on

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

Population and Development and the Beijing Platform for Action and the outcome documents of their review conferences

Target 7. Undertake reforms to give women equal rights to economic resources, as well as access to ownership and control over land and other forms of property, financial services, inheritance and natural resources, in accordance with national laws

Target 8. Enhance the use of enabling technology, in particular information and communications technology, to promote the empowerment of women

Target 9. Adopt and strengthen sound policies and enforceable legislation for the promotion of gender equality and the empowerment of all women and girls at all levels

The National Institution for Transforming India (NITI Aayog), has been assigned the responsibility of overseeing the implementation towards achievement of SDGs. A multi-disciplinary VNR Task Force was constituted by NITI Aayog to coordinate the review and process documentation.

India has achieved gender parity at primary education level and is on track to achieving parity at all education levels, the proportion of seats in the Lok Sabha held by women had only reached 11% and in Panchayati Raj Institutions by 46%. Though the sex ratio has improved marginally. Lowest was 927 women for 1000 male in 1991 which has marginally increased to 933 in 2001 and 940 in 2011) but, India is confronting the challenge of violence against women (domestic violence, sexual violence in public spaces, physical and mental exploitation at work place).

Thus, the Indian government has identified ending violence against women as a key national priority for the same it has targeted to end all forms of discrimination, all forms of violence and exploitation against girls/ women, eliminate all harmful practices like child, early and forced marriage, recognise and value unpaid care and domestic work through the provision of public services, social protection policies and promotion of shared responsibility within the household, enhance the usage of information and communications technology for empowerment of women.

4. Global Gender Gap Index: Status of India

4.4.1. World Economic Forum introduced the Global Gender Gap Index in 2006 as a framework for capturing the magnitude and scope of gender based disparities and tracking their progress

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

over time across four thematic dimensions and they are Economic Participation and Opportunity, Educational Attainment, Health and Survival and Political Empowerment.

Table 6. India's position in global Gender Gap

Year	Global Index	Economic Participation and Opportunity	Educational Attainment	Health and Survival	Political Empowerment
2020	112 (0.668)	149 (0.354)	112 (0.962)	150 (0.944)	18 (0.411)
2018	108 (0.665)	142 (.385)	114 (0.953)	147(0.940)	19 (0.382)
2016	87 (0.683)	136 (0.408)	113 (0.950)	142 (0.942)	9 (0.433)
2014	114 (0.6455)	134 (0.4096)	126 (0.8503)	141 (0.9366)	15 (0.3855)
2012	105 (0.6442)	123 (0.4588)	121 (0.8525)	134 (0.9312)	17 (0.3343)
2010	112 (0.6155)	128 (0.4025)	120 (0.8369)	132 (0.9312)	23 (0.2913)
2006	98 (0.601)	110 (0.397)	102 (0.819)	103 (0.962)	20 (0.227)

(Source: Global Gender Gap Report 2006, 2010, 2012, 2014, 2016, 2018 and 2020 by World Economic Forum)

The year 2016 witnessed an increase in the overall ranking which positioned India at 87 out of 144 countries and the reasons may be the Gross Domestic Product for the year which stood as US $ 2073.54 billions, population growth rate at 1.15%, population sex ratio (female to male) at 0.93 and human capital optimization at 57.73%. Another impertinent reason for the increase in the parity is associated with enrolment in primary education (1.01), secondary education (1.01) and tertiary education (0.94). Parity has also been maintained in healthy life expectancy (1.05) and sex ratio at birth (0.89).

India (112) maintains a stable ranking this year but its gap is directionally larger this year with a 33% gap yet to be bridged. The country records a decline in wage equality for similar work, succeeds in fully closing its tertiary education gender gap for the first time, and keeps primary and secondary education gaps closed for the third year running. However, it continues to rank third lowest in the world on Health and Survival, remaining the world's least-improved country on this sub index over the past decade. In fact, India actually widens the gender gap on this sub index this year. However, the condition of women in large fringes of India's society is precarious. Among the 153 countries studied, India is the only country where the economic gender gap is larger than the political gender gap.

4.4.2. South Asian Association for Regional Cooperation (SAARC) Nations is the regional intergovernmental organization and geopolitical union of states in South Asia. SAARC

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

comprises 3% of the world's area, 21% of the world's population and 4.21% (US$3.67 trillion) of the global economy, as of 2019. Table 7 points out that Maldives has gained first rank in achieving parity in educational attainment as compared to Pakistan that ranks 143 where the gap is 0.823 reason being the societal pressure in Pakistan which does not allow girls to attain education. India is improving in lowering the gender gap in providing education and penetrating in rural India by providing suitable infrastructure. Sri Lanka ranks first in maintaining gender parity in health and survival as compared to India which is also working towards achieving parity.

Table 7. Ranking of SAARC Nations in Global Gender Gap

Member States	**Global Index**	**Economic Participation and Opportunity**	**Educational Attainment**	**Health and Survival**	**Political Empowerment**
Afghanistan	NA	NA	NA	NA	NA
Bangladesh	50 (0.726)	141 (0.438)	120 (0.951)	119 (0.969)	7 (0.545)
Bhutan	131 (0.635)	130 (0.544)	116 (0.954)	144 (0.960)	132 (0.082)
India	112 (0.668)	149 (0.354)	112 (0.962)	150 (0.944)	18 (0.411)
Maldives	123 (0.646)	131 (0.518)	1 (1)	147 (0.953)	115 (0.111)
Nepal	101 (0.680)	101 (0.632)	133 (0.895)	131 (0.966)	59 (0.227)
Pakistan	151 (0.564)	150 (0.327)	143 (0.823)	149 (0.944)	93 (0.159)
Sri Lanka	102 (0.680)	126 (0.558)	88(0.988)	1 (0.980)	73 (0.193)

(Source: Global Gender Gap Report 2020 by World Economic Forum)

4.4.3. BRICS is the acronym for the economic bloc of countries consisting of Brazil, Russia, India, China and South Africa. Economists believe these four nations will become dominant suppliers of manufactured goods, services, and raw material by 2050 due to low labour and production costs. India is struggling to achieve gender parity among its counterparts which hold better position in all aspects. India ranks the lowest in comparison with other BRICS nations and needs to take effort to outshine its counterparts.

Table 8. Position of BRICS nation in Global Gender Gap

Member States	**Global Index**	**Economic Participation and Opportunity**	**Educational Attainment**	**Health and Survival**	**Political Empowerment**
Brazil	92 (0.691)	89 (0.653)	35 (1)	1 (0.980)	104 (0.133)
Russia (Russian Federation)	81 (0.706)	32 (0.749)	1(1)	1 (0.980)	122 (0.095)
India	112 (0.668)	149 (0.354)	112 (0.962)	150 (0.944)	18 (0.411)
China	106 (0.676)	91 (0.651)	100 (0.973)	153 (0.926)	95 (0.154)

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

South Africa	17 (0.780)	92 (0.649)	67 (0.994)	1 (0.980)	10 (0.497)

(Source: Global Gender Gap Report 2020 by World Economic Forum)

5. Conclusion and Recommendation

Women Empowerment is not merely a slogan but a prerequisite for an optimal capacity for overall development of a family, society, nation and a sustainable world. The paper has covered about affirmative legislation, plans and programmes initiated by the Government, NGOs and Corporate sector through their Corporate Social Responsibility over a very large spectrum starting from the Indian constitution to development in the recent years. It has been reflected by improvement in achieving gender parity in the category of educational attainment, showing improvement in political participation but showing a declining trend in economic participation and opportunity as well as in health and survival segment. In terms of Health and Survival, it is worse even among SAARC countries. As we have a decade towards achieving the SDG Goals, the Government needs to focus on these two segments more to achieve gender parity. There is urgent need of reassessing India's affirmative actions, giving fresh impetus on empowering girls and women not only through conventional education but also hand holding for skill acquisition and providing them with equitable economic opportunity with providing them with fitting medical and healthcare. The findings open up several avenues for policy discussion and intervention towards women empowerment.

References

Alsop, R. and Heinsohn, N. (2005) Measuring empowerment in practice: structuring analysis and framing indicators. World Bank

Basu, M. and Koolwal, G. B. (2005) Two concepts of empowerment: Some leads from DHS data on women's status and reproductive health. In S. Kishor (Ed.), A focus on gender: Collected papers on gender using DHS data, pg. 15-53

Batliwala, S. (1994) The meaning of women's empowerment: new concepts from action. In Population Policies Reconsidered: Health, Empowerment, and Rights, Sen G, Germain A, Chen LC (eds). Harvard Center for Population and Development Studies: Boston.

Batliwala, S. (2007) Taking the power out of empowerment: an experiential account. Development inPractice, 17(4/5), pg. 557–65.

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

Beti Bachao Beti Padhao Scheme (2019) Implementation Guidelines for State Government and UT Administrations, Governement of India pg. 1-103

Commission of Women and Development (2007) The Women Empowerment Approach: A Methodological Guide.

Government of India (2016) Sanction for Implementation of Mahila Police Volunteers during financial year 2016-17, pg. 1-17

Hindin, M (2000) Women's autonomy, women's status and fertility-related behavior in Zimbabwe. Population Research Policy Review, Vol.19, pg. 255–282

Lucia Hanmer & Jeni Klugman (2016) Exploring Women's Agency and Empowerment in Developing Countries: Where do we stand?, Feminist Economics, 22(1), pg. 237-263

Kabeer N. (1994) Reversed Realities: Gender Hierarchies in Development Thought. Verso: London.

Kabeer N. (1999) Resources, agency, achievements: reflections on the measurement of women'sempowerment. Development and Change 30(3), pg. 435–64.

Kabeer, N. (2008) Paid work, women's empowerment and gender justice, Pathways Working Paper3, Brighton: Pathways of Women's Empowerment.

Kabeer, N. (2011) Contextualising the economic pathways of women's empowerment, PathwaysPolicy Paper, Brighton: Pathways of Women's Empowerment.

Lee-Rife, SM. (2010) Women's empowerment and reproductive experiences over the lifecourse. Soc Sci Med

Mabsout, R (2011) Capability and Health Functioning in Ethiopian Households, Social Indicators Research, Vol 101, Issue 3, pg. 359-389

Malhotra A, Schuler SR, and Boender C. (2002) Measuring women's empowerment as variable in international development. World Bank Gender and DevelopmentGroup Background paper.

Mason, KO. (1995) International Union for the Scientific Study of Population. Gender and demographic change: What do we know?

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

Mathew, G. (2003) Keynote address in the workshop on "A Decade ofWomen's Empowerment through Local Governance" organized jointly byInstitute of Social Sciences and South Asia Partnership, Canada sponsored byInternational Development Research Centre

Ministry of Women and Child Development (2011) 12th Five Year Plan Report of the Working Group on Women's Agency and Empowerment, Government of India, pg. 1-118

One Stop Centre Scheme (2015) Government of India Guidelines pg. 1-12

Rowlands. (1997) Questioning Empowerment: Working with Women in Honduras. OxfamPublishing: Oxford.

Scheme for Working Women Hostel, 1972-73 Guidelines issued by Government of India, pg 1-28

Schuler, S.R. and Hashemi, S.M. (1994) Credit programs, women's empowerment andcontraceptive use in rural Bangladesh. Stud Fam Plann, 25(2), pg. 65-76.

Sen, G. (1997) Empowerment as an approach to poverty, Working Paper Series 97.07, backgroundpaper for the UNDP Human Development Report, New York: UNDP.

STEP (2009) Guidelines issued by Governement of India, pg 1-30

Sudarshan, R. M., Banerjee, M., Bhandari, L., Venkatraman, S. (2001), Outsourcing of Manufacturing to Households: Subcontracted Home Based Work in India, UNICEF Innocenti Research Centre (unpublished paper).

Swadhar Greh (2015) Scheme Guidelines issued by Government of India, pg 1-32

Tomar N. and Singh S. (2018), "How sensitive is the Indian Government Towards Women Empowerment in India" International Journal of Management, IT and Engineering, 08 (07), pp 66-83.

Ujjawala (2016) Scheme Guidelines, Government of India, pg 1-31it is not detail

Universalisation of Women Helpline (2015) Implementation Guidelines for State and Union Territories Administration, pg 1- 68

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

Upadhyay, U., Gipson, J., Withers, M., Lewis, S., Ciaraldi, E., Fraser, A., Huchko. M. and Prata, N (2014) Women's empowerment and fertility: A review of Literature, Social Science and Medicine, Vol. 115, pg. 111-120

Wieringa, S. (1994) Women's Interests and Empowerment: Gender Planning Reconsidered, Development and Change, 25(4), pg 829-848

Yount, K., Cheong, Y., Maxwell, L., Heckert, J., Martinez, E. And Seymour, G (2019) Measurement properties of the project level Women Empowerment in Agriculture Index, World Development, Vol. 124, pg. 1-19

Copyright © 2020Author

View publication stats

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

A Comparative Study of Image Dehazing Algorithms

Anil Singh Parihar, Yash Kumar Gupta, Yash Singodia, Vibhu Singh, Kavinder Singh
Machine Learning Research Lab, Dept. of Computer Science
Delhi Technological University
Delhi, India
parihar.anil@gmail.com, yashk.gupta97@gmail.com, yashsingodia@gmail.com, vibhusingh98@gmail.com,
kavinder85@gmail.com

***Abstract*—Haze poses a great challenge in modern-day applications. Removing haze is a challenging task because haze varies with the depth of the scenes in the image. Many automated systems like surveillance systems, object tracking, etc. use the dehazing methods internally to improve their overall performance in a hazy environment. Hence, recovering the original image from the hazy image is a crucial task. This paper provides a detailed survey of many algorithms that have been proposed for haze removal, which refines the image, either by using statistical observation of the scene or by network-based learning method. These algorithms have been assessed quantitatively and visually to compare based on their dehazing potential. The paper presents a comparative study of the various image dehazing methods having a different way of estimation of transmission and atmospheric light. This paper reviews different prior-based technique and learning-based technique image dehazing algorithms.**

***Index Terms*—Image dehazing, learning-based, CNN, Enhancement, Atmospheric scattering model**

I. Introduction

Image enhancement endeavours to improve the visual quality of an image. Image can degrade due to lighting, atmosphere, device issues and many other reasons. Researchers presented different approaches to deal with different issues and achieve image enhancement. Contrast enhancement [1]–[5], Haze removal [6]–[14], and Low-light enhancement [15]–[17] are a part of Image Enhancement.

The visibility of outdoor regions is deteriorated in adverse weather conditions with the presence of haze, fog, water droplets and various different particles in the atmosphere. The presence of haze is a dominant issue that decreases the contrast level and brings the noise to the image. Consequently, haze removal and improving the colour of the image becomes a difficult task. The volume of the haze or fog depends on the floating particles in the environment. These bits of particle absorb and scatter light, which reflects back from scene objects. The degree to which light scatters, rely on the separation gap between the object and the observer. Object tracking [18], Object detection [19], Action recognition [20], surveillance systems ,environment monitoring and other automated systems require clear and haze-free images to work properly. Moreover, It is difficult to perform any other vision-based task effectively in a hazy image. Thus, Haze removal becomes a necessity. Many researchers [21], [22] performed haze removal with the help of multiple images. These methods attempt to take more than one scene-image with an intent to learn the different features of the image. Separate images contain different features and thus, can help in capturing haze relevant feature effectively. However, it is sometimes not possible to have multiple images for feature extraction. It is a challenging task to distil out pertinent haze features from a single input image because of very less information provided by a single image. Various researchers [7], [8], [23]–[25] endeavoured to single image haze removal. The dehazed outputs produced by many single image dehazing algorithms appeared pleasing to the eye and the overall complexity also got reduced. Hence, addressing this problem is important in computer vision to recover images from the haze regions for useful information. The paper presents a contrasting examination of several image dehazing algorithms.

Section-II presents Atmospheric scattering model. Section-III contains Brief discussion of various single image dehazing methods. Section-IV propounds a detailed comparative analysis. Section-V is used for the conclusion.

II. Atmospheric Scattering Model

The Atmospheric scattering model [6] is extensively exploited to deal with hazy images. The model is stated as:

$$\mathbf{I}^c = \mathbf{R}^c\mathbf{T} + L^c(1 - \mathbf{T}) \tag{1}$$

here **I** depicts input-image, **R** denotes the scene radiance of the image, c denotes colour channels, and Transmission **T** depicts quantity of light that falls straight to lens of the

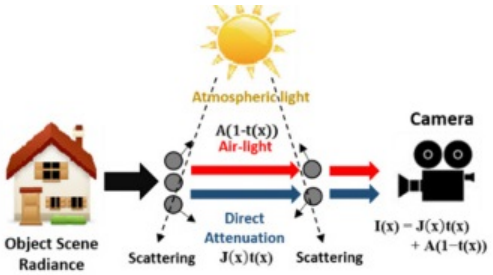


Fig. 1. Atmospheric Scattering Model [26]

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:00:15 UTC from IEEE Xplore. Restrictions apply.

camera without getting dispersed and L tells the atmospheric light. Also, the transmission is proportionate to the exponential distance between object and lens, therefore

$$\mathbf{T} = e^{-\beta \mathbf{D}} \tag{2}$$

here β is the atmosphere coefficient of scattering and **D** denotes the distance between the camera and the object. The model is widely employed in image dehazing to produce enhanced results **R** from input **I** by evaluating **T** and L.

III. LITERARY SURVEY

The main objective of image dehazing methods is to produce outputs that contains appropriate visible object in the image with almost no noise. This paper presents a brief deliberation of various image dehazing methods that are commonly used. Image dehazing can be used in variety of applications. Various methods are proposed by researchers to remove haze from an image and generate images with better image quality. The analysis contains prior-based and learning-based approaches. In the following section, various algorithms will be discussed.

A. DARK CHANNEL PRIOR

The dark channel prior (DCP) [7] forms on the statistical knowledge of outside scene without fog. It is built on the inspection that the local patchs Ω, which are not a part of sky region, have some pixels, with less intensity values in no less than one of the colour channel. These low intensities are accounted by the presence of shadows, colourful objects and dark surfaces. DCP is built on atmospheric model of image formation (1). The approach estimates Atmospheric light using dark-channel which is given as:

$$\mathbf{R}^{dark} = \min_{c \varepsilon (R,G,B)} (\min_{y \varepsilon \Omega} \mathbf{R}^c) \tag{3}$$

where $\mathbf{R}^{dark}$ denotes dark-channel, and Ω denotes a patch size. After estimation of dark-channel, it estimates Transmission map as:

$$\widehat{\mathbf{T}} = 1 - \omega \min_{c} (\min_{y \varepsilon \Omega} (\mathbf{I}^c / L^c)) \tag{4}$$

where $\widehat{\mathbf{T}}$ denotes Transmission map, and L^c denotes the channel-wise atmospheric light. DCP, assumes that the R^{dark} of a haze free image is nearly 0, and L^c is always non-negative.

$$\mathbf{R}^{dark} \rightarrow \min_{c \varepsilon R,G,B} (\min_{y \varepsilon \Omega} \mathbf{R}^c(y)) \tag{5}$$

Since, the Complete elimination of haze may appear displeasing to the eye. Thus, DCP preserves few traces of haze in the image to take care of the aerial perspective for distant objects. It is done by introducing a new parameter ω, which is set to 0.9.

$$\widehat{\mathbf{T}} = 1 - \omega \min_{c} (\min_{y \varepsilon \Omega} (\mathbf{I}^c(y) / L^c)) \tag{6}$$

The scene radiance can be recovered by using estimated Transmission, and Atmospheric light. Hence, It computes the final dehazed image as:

$$\mathbf{R} = \frac{\mathbf{I} - L}{max(\mathbf{T}, t_0)} \tag{7}$$

where t_0 denots the lower bound for Transmission, which keeps transmission at particular lower limit.

DCP work appropriately when images do not contain any white or brighter object. DCP generates artifacts while dealing with images having white object. In DCP, transmission is dependent on atmospheric light and prior. Thus, a improved prior can provide better enhancement.

B. COLOR ATTENUATION PRIOR

Colour attenuation prior (CAP) [8] is built on the scrutiny that, the difference in the intensity of brightness and that of an image saturation varies hazy regions of an image than haze-free regions greatly. For hazy parts, this difference is vast, whereas, the difference is low for the haze-free regions. Hence, It could be inferred that $\mathbf{D} \propto \rho \propto \mathbf{B}$ - **S**, where **D** symbolize the depth of scene, ρ is the haze concentration in the image, **S** is the intensity of saturation whereas **B** is the intensity brightness of the scene. CAP estimates depth map using the following learning model:

$$D(p) = \delta_0 + \delta_1 B(p) + \delta_2 S(p) + \varepsilon(p) \tag{8}$$

p is the position, in the image, $\delta_0, \delta_1, \delta_2$ are unknown parameter and θ_0 is equal to 1.0, θ_1 is equal to -1.0 and ϵ is a input image. The depth map can be estimated using equation (8). CAP estimates of atmospheric light by choosing the brightest 0.1% pixel in the depth and selecting the pixels intensity corresponding to them in the input image. Transmission map can be estimates by using Eq (3). Hence, The scene radiance can be estimated as:

$$\mathbf{R} = \frac{\mathbf{I} - L}{min(max(e^{-\beta \mathbf{D}}, 0.9), 0.1)} + L \tag{9}$$

It limits transmission map between a lower bound and upper bound to achieve natural enhancement. CAP performs better than DCP as it uses depth for transmission estimation. However, it sometimes provides limited enhancement. Since, it uses hard thresholding to limit the scene radiance.

C. RANKING CNN

Ranking Convolutional Neural Networks [11] attempts to cover the drawbacks of the results attained from [9] and [10].Any image has two features that play a major role in the formation of transmission map. They are called the structural and the statistical features. Some of the finest CNN may attain success in evaluating the structural features but most of them

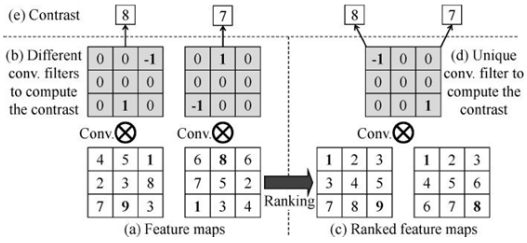


Fig. 2. Calculation of statistical feature (contrast) [11]

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:00:15 UTC from IEEE Xplore. Restrictions apply.

fall short while estimating statistical details. The Ranking layer was proposed keeping both these features in mind as it can estimate both the above said features. Ranking Layer takes feature maps as input and then return an output of ranked maps with the same shapes and sizes. The role of this layer is just to change the ordering of the input map. This aids the overall dehazing process by increasing its efficiency in calculation of features.

It can be observed from the 2 that for the calculation of the statistical feature other CNN may need many convolution filters but the ranking layer needs only one filter.This Ranking Layer is then inserted into the classical CNN to get an estimation over the pixel-wise transmission. The remaining part is a regression network followed by a Random Forest generator to increase performance. The obtained result is then passed for smoothing and rectified result is then coalesced with L, which is evaluated by calculating the 0.1% pixels having the highest intensities in the Dark Channel, to produce final results.

Ranking CNN attempts to learn haze characteristics using ranking layer. The other techniques are unable to distinguish statistical and structural details in the image. Ranking CNN evaluates statistical details with structural details from an image. While the other methods were taking many traversals of the filters to calculate the statistical details, the ordering of elements in the layer aids in the calculation of statistical details like contrast easily and does not require many traversals. This approach does not assumes that the transmission is constant locally instead it aims to find transmission at every pixel of the image, thereby increasing the overall image enhancement ability.

D. DEHAZE NET

DehazeNet [10] was proposed which being a [CNN] could be trained for memorizing haze characteristics and the effects they have in transmission. DehazeNet attempts to learn associations between hazy areas and their transmission. The attributes such as Dark Channel, Maximum Contrast, Colour Attenuation and Hue Disparity, if removed from an image makes the image haze-free. To improve the chances of convergence this method also brings with it a novel non-linear activation function known as Bilateral-Rectified Linear Unit (BReLU).

The structure of DehazeNet is a combination of many convolutional, pooling layers stacked back to back along with the presence of activation functions as shown in Fig. 3 The main feature of DehazeNet includes haze relevant characteristic extraction from the whole image using a non linear regression based on BReLU. Network learns the training parameters through Stochastic Gradient Descent (SGD) and Mean-Squared Error (MSE) is its loss function. The end result of this network is the efficient calculation of transmission map.

During the estimation of transmission near the sky region, some of the values of the outliers in CAP [8] fluctuates. However, DehazeNet is able to calculate the transmission with greater precision. Also, CAP performance degrades badly when the concentration of haze increase which is not observed in DehazeNet. Further, BReLU is an efficient activation function that helps in image restoration and conversion.

E. MULTISCALE CNN

To evaluate the transmission MSCNN [9] presented a two step approach. First, a rough outline depicting the transmission is obtained utilizing a Coarse-Scaled Network. The produced rough result containing the overall details of the image is then passed along to the second phase. Convolution, pooling, upsampling are some of the operations performed during the first phase. Second, the Fine-Scaled Network enhances the overall structure of the intermediate passed onto it. The coarse transmission is provided here as an add-on feature map. The transmission here gets refined due to the features learned by the fine-scaled model through forward and backward iterations. MSCNN model is as shown in Fig. 4.

Some methods depends on prior based knowledge and fails as compared to MSCNN because MSCNN tries to learn the haze relevant features with the help of an end to end CNN. MSCNN method includes the effects of priors and other details acquired by the forward and backward iterations of the network. Hence, it provides a better haze representation as compared to other dehazing techniques.

F. AOD NET

All-In-One Dehazing Network (AOD-Net) [12] evaluates haze-free results directly from the foggy image without separate evaluation of transmission and airlight. AOD Net changes the atmospheric model into a K-module based model. The

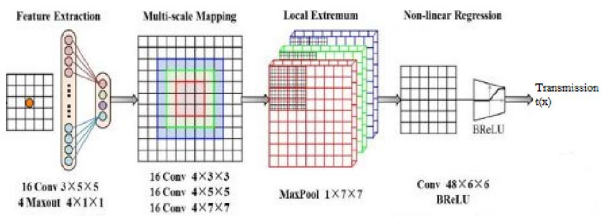


Fig. 3. Architecture of DehazeNet [10]

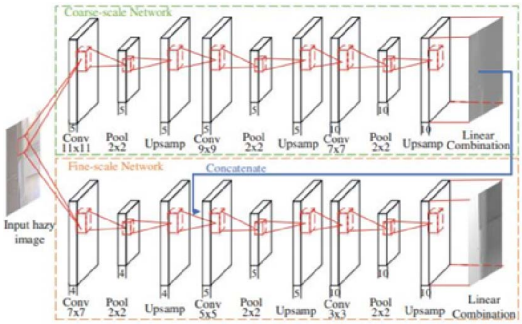


Fig. 4. MSCNN model [9]

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:00:15 UTC from IEEE Xplore. Restrictions apply.

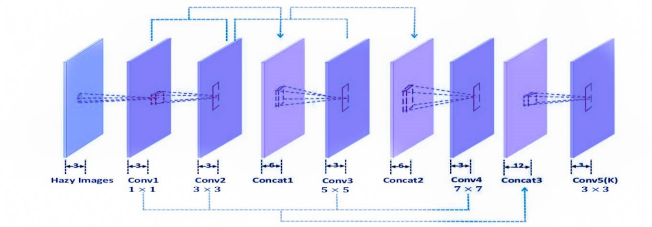


Fig. 5. Architecture of AODNet [12]

K-module based model incorporates the effect of the transmission, and airlight, Using Eq. (1).

$$\mathbf{R} = \frac{1}{\mathbf{T}}\mathbf{I} - L\frac{1}{\mathbf{T}} + L \tag{10}$$

The purpose here is to entangle L and $\frac{1}{T}$ into K.

$$\mathbf{R} = \mathbf{KI} - \mathbf{K} + b \tag{11}$$

where b has default value 1 and is a bias.Here,

$$\mathbf{K} = \frac{\frac{1}{\mathbf{T}}(\mathbf{I} - L) + (L - b)}{\mathbf{I} - 1} \tag{12}$$

After the model is converted as per the K module equation, the next step is the clean image conversion module that incorporates many pixel wise additions and multiplication layers. The results then obtained as the dehazed outputs. Because of the fact that the dehazed output is produced by a single network run, AOD Net has its application that it can be inserted with other learning based models for faster overall results. Fig. 5 shows AOD Net framework.

AOD Net is robust in cases when there are objects with colour alike atmospheric light. In those cases DehazeNet [10], MSCNN [9] show very less details. Also, the artifact problems are present in DCP [7], CAP [8], DEHAZENet [10], MSCNN [9], however AOD Net is artifact free method.

G. DCPDN (Densly Connected Pyramid Dehazing Network)

Zhang, and Patel [13] deveoped an end-to-end model for removal of haze using Generative Adversarial Network(GAN). The model learns structural relation from the images. The model consists of encoder-decoder structure which can learn transmission, airlight and can dehaze the image simultaneously. Transmission is better calculated with the help of multi-pool mapping. Also, atmospheric model has been included for better optimization for overall learning process. DCPDN has less ability for being temporally coherent. Fig. 6 represents the architecture of the model which contains 4 modules : 1. Pyramid Structure Transmission Estimation Map. 2. Atmospheric light estimation module. 3. Overall Dehazing. 4. GAN Joint Discriminator .

It optimizes GAN using the edge-preserving loss function. In practice they concatenated images free from fog with the transmission and then leveraged use of joint-discriminator loss function as:

$$\begin{aligned}\min_{G_t,G_d}\max_{D_j \text{ oint}} \mathbb{E}_{I\sim p_{\text{data }(I)}}\left[\log\left(1 - D_j \text{ oint }(G_t(I))\right)\right] + \\ \mathbb{E}_{I\sim p_{\text{data }(I)}}\left[\log\left(1 - D_j \text{ oint }(G_d(I))\right)\right] + \\ \mathbb{E}_{t,J\sim p_{\text{data }(t,J)}}\left[\log D_{\text{joint}}\left(t, J\right)\right]\end{aligned} \tag{13}$$

Several CNN-based methods like Ranking-CNN [11], MSCNN [9], restricts their ability by only taking into account of transmission map for their CNN framework whereas AOD-Net [12] solves this issue with their light-weight CNN.

H. FFA-NET

Qin *et al.* [14] presented a fusion end-to-end attention model for image dehazing. It uses Channel Attention and Pixel Attention modules which considers features based on different channels. The use of residual learning is stated to bring less vital features like thin haze. Feature Fusion Module is used to learn different features on different channels. It provides more weight to important features like thick haze regions with fusion of weights from different channels.

Fig. 7 shows Feature Attention module in FFA-Net. It consists of with basic residual learning which consists N group architectures which learns the less important features like low frequency and hazy region to avoid through multiple skip connections. The feature attention module improves the performance of the algorithm which consists of channel and pixel attention modules. Channel attention relies on the fact that the channels of image consists of different features and spatial-information is obtained by using global average pooling, with

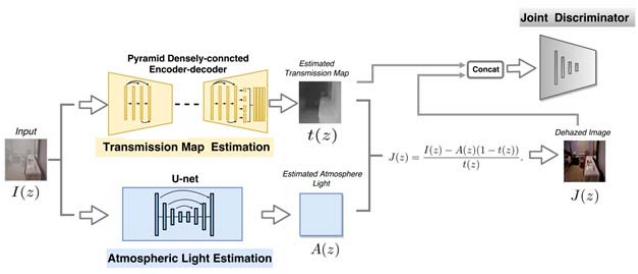


Fig. 6. Architecture of DCPDN Model [13]

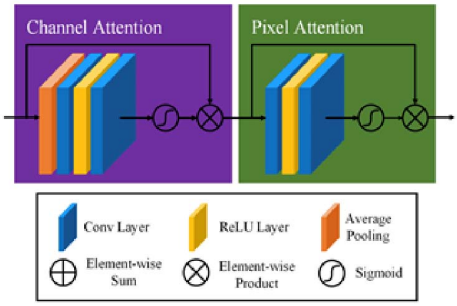


Fig. 7. Feature Attention Module [14]

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:00:15 UTC from IEEE Xplore. Restrictions apply.

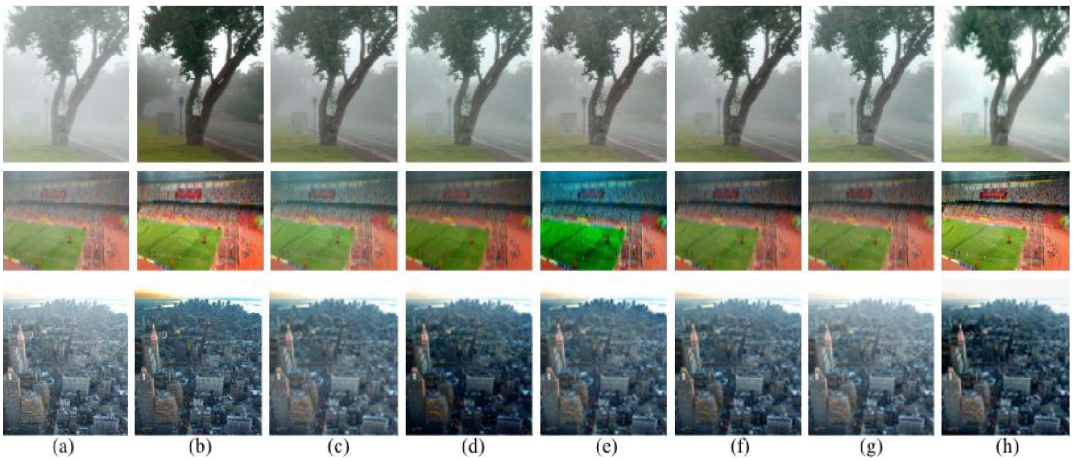


Fig. 8. Images (a) input; and result of (b) Dark Channel Prior [7]; (c) Colour Attenuation Prior [8]; (d) AOD NET [12]; (e) MSCNN [9] ; (f) Dehaze NET [10]; (g) FFA NET [14]; (h) DCPDN [13].

TABLE I
MEAN VALUE OF QUANTITATIVE MEASURES

Metrics	***He et al.'s [7]***	***Zhu et al.'s [8]***	***Cai et al.'s [10]***	***Ren et al.'s [9]***	***Li al.'s [12]***	***Zhang et al.'s [13]***	***Qin et al.'s [14]***
PSNR	17.71655	13.9468125	13.42545	14.71495	13.96785	13.06925	13.1842375
SSIM [27]	0.994475	0.98865	0.984375	0.9918375	0.9892375	0.9818625	0.9703
WPSNR [28]	20.8030625	17.9290625	16.6680625	18.91295	17.8942375	15.81585	15.12675
MSE	0.117425	0.2021125	0.2319625	0.1611125	0.175525	0.250725	0.3226125

convolution layers and element wise multiply with the weights of channels as:

$$g_{\mathrm{c}} = H_{\mathrm{p}}(F_{\mathrm{c}}) = \frac{1}{Height * Width} \sum_{i=1}^{Height} \sum_{j=1}^{Width} I_{\mathrm{c}}(i,j) \quad (14)$$

$$CA_{\mathrm{c}} = \sigma(Conv(\delta(Conv(g_{\mathrm{c}})))) \quad (15)$$

$$F^{*}{}_{\mathrm{c}} = CA_{\mathrm{c}} \otimes F_{\mathrm{c}} \quad (16)$$

,where I_c denotes $\mathrm{c_{th}}$ channel of Image with pixel position at (i,j), H_p indicates the global average pooling layer, σ is a sigmoid function, δ denotes the ReLu function, CA_c denotes the channel weights which is multiplied with input F_c.

Pixel Attention module is further used by the network to learn more information about thick hazy regions. It consists of convolution lanes and element-wise multiplication and outputs the Feature Attention module.

$$PA = \sigma(Conv(\delta(Conv(F^{*})))) \quad (17)$$

$$\tilde{F} = F^{*} \otimes PA \quad (18)$$

,where PA_c denotes output of Pixel Attention, which is further element-wise multiplied to obtain Feature Attention .

Previous CNN-based models like Ranking-CNN [11], DehazeNet [10], MSCNN [9], AOD-Net [12] handle the attention modules for channel and pixel-wise as uniform, but, as thin haze and thick haze should have different weighting factors and with the heterogeneous nature of haze distribution along different channels are also taken in consideration.

IV. QUALITATIVE AND QUANTITATIVE ANALYSIS

This section presents the analysis of above-discussed methods. For computing differences and evaluate efficiencies among different dehazing methods, both quantitative and qualitative analysis are necessary. Fig. 8 depicts the outputs that are produced by distinct dehazing techniques. Fig. 8(a) shows the input image with haze. Fig. 8(b-h) represents output of He [7], Zhu [8], Li [12], Ren [9], Cai [10], Qin [14], Zhang [13] respectively. It can be noted that Fig. 8(b) is removing haze efficiently but at times suffer from halo-artifacts near the boundaries of the subject such as tree or building. Fig. 8(c) does not suffer from excess saturation and brings out the real colours. The outputs of Fig. 8(d) have sharper details and maintains the colour accuracy of the image.In the presence of objects having colour close to white light, this method shows robustness and performs efficiently. Fig. 8(f) eradicates haze nicely and the colour distribution is also uniform thereby producing good results. But as the number of hazy images increases, the visual evaluation becomes harder and biased. Therefore it is required to present the quantitative evaluation of the above-mentioned methods. Learning-based techniques require a lot of data for learning complex features so after research, the datasets that are used in these methods for learning haze relevant features are: RESIDE [29], NYU Depth v2 [30], Middlebury Stereo [31], NTIRE [32], FRIDA [33], I-HAZE [34], and O-HAZE [35]. The codes and weights given by authors are used for fair comparison. Further, the image quality is tested using metrics such as Structure Similarity In-

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:00:15 UTC from IEEE Xplore. Restrictions apply.

dex(SSIM) [27], Peek Signal-to-Noise Ratio(PSNR), Weighted Peek-Signal-to-Noise Ratio(WPSNR) [28] , and Mean Squared Error. Since these are widely used metrices for analyzing image dehazing methods. The synthetic image data is used from the project sites of [36], [37] to gather the result with application of those metrics. Quantitative results are compared in Table 1. The value of these metrics are averaged out for the images of data set.

V. CONCLUSION

After analyzing Image dehazing approaches, it can be observed that each approach contains their own limitations. This paper depicts an extensive inspection of various state of art image dehazing approaches. It is observed that the statistical methods fail in regions where the object's colour was identical to the atmospheric light. Statistical methods performs better in quantitative analysis. However, in Qualitative analysis these methods shows various artifacts. The algorithms either produced a halo effect or erased the object. The learning based methods performed efficiently on most of the images. The contrast of learning based methods is similar to the statistical methods. The learning based approach performs appreciably in qualitative analysis. The performance of statistical methods can be improved using improved prior knowledge and parameters. Learning-based methods can overfit on a dataset thus, it must consider the overfitting conditions appropriately.

REFERENCES

[1] A. S. Parihar and O. P. Verma, "Contrast enhancement using entropy-based dynamic sub-histogram equalisation," *IET Image Processing*, vol. 10, pp. 799–808(9), November 2016.

[2] A. S. Parihar, "Histogram modification and dct based contrast enhancement," in *2018 2nd International Conference on Inventive Systems and Control (ICISC)*, 2018, pp. 614–618.

[3] A. S. Parihar, "Fuzzy adaptive gamma correction for contrast enhancement," in *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, 2017, pp. 625–629.

[4] A. S. Parihar, O. P. Verma, and C. Khanna, "Fuzzy-contextual contrast enhancement," *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1810–1819, 2017.

[5] A. S. Parihar, O. P. Verma, and D. Yadav, "Image contrast enhancement using differential evolution," in *Advances in Communication, Devices and Networking*. Singapore: Springer Singapore, 2018, pp. 517–526.

[6] E. J. McCartney, "Optics of the atmosphere: scattering by molecules and particles," *New York, John Wiley and Sons, Inc., 1976. 421 p.*, 1976.

[7] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010.

[8] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Transactions on image processing*, vol. 24, no. 11, pp. 3522–3533, 2015.

[9] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *European conference on CV*. Springer, 2016, pp. 154–169.

[10] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An end-to-end system for single image haze removal," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187–5198, 2016.

[11] Y. Song, J. Li, X. Wang, and X. Chen, "Single image dehazing using ranking convolutional neural network," *IEEE Transactions on Multimedia*, vol. 20, no. 6, pp. 1548–1560, 2017.

[12] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," in *IEEE Conference on Computer Vision*, 2017, pp. 4770–4778.

[13] H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," in *IEEE Conference on CVPR*, 2018, pp. 3194–3203.

[14] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "Ffa-net: Feature fusion attention network for single image dehazing," *arXiv preprint arXiv:1911.07559*, 2019.

[15] K. Singh and A. S. Parihar, "A comparative analysis of illumination estimation based image enhancement techniques," in *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, 2020, pp. 1–5.

[16] A. S. Parihar and K. Singh, "A study on retinex based method for image enhancement," in *2018 2nd International Conference on Inventive Systems and Control (ICISC)*, Jan 2018, pp. 619–624.

[17] A. S. Parihar and K. Singh, "Illumination Estimation for Nature Preserving low-light image enhancement," *10.36227/techrxiv.12236780.v1*, 5 2020. [Online]. Available: https://www.techrxiv.org/articles/Illumination_Estimation_for_Nature_Preserving_low-light_image_enhancement/12236780

[18] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. Torr, "Fully-convolutional siamese networks for object tracking," pp. 850–865, 2016.

[19] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," pp. 779–788, 2016.

[20] S. Ji, W. Xu, M. Yang, and K. Yu, "3d convolutional neural networks for human action recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 221–231, 2012.

[21] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, "Instant dehazing of images using polarization," in *IEEE Conference on CVPR*, vol. 1. IEEE, 2001, pp. I–I.

[22] Y. Y. Schechner, S. G. Narasimhan, and S. Nayar, "Polarization-based vision through haze," *Applied optics*, vol. 42, no. 3, pp. 511–525, 2003.

[23] J.-P. Tarel and N. Hautiere, "Fast visibility restoration from a single color or gray level image," in *IEEE Conference on Computer Vision*, 2009, pp. 2201–2208.

[24] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *IEEE conference on computer vision*, 2013, pp. 617–624.

[25] C. O. Ancuti and C. Ancuti, "Single image dehazing by multi-scale fusion," *IEEE Transactions on Image Processing*, vol. 22, no. 8, pp. 3271–3282, 2013.

[26] C.-C. Tsai, C.-Y. Lin, and J.-I. Guo, "Dark channel prior based video dehazing algorithm with sky preservation and its embedded system realization for adas applications," *Optics express*, vol. 27, no. 9, pp. 11 877–11 901, 2019.

[27] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[28] J. Mannos and D. Sakrison, "The effects of a visual fidelity criterion of the encoding of images," *IEEE transactions on Information Theory*, vol. 20, no. 4, pp. 525–536, 1974.

[29] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 492–505, 2018.

[30] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from rgbd images," in *European conference on CV*. Springer, 2012, pp. 746–760.

[31] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *IEEE Conference on CVPR*. IEEE, 2007, pp. 1–8.

[32] C. Ancuti, C. O. Ancuti, and R. Timofte, "Ntire 2018 challenge on image dehazing: Methods and results," in *IEEE Conference on CVPR Workshops*, 2018, pp. 891–901.

[33] J.-P. Tarel, N. Hautiere, L. Caraffa, A. Cord, H. Halmaoui, and D. Gruyer, "Vision enhancement in homogeneous and heterogeneous fog," *IEEE Intelligent Transportation Systems Magazine*, vol. 4, no. 2, pp. 6–20, 2012.

[34] C. O. Ancuti, C. Ancuti, R. Timofte, and C. D. Vleeschouwer, "I-haze: a dehazing benchmark with real hazy and haze-free indoor images."

[35] C. O. Ancuti, C. Ancuti, R. Timofte, and C. De Vleeschouwer, "O-haze: a dehazing benchmark with real hazy and haze-free outdoor images," in *IEEE CVPR*, 2018, pp. 754–762.

[36] R. Fattal, "Dehazing using color-lines," *ACM transactions on graphics (TOG)*, vol. 34, no. 1, pp. 1–14, 2014.

[37] D. Berman, S. Avidan *et al.*, "Non-local image dehazing," in *Proceedings of the IEEE conference on CVPR*, 2016, pp. 1674–1682.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:00:15 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

A Detailed Study on Generative Adversarial Networks

Shailender Kumar
Dept. of Computer Science and Engineering
Delhi Technological University
Delhi, India
shailenderkumar@dce.ac.in

Sumit Dhawan
Dept. of Computer Science and Engineering
Delhi Technological University
Delhi, India
sumitdhawan6@gmail.com

***Abstract*—From past decades, along with increase in computing power, different generative models have also been developed in the area of machine learning. Among all such generative models, one very popular generative model, called as Generative Adversarial Networks, has been introduced and researched upon, in past few years only. It is based on the concept of two adversaries constantly trying to outperform each other. Objective of this review is to extensively study the GAN related literature and provide a summarized form of the studied literature available on GAN, including the concept behind it, its objective function, proposed modifications in base model, and recent trends in this field. This paper will help in giving a thorough understating of GAN. This paper will give an overview of GAN and discuss the popular variants of this model, its common applications, different evaluation metrics proposed for it, and finally its drawbacks, conclusion of the paper and future course of action.**

Keywords—deep learning; generative adversarial networks; neural network; generative model; image processing

I. INTRODUCTION

There has been many advancements in the field of machine learning, especially deep learning as more and more computing or processing power is becoming available. Deep Learning helps us in extracting high level, useful and abstract features from the input data and using those features in classification and generation tasks. This approach is commonly known as representation learning and is based on how the human mind learns anything. Concept of generative models (based on deep learning) forms the basis of Generative Adversarial Networks (or GANs). Traditional generative models like Restricted Boltzmann machine [15] and Variational Auto Encoder [14] were based on concepts like Markov chains and maximum likelihood estimation. Based on the distribution of input data, they estimate the distribution of generated data, but as a result of not being good in generalization, their performance and outcome is affected. A new concept in the field of generative models, called as Generative Adversarial Networks (GANs) was introduced in the year of 2014 by Goodfellow et al. [1]. It consists of one generator and one discriminator, which are adversaries of each other and thus, constantly trying to outperform each other and thus, improving themselves in the process of doing so. GANs are based on learning the joint probability distribution.

Image Synthesis is a problem which has been researched a lot upon. It means generating an image based on the hidden and visible features of an already existing image. GANs are being used largely in the area of image synthesis (or image processing, in general) as they have been proven to work really well with the images. GANs consist of two models which are trained simultaneously against each other. Generator is tasked with generating new data points based on how the data points are distributed in input samples and thus, deceiving the Discriminator with the generated/fake data points as real data points. Discriminator is tasked with catching the bluff of Generator by classifying the received data points as generated or a real (taken directly from the input sample space). It is similar to a zero-sum game between two opponents. Models are trained using back-propagation [17] and dropouts (to prevent over-fitting).

GANs can be classified based on their architecture and loss function used to train the generator [16]. It has been shown that GANs can produce pretty good and high quality images which seem convincing enough to be considered as real images. GANs have grown exponentially since they were first proposed in 2014. As of now, there are many variants of generative adversarial networks, each created specifically for a specific problem in multiple research areas like image synthesis, style transfer, image enhancement, image to text conversion, text to image conversion, detecting an object, etc., making GANs a hot topic for research purpose as well as in real world applications. Despite all the growth and research going on, GANs suffer from some problems and have shortcomings [9] like vanishing gradient, mode collapse, non-convergence, absence of universal performance evaluation metrics. Also, various solutions have been proposed for above mentioned drawbacks.

This survey analyses and presents theory, applications, shortcomings, state of the art variants of GANs and recent trends in this field. This survey is organized in parts as follows. Section II defines the background, structure and loss function of GANs. A comparison of variants of GANs is presented in Section III. Section IV discusses different evaluation metrics. The applications of GANs are given in Section V. Section VI

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:56:00 UTC from IEEE Xplore. Restrictions apply.

discusses the shortcomings of GANs. Finally, Section VII gives the conclusions and future scope.

II. THEORY AND STRUCTURE OF GAN

A. Basic Theory

Behind the motivation for GANs, is Nash equilibrium of Game Theory [1], which is a solution for a non-cooperative game between two adversaries, in which each player already knows all the strategies of other player, therefore no player gains anything by modifying their own strategy.

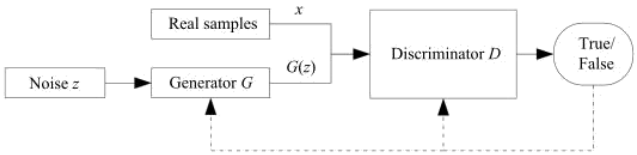


Fig. 1. Structure of Generative Adversarial Networks

The main aim of GAN is to arrive to this Nash equilibrium. Fig. 1 depicts the basic structural model of a standard GAN. Any function which can be differentiated can be used as the function for equations of generator as well as discriminator. Here, G and D are two differentiable functions that represent the generator and the discriminator respectively. Inputs given to D is x (real data), and z (random data or simply, noise). Output of G is fake data produced as per the probability distribution of actual data (or pdata), which is termed as G(z). If actual data is given as input to Discriminator, it should classify the input data as real data, labeling it 1. If fake or generated data is given as input to Discriminator, it should classify the input data as fake data labeling it 0. Discriminator strives to classify the input data correctly as per the source of data, On the other hand, Generator strives to deceive the Discriminator by making generated data G(z) similar and in line with the real data x. This game like adversarial process helps in improving the performance of both Discriminator and Generator slowly and gradually throughout the process. Therefore, slowly, Generator is able to generate better images that look more real, because it has to fool the improved and more efficient Discriminator (when compared to the previous iteration of training).

B. Loss Function

In this part of the section, loss function and learning methodology of GANs will be discussed. Following is the mathematical representation of loss function (or objective function) of GANs, as discussed in [1].

$$min_G max_D V(D, G) = E_{x\sim p(x)}[log\, D(x)] + E_{x\sim p(z)}[log(1-D(G(z)))] \quad (1)$$

where,
x - Obtained by sampling the distribution of real data or p(x).
z - Obtained by sampling the distribution of prior data or p(z) (which can be Gaussian or Uniform distribution).
E[x] - Expected value of any random variable x.
D(x) - The probability that x is sampled from actual data and not from the generated data.

Main aim of Discriminator is to maximize the objective function by minimizing D(G(z)) along with maximizing D(x), which signifies Discriminator is classifying real data as real and fake data as fake. Main aim of Generator is to minimize the objective function by maximizing D(G(z)) which signifies that Discriminator is classifying generated data (which is fake) as real. The training data consists of real data as well as generated data. Discriminator is initially trained on real as well as fake data. After that, its weights are made non trainable. Then the generator is trained. This process is repeated until required level of performance is achieved. In initial phase of learning, due to poor performance of Generator, Discriminator can catch the lie and reject generated samples easily as they are easily distinguishable from the real samples. Therefore, saturation of log(1 − D(G(z))) happens and derivative tends to 0, preventing gradient descent from occurring. So, to correct this, Generator is now trained to maximize the value of log(D(G(z))) instead of minimizing the value of (1 − D(G(z))). This modified objective function results in the derivative to be high and provides much better gradients in the initial phases of learning

III. GAN MODELS

Since the idea of GAN was proposed, many variations have been done in the original model, resulting in different models of GANs. These variations include the improvement of structure for efficiency, changes done for any specific application or problem statement like, style transfer from one image into another [18], image enhancement [13], completing an incomplete image [20-21], and generating an image from text [19].

A. Fully Connected or Vanilla GAN

It was first introduced as a baseline model for GANs in 2014 by Goodfellow et al. [1]. As the overlap of probability distributions of actual data and fake data is very little, the measure of similarity between two distributions, also known as the Jensen-Shannon divergence, may become constant, which results in the problem of vanishing gradient. It doesn't perform that well in the case of more complex images.

B. Deep Convolutional GAN (DCGAN)

Another variant of GANs was proposed by Radford et al. (2015) [4], known as Deep Convolutional Generative Adversarial Networks (DCGANs), which is based on Convolutional Neural Networks with following changes:

- Removing fully connected hidden layers.
- In the generator, a fractional-strided convolutions are used instead of pooling layers. In the discriminator, strided convolutions are used instead of pooling layers.
- Using the concept of batch normalization in generator as well as discriminator.
- Applying ReLU activation function in all but the last layer of generative model. LeakyReLU activation function is applied in each and every layer of the discriminator.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:56:00 UTC from IEEE Xplore. Restrictions apply.

C. Wasserstein GAN (WGAN)

Wasserstein GAN (WGAN) was proposed by Arjovsky et al. [5] to resolve the issue of vanishing gradient. Instead of the Jensen-Shannon divergence, it uses the Earth-Mover distance, for calculating the similarity between probability distribution of real or actual data and fake or generated data. Discriminator is represented using f, which is a critic function, and is built on Lipschitz constraint. WGAN, though useful in a comparatively stable training of GANs, still produces data samples which are low in quality and even sometimes fail in the process of converging.

D. Wasserstein GAN with Gradient Penalty (WGAN-GP)

This model can be considered as an improvement to the basic WGAN, proposed by Gulrajani et al. [12]. They found that the weight clipping in GAN, which is done in order to apply a Lipschitz constraint on the critic, leads to failure in training. This causes abnormal behavior. So an alternative was proposed to apply the Lipschitz constraint on critic, that is, it works by penalizing the value of norm of the gradient of the f (or critic function) with respect to the input, instead of clipping of weights. This method tends to converge quickly and generates samples of relatively high quality when compared to WGAN version with weight clipping.

E. Least Square GAN (LS-GAN)

In conventional GANs, during the process of back-propagation, sigmoid cross entropy loss function is used. However, sometimes, during the learning process, the issue of vanishing gradient occurs due to this particular loss function. As a solution, Mao et al. [22] came up with LS-GAN. It uses the least square loss function. LS-GANs generate higher quality images and also are more stable during the whole learning process compared to normal GANs.

F. Semi-GAN (SGAN)

In basic GAN, during training, only one class label is required which is used to specify data source (real or fake). Odena [10] proposed Semi GAN (SGAN) by further adding class labels (say, N) of actual data while discriminator D is trained. After training, D will classify the input to one of the N+1 classes, where the extra class is used to specify the data source as per G. It is shown that this method can create a better classifier which generates samples of high quality when compared to standard GAN, as it is trained to learn the features and associates them with correct class labels.

G. Conditional GAN (C-GAN)

In conventional GAN, only latent space is provided to generator. The conditional GAN, as proposed by Mirza [11] changes that by providing an additional parameter (label y) along with latent space, to the generator and train it to produce corresponding images. Discriminator is provided with real images and label as its input, to distinguish real images better. It is shown that this model can produce digits similar to those of MNIST dataset, conditioned on class labels (0, 1, 2, 3…9). Hence, the name is Conditional GAN.

H. Bidirectional GAN (BiGAN)

In normal GANs, the generator learns by mapping latent feature vector to real data probability distribution. But it lacks an efficient mechanism which maps the real data to data in latent space. Bidirectional GANs (BiGANs) was proposed by Donahue et al. [23]. It maps the actual data probability distribution on to latent space, hence helping to learn to how to extract relevant features.

IV. EVALUATION METRICS

Apart from the manual inspection and evaluation, of samples produced by Generator module of GANs, there are some quantitative measures like Average Log Likelihood [24], Inception Score (IS) [9], Wasserstein metric [5], Frechet Inception Distance (FID) [25], etc. IS should be high and FID should be low for high quality generated images. Based on these evaluation metrics, GAN model's performance can be better judged and thus can be improved by introducing the required modifications in model.

Log-likelihood (also called as Kullback-Leibler divergence or KL divergence) is one of the widely used standard metrics for evaluating and measuring the performance of generative models [24]. It helps in measuring the likelihood of generated distribution being consistent with actual data distribution. Maximum likelihood or 0 KL divergence will produce perfect samples which will seem real.

Inception Score is probably one of the most used evaluation metric for GAN evaluation, and it was proposed by Salimans [9]. A pre-trained neural network (for example, Inception v3 model) is used to capture the relevant and important properties of generated samples, i.e., image quality and image diversity. High image quality is represented by narrow distribution of label and high image diversity is represented by uniform distribution of labels of all samples combined. As both distributions are very much dissimilar (one is narrow, another is uniform), KL divergence should be high, further giving good Inception score.

$$KL\ divergence = KL(C||M) = p(y|x) * (log(p(y|x)) - log(p(y))) \quad (2)$$

where,
C is conditional probability distribution, p(y|x) and,
M is marginal probability distribution, p(y)

Inception score doesn't measure the variations in fake images if contrasted with actual images, therefore Heusel et al. [25] introduced Frechet Inception Distance. It works on the features and transforms generated data and maps it to feature space. This feature space is obtained by the last pooling layer, just before output classification. It is summarized by calculating the covariance as well as mean for fake and actual data. By calculating this Frechet Inception Distance (also known as Wasserstein-2 distance) between these Gaussian distributions, quality and variation of produced data is measured.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:56:00 UTC from IEEE Xplore. Restrictions apply.

As of yet, there is no unanimous decision regarding which one is a better evaluation measure. Different scores focus on different aspects of the image generation process. But some metrics seem more plausible than others, like FID has the advantage of being more robust to noise. And as claimed in [25], FID can tell us how similar are real and fake images, and this measure is considered more efficient than IS.

V. APPLICATIONS OF GANS

GANs are being used widely in fields like Image processing and Natural Language Processing. Following are few of the main applications of GANs apart from classical application of generating data similar to the input data.

1) Inter-domain conversion of content of one image to another using CGAN was proposed by Isola et al. [18]. It is named pix2pix which is effective at generating photos from label maps and colorizing images.
2) Zhu et al. [18] propose CycleGAN, which can also be used for translating image from one domain to another, in the absence of a training example of paired images. It can perform style transfer, photo enhancement, object transfiguration, attribute transfer, etc.
3) Solution to the problem of Image super-resolution was provided in [13], where Ledig et al. present SRGAN. VGG network [26] is used as the discriminator. It is shown in [13] that it is possible to generate photorealistic image with high resolution. But the texture information which SRGAN generates, doesn't seem that much real, and is noisy. Therefore, Wang et al. [29] propose an Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) which aims to improve the network's architecture.
4) Zhang et al. [28] propose that GAN can generate such sentences which seem too real by utilizing the concept of long short term memory (LSTM). Li et al. [30] use GANs to perform speech to text conversion by capturing the relevance of the dialogue. SeqGAN [31] uses reinforcement learning to not only generates speech but even poems and music.
5) Antipov et al. [27] introduced a model on the concept of GAN that can be used for automatic face aging. It produces synthetic images. Facial attributes are modified in similar kinds of work but in this, emphasis is given on preserving basic facial attributes and original person's identity in aged version of the face.



Fig. 2. Few examples of applications of GANs [8]

VI. DRAWBACKS

The main shortcoming in the training process of GANs is that there is always a risk of mode collapse. It occurs when data produced by GANs is mostly focused on very less modes or sometimes even on one single mode, which results in comparatively less diverse samples. A solution to the above problem is batch regularization or mode regularization. In this, decent size of batches of data samples or various data points are included so that the diversity is improved. Another solution is proposed in [33], in which the samples generated by different models are combined together. Moreover, optimizing the objective function (like done in WGAN [5]) can resolve the problem as well.

The lack of stability in the process of training is a big issue in itself. The GAN model's parameters oscillate, destabilize and never converges to the Nash equilibrium, that is each opponent always countermeasures other opponent actions, making the models harder to converge. In contrast to the other generative models, GANs evaluation problem is much more challenging as there is no consensus on what evaluation metrics should be used for GANs [34]. Another shortcoming of GANs is the problem of diminished gradient. It occurs if the discriminator gets very powerful while training, causing the gradient of generator's function to reduce and slowly vanish, and in turn, the generator learns nothing. This imbalance between models of generator and discriminator, results in overfitting. Also, GANs are very much sensitive to the selection of hyper parameters.

VII. CONCLUSION AND FUTURE SCOPE

This paper analyses and summarizes the background of GANs, its basic theory, structure, variants, evaluation metrics, applications, shortcomings and future scope. It gives an overall and broad understanding of the literature present on GAN. A wide variety of applications of GANs is shown in this paper. New as well as better solutions to new and existing challenges in GANs are yet to be explored with an aim to increase the efficiency of GANs. The field of GAN is a promising research area, but even with all the development in its research area, it has its own problems like unstable training, non-convergence, lack of consensus on an evaluation metric, requirement of more computing power and the model's complexity. In conclusion, GAN is an interesting and useful research field with many applications but also, a lot of work needs to be done to overcome existing challenges, as it is still comparatively a new field.

New work is being continuously done to overcome the limitations of GANs. For example, WGAN can resolve the mode collapse problem as well as the training instability problem, but only partially. Therefore, preventing the problem of mode collapse in GANs remains an open research problem. Also, other research areas include the presence of Nash equilibrium and the theory of the convergence of a GAN model. GANs are being widely utilized in the area of Computer Vision but comparatively not that much in other fields like Natural Language Processing. This limitation is

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:56:00 UTC from IEEE Xplore. Restrictions apply.

because of the different properties related to image and non-image data. GANs can be used for interesting applications in various fields, therefore research is going on for that and also on how to increase the efficiency and improve the performance of GAN.

REFERENCES

[1] I. Goodfellow et al., ‘‘Generative adversarial nets,’’ in Proc. Adv. Neural Inf. Process. Syst., 2014, pp. 2672–2680. [Online]. Available: http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf.

[2] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. New York, USA: MIT Press, 2016.

[3] I. Goodfellow, “NIPS 2016 tutorial: generative adversarial networks,” arXiv: 1701.00160, 2016.

[4] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” arXiv: 1511.06434, 2015.

[5] Arjovsky, Martin, Soumith Chintala, and Léon Bottou. "Wasserstein gan." arXiv preprint arXiv:1701.07875 ,2017.

[6] L. J. Ratliff, S. A. Burden, and S. S. Sastry, “Characterization and computation of local Nash equilibria in continuous games,” in Proc. 51st Annu. Allerton Conf. Communication, Control, and Computing (Allerton), Monticello, IL, USA, 2013, pp. 917−924.

[7] Hitawala, Saifuddin. "Comparative Study on Generative Adversarial Networks." arXiv preprint arXiv:1801.04271, 2018.

[8] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In CVPR, 2017

[9] Salimans, Tim, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. "Improved techniques for training gans." In Advances in Neural Information Processing Systems, pp. 2234-2242. 2016.

[10] A. Odena, “Semi-supervised learning with generative adversarial networks,” arXiv: 1606.01583, 2016.

[11] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” arXiv: 1411.1784, 2014.

[12] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville. “Improved Training of Wasserstein GANs”. arXiv preprint arXiv:1704.00028, 2017

[13] Ledig, Christian, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew P. Aitken et al. "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network." In CVPR, vol. 2, no. 3, p. 4. 2017.

[14] Kingma, Diederik P., and Max Welling. “Auto-Encoding Variational Bayes.” ArXiv:1312.6114 [Cs, Stat], May 2014. arXiv.org, http://arxiv.org/abs/1312.6114.

[15] Fischer, Asja, and Christian Igel. “An Introduction to Restricted Boltzmann Machines.” Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, edited by Luis Alvarez et al., Springer, 2012, pp. 14–36. Springer Link.

[16] Wang, Zhengwei, et al. “Generative Adversarial Networks: A Survey and Taxonomy.” ArXiv:1906.01529 [Cs], June 2019. arXiv.org, http://arxiv.org/abs/1906.01529.

[17] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, ‘‘Learning representations by back-propagating errors,’’ Nature, vol. 323, pp. 533-536, Oct. 1986.

[18] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” arXiv preprint arXiv:1703.10593v6, 2017.

[19] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, “Generative adversarial text to image synthesis,” arXiv preprint arXiv:1605.05396, 2016.

[20] Z. Chen, S. Nie, T. Wu, and C. G. Healey, “High resolution face completion with multiple controllable attributes via fully end-to-end progressive generative adversarial networks,” arXiv preprint arXiv:1801.07632, 2018.

[21] Y. Li, S. Liu, J. Yang, and M.-H. Yang, “Generative face completion,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3911–3919.

[22] Mao, Xudong, et al. “Least Squares Generative Adversarial Networks.” ArXiv:1611.04076 [Cs], Apr. 2017. arXiv.org, http://arxiv.org/abs/1611.04076.

[23] Donahue, Jeff, et al. “Adversarial Feature Learning.” ArXiv:1605.09782 [Cs, Stat], Apr. 2017. arXiv.org, http://arxiv.org/abs/1605.09782.

[24] Hamid Eghbal-zadeh, Gerhard Widmer, “Likelihood estimation for generative adversarial networks”, 2017.

[25] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, Gans trained by a two time-scale update rule converge to a local nash equilibrium, in: Advances in Neural Information Processing Systems, 2017, pp. 6629–6640.

[26] Simonyan, Karen, and Andrew Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition.” ArXiv:1409.1556 [Cs], Apr. 2015. arXiv.org, http://arxiv.org/abs/1409.1556.

[27] Antipov et al., “Face aging with conditional generative adversarial networks”, 2017.

[28] Y. Z. Zhang, Z. Gan, and L. Carin, “Generating text via adversarial training,” Proc. Workshop on Adversarial Training, Barcelona, Spain, 2016

[29] Wang, Xintao, et al. “ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks.” ArXiv:1809.00219 [Cs], Sept. 2018. arXiv.org, http://arxiv.org/abs/1809.00219.

[30] J. W. Li, W. Monroe, T. L. Shi, S. Jean, A. Ritter, and D. Jurafsky, “Adversarial learning for neural dialogue generation,” arXiv: 1701.06547, 2017.

[31] L. T. Yu, W. N. Zhang, J. Wang, and Y. Yu, “SeqGAN: sequence generative adversarial nets with policy gradient,” arXiv: 1609.05473, 2016.

[32] Killoran, Nathan, et al. “Generating and Designing DNA with Deep Generative Models.” ArXiv:1712.06148 [Cs, q-Bio, Stat], Dec. 2017. arXiv.org, http://arxiv.org/abs/1712.06148.

[33] A. Ghosh, V. Kulharia, V. P. Namboodiri, P. H. S. Torr, and P. K. Dokania, ‘‘Multi-agent diverse generative adversarial networks,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 8513–8521

[34] Borji, Ali. “Pros and Cons of GAN Evaluation Measures.” ArXiv:1802.03446 [Cs], Oct. 2018. arXiv.org, http://arxiv.org/abs/1802.03446.

[35] Vijayakumar, T. (2019). Comparative study of capsule neural network in various applications. Journal of Artificial Intelligence, 1(01), 19-27

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:56:00 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

Adaptive Droop based Control for Active Power Sharing in Autonomous Microgrid for Improved Transient Performance

Surya Prakash, Student Member IEEE, Vaibhav Nougain, and Sukumar Mishra, *Senior Member, IEEE*

***Abstract*—This paper presents a decentralised adaptive droop based control for active power sharing between parallel inverters in autonomous microgrids. The work is focused on improving the transient performance of the system proposing a robust controller. The analysis does not require load current measurement where the Lyapunov analysis based control takes load current as a disturbance to the system. The mathematical analysis takes a more practical form of *LC* filter configuration where the dynamics due to resistances across the *LC* is also taken into consideration. Controller gain conditions are taken into consideration employing a detailed Lyapunov analysis. A comparative study of the conventional droop control with the proposed adaptive droop control is included in the work. MATLAB has been employed as the simulink platform for the comparative study. In addition, for experimental verification, FPGA based hardware setup has been developed and the related results are presented to verify the real-time operation of proposed adaptive control.**

***Index Terms*—Point of Common Coupling (PCC), Low Voltage (LV), Medium Voltage (MV), Backstepping Control, Voltage Source Converter (VSC), Phase Locked Loop (PLL), Constant Power Load (CPL), Adaptive Droop**

NOMENCLATURE

p	Instantaneous active power
q	Instantaneous reactive power
P	Filtered active power
Q	Filtered reactive power
V_{odq}	PCC point voltage in *dq* frame of reference
V^*_{odq}	Virtual PCC point voltage in *dq* frame of reference
V_{idq}	Inverter output voltage in *dq* frame of reference
I_{idq}	Inverter current in *dq* frame of reference
I^*_{idq}	Virtual Inverter current in *dq* frame of reference
I_{odq}	Load current in *dq* frame of reference
L_f	Filter Inductance
R_f	Resistance associated with the filter inductance
C_f	Filter Capacitance
R_d	Equivalent leakage Resistance associated with the filter capacitance
ω^*	Nominal system frequency
ω_o	Reference system frequency
m	Static droop for Real Power sharing
α	Adaptive droop for Real Power sharing
e_1, e_2	Current Tracking Errors
e_3, e_4	Voltage Tracking Errors

This work was supported in part by the Science & Energy Research Board (SERB), Department of Science & Technology (DST), Govt. of India through the Distinguished Investigator Award (DIA) scheme (File number - DIA/2018/000032).

The authors are with the Department of Electrical Engineering, Indian Institute of Technology Delhi, India. (e-mail: eeesuryaprakash@gmail.com, nougainvaibhav@gmail.com and sukumar@ee.iitd.ernet.in)

I. INTRODUCTION

THE penetration of power electronic interfaces in the power system has changed the course of how electricity is perceived. Power electronic interfaces have made it possible for the renewable sources (solar, wind, fuel cell etc.) to be either interface with the conventional utility grid or to be able to get locally consumed upon generation forming an autonomous microgrid [1]. The sources of such microgrid, i.e., distributed generators (DGs) are interfaced with Voltage source converters (VSCs). The VSCs, running in parallel, is responsible for controlling the frequency and voltage profile of the system due to the absence of grid [2]. The multi-DG configuration, in parallel shares the total power of the microgrid. The discussed control objective could be met by employing centralised [3], distributed [4] or decentralised methods [5].

The primary challenge for the autonomous microgrid is maintaining voltage and frequency of operation for proper operation of the loads. Once the primary challenge is dealt, the secondary challenge surfaces which are meeting the load demand in a microgrid. Once the secondary problem is addressed, the tertiary and final challenge is, maintaining the power-sharing between different sources proportionately to ensure the sources not hitting their current limit for their operational longevity. The droop control based power-sharing is widely accepted due to its decentralised nature [6]-[7]. However, droop based approach has the inherent problem of a trade-off between the voltage regulation and power-sharing accuracy [8]. To elaborate, a larger value of droop gives better power-sharing. However, the oscillations increase for such cases [9]. Transient operations in power-sharing in such microgrid setups cause overstressing, which affects the life of DG units. Mitigation of such transients is necessary to ensure the longevity of VSCs [10]-[11]. The transient and steady-state performance of the system is dependent on the control architecture employed. The conventional Proportional-Integral (PI) control has **I** (Integral) which is responsible for the robust steady-state performance. However, it is also responsible for the sluggish transient performance of the controller. The hierarchical droop based control [12]-[14] has been discussed employing radial basis

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Exeter. Downloaded on May 05,2020 at 15:09:36 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

function neural network (RBFNN). However, such schemes are sluggish due to multiple control layers. The idea of single control layer modern non-linear controllers have surfaced in the power domain for improved transients performance of such microgrid setups. Sliding mode control (SMC) [15]-[16] has been proposed to improve the transient response of the system. However, SMC inherently suffers from the problem of chattering along the sliding surface. The authors in [17] employ Adaptive Neural Network (ANN) control to conclude improved transient performance. However, the gain re-tuning for different configurations is a problem of control in [17]. The authors in [18] introduce a robust Takagi-Sugeno (T-S) fuzzy control for a parallel islanded AC microgrid. However, the complexity of such control makes its implementation difficult [19].

The idea of the proposed work is improvement of transient performance for active power sharing employing an Adaptive control. Virtual inputs [20] are introduced to frame a non-linear systematic recursive control design. The transient performance of such virtual input based control has been validated in [21]. The proposed adaptive control is a simple, robust approach without the problem of chattering as in SMC. A detailed Lyapunov analysis is used to design the gain conditions for the proposed control. The performance of the proposed adaptive control is tested under dynamic AC (RLC) transients, CPL transients and fault transients.

The rest of the paper is organised as follows. Section II presents the system architecture of the autonomous microgrid. Section III gives the modelling of parallel VSCs, including system dynamics, control problem formulation, control synthesis, generation of firing signal and tuning of control gains. Section IV gives the simulation results. Section V presents the hardware results for the proposed adaptive control.

II. System Architecture

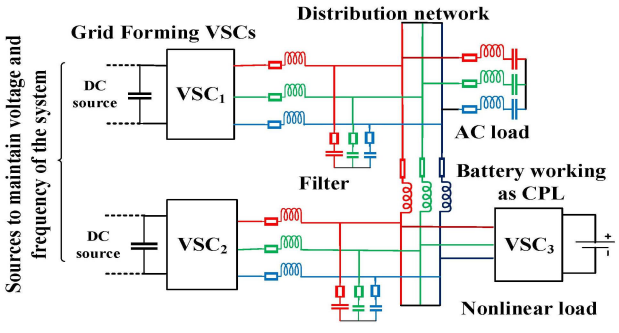


Fig. 1: Autonomous Microgrid setup

The system under consideration is an autonomous microgrid with its abstract presentation depicted in Fig. 1. Since the network is a distribution type in nature, the operating voltage of the system is 400V (phase to phase). The coupling between active and reactive power is not taken into consideration since the filter inductance value is considerable in such cases [22]. A Constant Power Load (CPL) has been taken into account due to its nature of negative incremental resistance which tends to destabilise the distribution network [23]. The different loads are distributed in nature, and hence, if the load current is a requirement for controlling the autonomous microgrid, a large number of current sensors are required. The proposed analysis does not require load current measurement, saving the cost of multiple load current sensors. The DG units and the loads are interfaced employing filter and distribution network, as shown in Fig. 1. The filter configuration employed is *LC* type where the resistances R_f (across L_f) is also included in the dynamics of the proposed analysis. The control architecture is built around the degree of freedom of the filter parameters (L_f, R_f and C_f). The power electronic interfaces, i.e., VSCs, are modulated using a control strategy obtained from Lyapunov based Adaptive technique which has been discussed in the following section.

III. Modelling of Parallel VSCs in Autonomous Microgrid

The control objective of the VSCs operating in parallel is to maintain the voltage and frequency of the autonomous microgrid, sharing the total load power in the microgrid. The instantaneous active power, p shared by each VSC is the microgrid can be calculated as

$$p = V_{od} I_{id} + V_{oq} I_{iq} \tag{1}$$

where the parameters employed can be seen in Fig. 3. The fundamental component of power, P is obtained using a low pass filter and is expressed as

$$P = \frac{\omega_c}{s + \omega_c} p \tag{2}$$

where ω_c is the cut-off frequency of the low pass filter.

The approach applied widely for such control objective is Proportional-Integral (PI) based static droop control [24]-[25]. The proposed adaptive method employs an adaptive droop control governed by,

$$\omega_0 = \omega^* - (m + \alpha) P \tag{3}$$

where ω_0 is the reference frequency of the autonomous microgrid, ω^* is the nominal frequency which is taken to be 314 *rad/s*, m is the static droop and α is the adaptive droop. Intuitively, the adaptive droop term is responsible for the improved transient response of the power-sharing in the autonomous setup. The idea behind introducing α is, increasing static droop, m improves the power-sharing; however, a low-frequency is introduced in the system if the value of static droop is very high [24]. Hence, a low value of m with adaptive α improves the power-sharing and helps to eliminate the problem of low-frequency oscillations in the system.

A. Control Design Using Lyapunov analysis based Adaptive Technique

The control design is derived using Lyapunov energy function of the autonomous system. The analysis employs adaptive technique to introduce parameter estimate ($\hat{\alpha}$) for adaptive term (α) and virtual control inputs (i^*_{1dq}, V^*_{odq}) in

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Exeter. Downloaded on May 05,2020 at 15:09:36 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

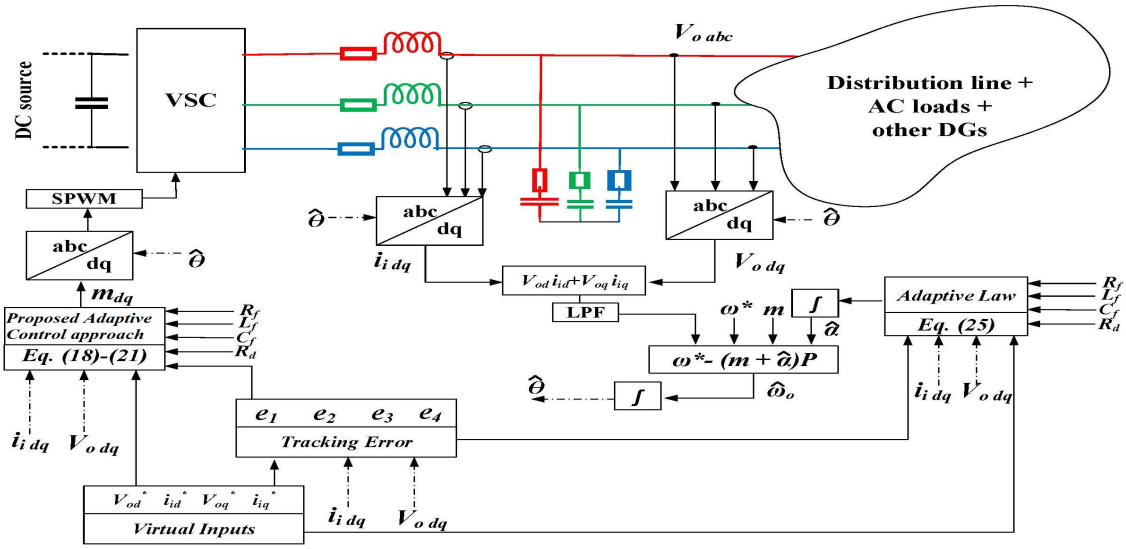


Fig. 2: Proposed Adaptive Control

the system dynamics [20]-[21], [26] (the terms are elaborated subsequently). The control technique takes load current (i^*_{odq}) as disturbance focusing on the notion of boundedness of the tracking error (e_1, e_2, e_3, e_4) dynamics. The results ensure exponential convergence of the tracking errors to a bound. Control gains introduced (γ, k_1, k_2, k_3, k_4) in the analysis are employed to vary the bound for the errors.

B. Mathematical analysis

The dynamics of the 3ϕ autonomous inverter setup is given using Kirchoff's law where the system is transformed into the direct-quadrature (*d-q*) form from the 3ϕ form. R_f are considered in the dynamics to present a more practical filter configuration. The system after the *d-q* transformation can be represented with the following differential equations where the notations are as indicated in Fig. 3.

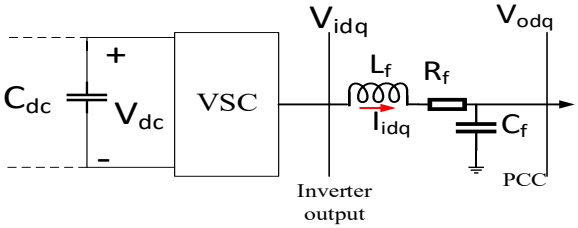


Fig. 3: The direct-quadrature equivalent of the 3ϕ (abc) single VSC configuration

1) System Dynamics:

$$\dot{i}_{1d} = -R_f i_{1d}/L_f - V_{od}/L_f + V_{id}/L_f + \omega_0 i_{1q} \quad (4)$$

$$\dot{i}_{1q} = -R_f i_{1q}/L_f - V_{oq}/L_f + V_{iq}/L_f - \omega_0 i_{1d} \quad (5)$$

$$\dot{V}_{od} = i_{1d}/C_f - i_{od}/C_f + \omega_0 V_{oq} \quad (6)$$

$$\dot{V}_{oq} = i_{1q}/C_f - i_{oq}/C_f - \omega_0 V_{od} \quad (7)$$

The discussed technique does not require load current measurement and hence i_{od} and i_{oq} are considered to be disturbances with an upper bound.

$$|i_{od}| < \bar{d}_1 \qquad |i_{oq}| < \bar{d}_2 \quad (8)$$

2) Control Problem Formulation: The closed loop system is considered for the control problem formulation. The tracking error dynamics for the system is defined as,

$$\begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} = \begin{bmatrix} i_{id} - i_{id}* \\ i_{iq} - i_{iq}* \\ V_{od} - V_{od}* \\ V_{oq} - V_{oq}* \end{bmatrix}$$

where $i_{1d}*$ and $i_{iq}*$ are the virtual input current references to the *d-axis* and *q-axis*; $V_{od}*$ and $V_{oq}*$ are the static virtual *d-axis* and *q-axis* voltage references. Inclusion of the tracking error dynamics modify eq. (4)-(7) to represent the closed loop error dynamics for the system (eq. (9)-(12)).

$$\dot{e}_1 = -R_f e_1/L_f - R_f i^*_{1d}/L_f - V_{od}/L_f + V_{id}/L_f + \omega_0 i_{1q} - \dot{i}^*_{1d} \quad (9)$$

$$\dot{e}_2 = -R_f e_2/L_f - R_f i^*_{1q}/L_f - V_{oq}/L_f + V_{iq}/L_f + \omega_0 i_{1d} - \dot{i}^*_{1q} \quad (10)$$

$$\dot{e}_3 = e_1/C_f + i^*_{1d}/C_f + \omega_0 e_4 + \omega_0 V^*_{oq} - i_{od}/C_f \quad (11)$$

$$\dot{e}_4 = e_2/C_f + i^*_{1q}/C_f - \omega_0 e_3 - \omega_0 V^*_{od} - i_{oq}/C_f \quad (12)$$

3) Adaptive Control Synthesis: The Lyapunov Energy Function for the state dynamics of the autonomous inverter is defined as;

$$V_1(e_1, e_2, e_3, e_4) = \frac{1}{2}L_f e_1^2 + \frac{1}{2}L_f e_2^2 + \frac{1}{2}C_f e_3^2 + \frac{1}{2}C_f e_4^2 \quad (13)$$

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Exeter. Downloaded on May 05,2020 at 15:09:36 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

where the energy function is positive definite and radially unbounded. The derivative of the energy function is further realized as;

$$\dot{V}_1(e_1,e_2,e_3,e_4) = L_f e_1\dot{e}_1 + L_f e_2\dot{e}_2 + C_f e_3\dot{e}_3 + C_f e_4\dot{e}_4 \quad (14)$$

Parameter estimation error for the adaptive term is defined as,

$$\tilde{\alpha} = \alpha - \hat{\alpha} \quad (15)$$

where $\hat{\alpha}$ is the known estimate of adaptive droop, α.

$$\eta = m + \hat{\alpha} \quad (16)$$

where η is a parameter used for simplification. The estimated reference system frequency is defined as,

$$\hat{\omega_o} = \omega^* - \eta P \quad (17)$$

The direct-quadrature derivative of virtual current inputs (i^*_{1d} and i^*_{1q}) and the direct-quadrature modulated inverter output voltages (V_{id} and V_{iq}) of the autonomous inverter are defined as,

$$V_{id} = R_f i^*_{1d} + V^*_{od} + L_f \dot{i}^*_{1d} - \hat{\omega_o} L_f i^*_{1q} - k_1 e_1 \quad (18)$$

$$V_{iq} = R_f i^*_{1q} + V^*_{oq} - L_f \dot{i}^*_{1q} - \hat{\omega_o} L_f i^*_{1d} - k_2 e_2 \quad (19)$$

$$i^*_{1d} = -C_f \hat{\omega_o} V^*_{oq} - k_3 e_3 \quad (20)$$

$$i^*_{1q} = -C_f \hat{\omega_o} V^*_{od} - k_4 e_4 \quad (21)$$

Hence, incorporating the derivative of virtual current inputs and the modulating inverter output voltages (eq. (18)-(21)), the derivative of the energy function is given as;

$$\begin{aligned}\dot{V}_1(e_1,e_2,e_3,e_4) = &-(R_f+k_1)e_1^2 - (R_f+k_2)e_2^2 - k_3 e_3^2 \\ &-k_4 e_4^2 - e_3 i_{od} - e_4 i_{oq} + L_f e_1 i^*_{1q}\tilde{\alpha}P - L_f e_2 i^*_{1d}\tilde{\alpha}P \\ &+C_f e_3 V^*_{oq}\tilde{\alpha}P - C_f e_4 V^*_{oq}\tilde{\alpha}P\end{aligned} \quad (22)$$

Now, the idea is to introduce another energy function to take the terms with $\tilde{\alpha}$ into consideration.

$$V_2(e_1,e_2,e_3,e_4,\tilde{\alpha}) = V_1 + 0.5\tilde{\alpha}^2 \quad (23)$$

Taking the derivative of new energy function,

$$\dot{V}_2(e_1,e_2,e_3,e_4,\tilde{\alpha}) = \dot{V}_1 - \tilde{\alpha}\dot{\hat{\alpha}} \quad (24)$$

The derivative of the estimate of adaptive term is defined as,

$$\dot{\hat{\alpha}} = -P(L_f e_1 i^*_{1q} + L_f e_2 i^*_{1d} - C_f e_3 V^*_{oq} + C_f e_4 V^*_{od}) \quad (25)$$

The idea of designing $\dot{\hat{\alpha}}$ is to shape the energy function, V_2 eliminating the terms which are impeding the decay of energy function derivative. Accounting eq. (25), the derivative of the energy function is modified to the form,

$$\dot{V}_2(e_1,e_2,e_3,e_4,\tilde{\alpha}) = -(R_f+k_1)e_1^2 - (R_f+k_2)e_2^2 - k_3 e_3^2 - k_4 e_4^2 - e_3 i_{od} - e_4 i_{oq} \quad (26)$$

The idea is to upper bound the derivative of the energy function considering the direct-quadrature profile of load current and their derivative as disturbances in the system.

$$\dot{V}_2(e_1,e_2,e_3,e_4,\tilde{\alpha}) \le -(R_f+k_1)e_1^2 - (R_f+k_2)e_2^2 - k_3 e_3^2 - k_4 e_4^2 + |e_3|\bar{d}_1 + |e_4|\bar{d}_2 \quad (27)$$

The evaluation of control gains (k_1–k_4) is based on the philosophy of exponential convergence of tracking errors (e_1–e_4) towards a bounded region around equilibrium point. The exponential convergence of tracking errors is guaranteed if the derivative of the energy function can be expressed in the form, $\dot{V}_2 \le -\gamma V_2 + c$ ensuring uniform ultimate boundedness of the energy function and the tracking errors.

$$\dot{V}_2 \le -(R_f+k_1)e_1^2 - (R_f+k_2)e_2^2 - \gamma C_f e_3^2 - \gamma C_f e_4^2 - |e_3|\left[(k_3-\gamma C_f)|e_3| - \bar{d}_1\right] - |e_4|\left[(k_4-\gamma C_f)|e_4| - \bar{d}_2\right] \quad (28)$$

The control gains (k_1–k_4) are defined as,

$$k_1 = k_2 = \gamma L_f - R_f \quad (29)$$

$$k_3 \ge \gamma C_f + \bar{d}_1/|e_3| \quad (30)$$

$$k_4 \ge \gamma C_f + \bar{d}_2/|e_4| \quad (31)$$

4) Generation of the Firing Signal for the inverter: The autonomous inverter is fired using the proposed control philosophy. The SPWM technique employed for firing includes generation of m_{dq}, inverse Park's Transformation (m_{dq} to m_{abc}), comparing with carrier sinusoid to generate the firing signal. The generation of m_{dq} is defined employing eq. (18)-(19) as,

$$\begin{bmatrix} m_d & m_q \end{bmatrix} = \frac{2}{V_{dc}} \begin{bmatrix} V_{id} & V_{iq} \end{bmatrix}$$

5) Eigen Value Synthesis: The control objective of exponential convergence of tracking error towards a bound requires tuned value of introduced control gains (k_1–k_4). Eq(29)-(31) give the range of control gains. The dynamics is then employed to do the eigenvalue synthesis studying variation of control gains on the system performance. The synthesis gives the idea of small-signal stability of the system with the variation of control gain parameters. The state-space representation is expressed as,

$$\dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{B}u(t) + \boldsymbol{D}d(t) \quad (33)$$

where $\boldsymbol{A}$ is the state matrix, $\boldsymbol{B}$ is the input matrix and $\boldsymbol{D}$ is the disturbance matrix [see eq. (32)]. $x = [e_1\ e_2\ e_3\ e_4]^T$ gives the state variables, $u = [V^*_{od}\ V^*_{oq}]^T$ gives the inputs, $d = [d_1\ d_2\ d_3\ d_4]^T$ gives the disturbances in VSC state-space representation. The eigenvalues are plotted for varying control gains. Fig. 4 shows the eigenvalue loci for varying k_1,k_2,k_3,k_4. Table I & II shows the eigenvalues for different k_1,k_2,k_3,k_4. It can be seen from table I that $k_{1min} > 0$. Similar studies are presented subsequently to verify the values of k_{2min}, k_{3min}, k_{4min} for system stability.

TABLE I: Eigen value loci with the variation of gain parameters, k_1 and k_2

λ	k_1=1000	k_1=500	k_1=0
λ_1	-471.23	-471.08	52.57
λ_2, λ_3	-436.61± 399.13i	-436.61±399.17i	-426.49±392.64i
λ_4	-1443.82	-848.15	-546.69
λ	k_2=1000	k_2=500	k_2=0
λ_1	-471.23	-471.08	52.57
λ_2, λ_3	-436.61± 399.13i	-436.61±399.17i	-426.49±392.64i
λ_4	-1443.82	-848.15	-546.69

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Exeter. Downloaded on May 05,2020 at 15:09:36 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

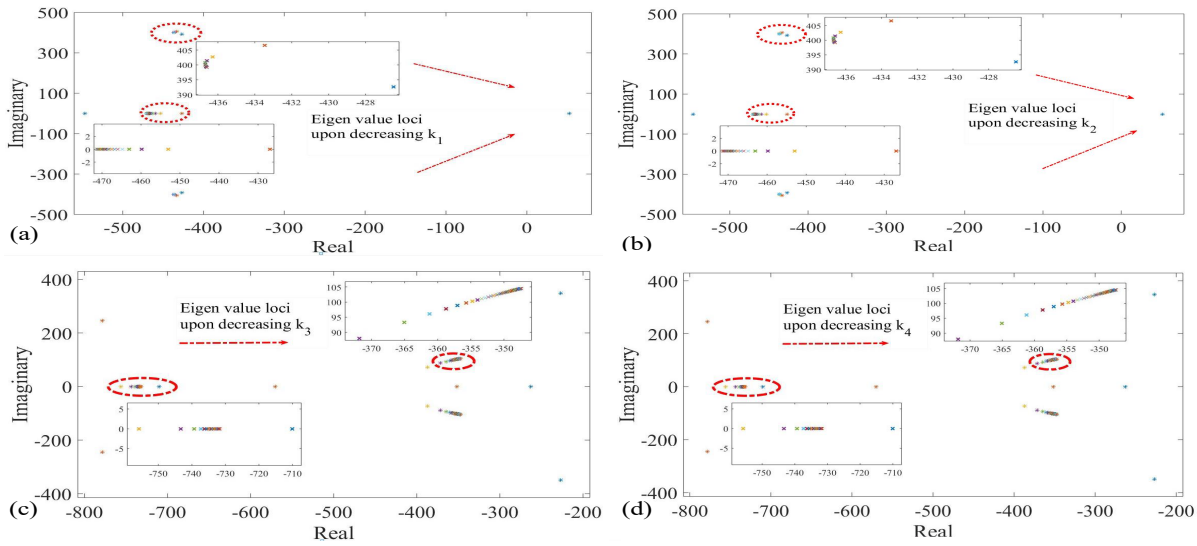


Fig. 4: Eigen Value loci for variation of (a) k_1, (b) k_2, (c) k_3, (d) k_4

$$\frac{d}{dt}\begin{bmatrix}\Delta e_1\\ \Delta e_2\\ \Delta e_3\\ \Delta e_4\end{bmatrix} = \begin{bmatrix} -\frac{R_f+k_1}{L_f} - \eta V_{od} e_2 & \hat{\omega_o} - \eta V_{oq} e_2 & -\frac{1}{L_f} - \eta i_{1d} e_2 & -\eta i_{1q} e_2 \\ \hat{\omega_o} - \eta V_{od} e_1 & -\frac{R_f+k_2}{L_f} - \eta V_{oq} e_1 & -\eta i_{1d} e_1 & -\frac{1}{L_f} - \eta i_{1q} e_1 \\ \frac{1}{C_f} - \eta V_{od} e_3 & -\eta V_{oq} e_4 & -\frac{k_3}{C_f} - \eta i_{1d} e_4 & \hat{\omega_o} - \eta i_{1q} e_4 \\ \eta i_{1d} e_3 & \frac{1}{C_f} + \eta V_{oq} e_3 & -\hat{\omega_o} + \eta i_{1d} e_3 & -\frac{k_4}{C_f} + \eta i_{1q} e_3 \end{bmatrix} \begin{bmatrix}\Delta e_1\\ \Delta e_2\\ \Delta e_3\\ \Delta e_4\end{bmatrix} + \eta \begin{bmatrix} -i_{1d} e_2 & -i_{1q} e_2 \\ -i_{1d} e_1 & -i_{1q} e_1 \\ -i_{1d} e_4 & -i_{1q} e_4 \\ i_{1d} e_3 & i_{1q} e_3 \end{bmatrix} \begin{bmatrix}\Delta V_{od}^* \\ \Delta V_{oq}^*\end{bmatrix} + \begin{bmatrix} 0 & 0 \\ -\frac{1}{C_f} & 0 \end{bmatrix} \begin{bmatrix}\Delta d_1 \\ \Delta d_2\end{bmatrix} \quad (32)$$

TABLE II: Eigen value loci with the variation of gain parameters, k_3 and k_4

λ	k_3=1000	k_3=500	k_3=0
λ_1	-731.71	-531.83	-262.70
λ_2, λ_3	-347.55± 390.81i	-347.87± 104.24i	-226.64± 349.02i
λ_4	-2105.25	-1052.61	-710.05
λ	k_4=1000	k_4=500	k_4=0
λ_1	-731.71	-531.83	-262.70
λ_2, λ_3	-347.55± 390.81i	-347.87± 104.24i	-226.64± 349.02i
λ_4	-2105.25	-1052.61	-710.05

IV. Simulation Results and Discussion

MATLAB/Simulink platform is employed for the simulation-based validation. The system considered is shown in Fig. 1. The simulation events considered for the comparison between the proposed control and the conventional PI control are elaborated in Table III. The events are kept similar to ensure proper comparison of control actions.

A. Simulation Events

1) AC load transients: Initially, both VSCs operating in parallel are contributing *320W* each (*m=22.86e-4*, α*=0*) when the AC load is consuming *640W* as can be seen in Fig. 6(a). A load transient of *640W* occurs at *t=1.3s* making the total load *1280W* as explained in Table III. The value of the adaptive droop increases to *10.27e-4* from *0* for both VSCs

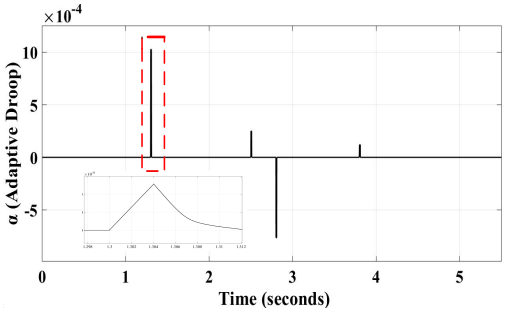


Fig. 5: Variation of adaptive droop, α with time with Event transients indicated in Table III

TABLE III: Event transients captured for comparison in Fig. 5

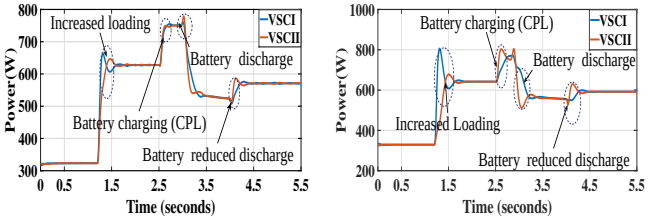
Parameter	Time of operation	Value
$P_1 + P_2$ (W)	0.0 s to 1.3 s	640 W (320 W each)
$P_1 + P_2$ (W)	1.3 s to 2.5 s	1280 W (640 W net change)
$P_1 + P_2$ (W)	2.5 s to 2.8 s	1480 W (200 W net change)
$P_1 + P_2$ (W)	2.8 s to 3.8 s	1020 W (460 W net change)
$P_1 + P_2$ (W)	3.8 s to 5.5 s	1120 W (100 W net change)

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Exeter. Downloaded on May 05,2020 at 15:09:36 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

and again reduces to *0*. The idea is to obtain better power sharing and also to eliminate the low-frequency oscillations with the adaptive term. The variation of α can be seen in Fig. 5. The proposed adaptive control also mitigates the transients better than PI control as can be seen, comparing Fig. 6(a) and Fig. 6(b). The peak overshoot for the proposed control is lesser than that for PI control. Fig. 6(c) shows the frequency profile for the proposed control and Fig. 6(d) for PI control. Similar to the power response, the proposed control mitigate the transients in frequency in a better manner.

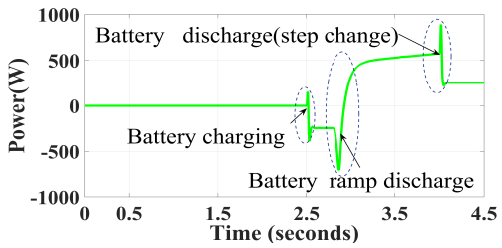
2) CPL transients: A *42V* lead-acid battery connected on the DC side of VSC_3 acts as CPL for the autonomous microgrid. At *t=2.5s*, CPL switching of *200W* (battery charging) occurs in the system. Another CPL switching of *460W* (battery discharging) occurs at *t=2.8s*. Fig. 7 validates the performance of proposed controller mitigating the CPL transients. The performance is validated to be better than the conventional PI control.



(a) Active power(proposed control) (b) Active power(PI)

(c) Frequency(proposed control) (d) Frequency(PI)

Fig. 6: MATLAB/simulink responses for transient events explained in Table III

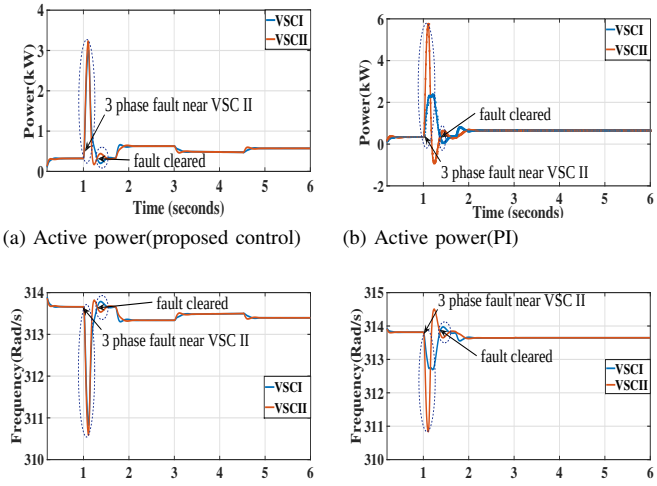


(a) CPL response

Fig. 7: MATLAB/Simulink responses of CPL operation explained in Table III

3) Fault transients: The case of transient mitigation during an AC faults contingency is discussed here. A 3ϕ fault occurs near VSC_2 for *100ms* at *t=1s* and is cleared at *t=1.1s* [see Fig. 8]. The contribution to the fault is dependent on the control employed. It can be seen that power-sharing has a better profile for the proposed control than for PI control. Fig. 8(a) shows both VSCs contributing *P=3.2 kW* whereas Fig. 8(b) shows VSC_2 contributing *P=5.8 kW* alone while VSC_1 contributing around *P=2 kW*. The frequency output of both VSCs are shown in Fig. 8(c) & (d). The proposed controller has better performance in terms of maintaining the same frequency to the distribution system. As it can be seen that during the transient, both VSC maintain the almost identical frequency when the proposed controller is applied. However, with the conventional PI controller VSC, I & VSC II have different frequency response during the transient as the drop in VSC II frequency output is higher than the VSC I frequency output. The settling time for the proposed control is also lesser than that for PI control. It means that for power restoration after a fault isolation, the proposed approach restores the system at *t=1.45s* whereas the conventional PI restores the system at *t=1.92s*.

However, the steady-state error is zero for the case of PI control, whereas the steady-state error is in a bound for the proposed approach. The proposed analysis gives uniform ultimate boundedness of the tracking errors. Hence, the error converges exponentially to a bound around the equilibrium point (zero).



(a) Active power(proposed control) (b) Active power(PI)

(c) Frequency(proposed control) (d) Frequency(PI)

Fig. 8: MATLAB responses for 3 phase ground fault.

V. Experimental Results

A. Experimental setup

The hardware validation of the proposed adaptive backstepping is carried out using an FPGA based NI- PXIe system. The scaled-down assembly for hardware validation is shown in Fig. 9 where the system shown in Fig. 1 has been implemented in a lab scale setup at *110V*. The two VSCs responsible for maintaining the voltage and frequency of the system and also for sharing the load proportionately are shown in the figure as VSC1 and VSC2. For constant power load (CPL), series combination of *12V* lead-acid batteries are connected on the DC side of a VSC through a bi-directional converter. The voltages and currents measurements are taken employing voltage sensors (LV25-P) and current sensors (LA55-P), shown in Fig. 9. Also, the RLC type loads connected to the system is shown as load banks in the figure. FPGA board (NI sbRIO 9606 and

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Exeter. Downloaded on May 05,2020 at 15:09:36 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

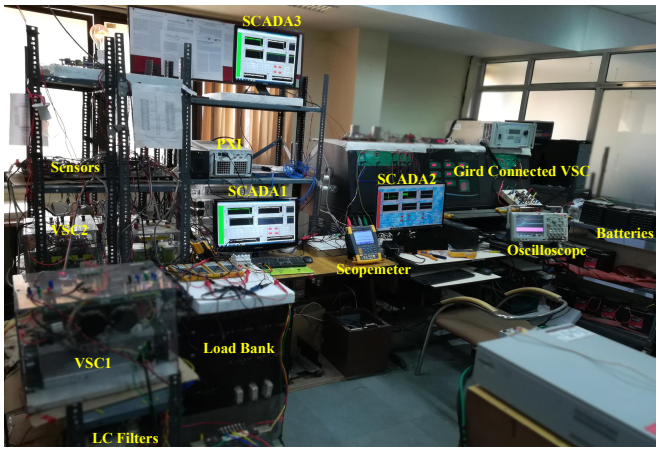
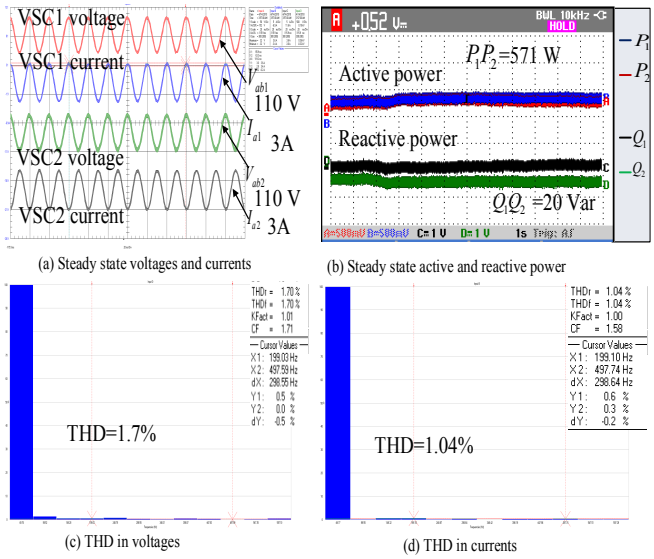


Fig. 9: Hardware setup for experimental validation

9683) is used for data acquisition and processing of signals. NI LabView is employed to design the control algorithms shown as SCADA1, SCADA2 and SCADA3 in Fig. 9.



(a) Steady state voltages and currents

(b) Steady state active and reactive power

(c) THD in voltages

(d) THD in currents

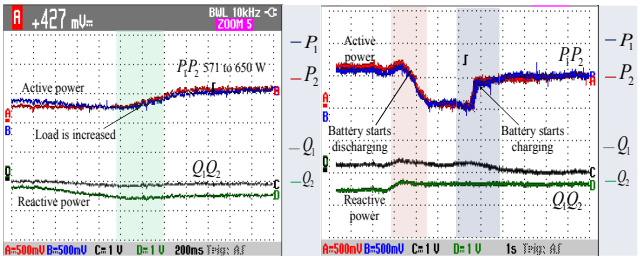
Fig. 10: Experimental Results (steady state)

B. Steady state response (VSC_1 & VSC_2)

Fig. 10(a) shows the voltage and current profile of VSCs running in parallel. Fig. 10(a) includes the VSC_1 and VSC_2 voltages and currents while the system is operating in steady-state. The output voltage of VSCs is at 110 V while both are supplying 3 A current to the loads connected to the system. The active and reactive powers of both VSC combined is shown in Fig. 10(b). The net active power transferred by VSC1 and VSC2 is of 571 W and reactive power transferred is of 20 Var. The VSCs share *571W* as seen in Fig. 10(b). Total Harmonic Distortion (THD) in AC phase voltage for the proposed backstepping control is 1.7% (see Fig. 10(c)) while THD in the current plot is 1.04%. (see Fig. 10(d)).

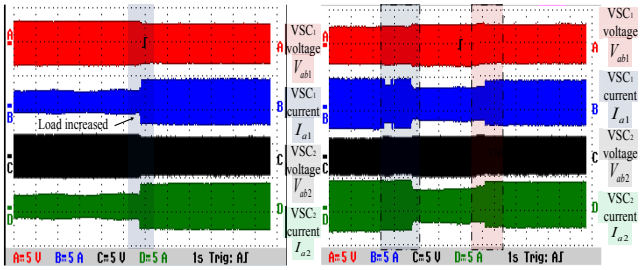
C. Dynamic state response (VSC_1 & VSC_2)

The performance of power-sharing during a load transient (*571W* to *650W*) is studied next. Fig. 11(a) shows the pro-



(a) Active and reactive power sharing in dynamic loading

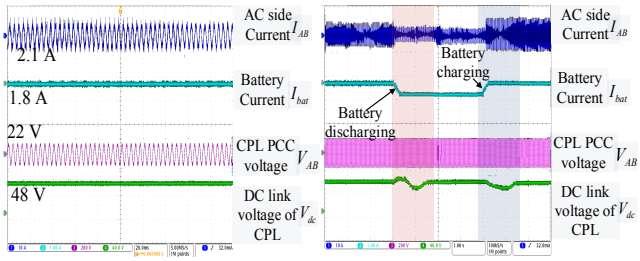
(b) Active and reactive power sharing in CPL loading



(c) Voltages and currents in increased loading

(d) Voltages and currents in CPL loading

Fig. 11: Experimental Results (load changes)



(a) Steady state operation of battery discharging

(b) Battery changed from charging to discharging

Fig. 12: Experimental Results (load changes)

portional power-sharing profile of both VSCs for the load transient. Fig. 11(c) shows the voltage and current profile of VSCs for such an operation. It can be seen that the current by each VSC increases from *3A* to *4.1A*. The loads are increased to make the loading dynamic, and it can be seen that the active power is increased from 571 W to 650 W and similarly the reactive power demand is also increased from 20 Var to 32 Var. However, this increased in active power demand is proportionally shared by both VSCs, which support the claim of the proposed control philosophy.

D. Dynamic state response with CPL (VSC_1, VSC_2 & VSC_3)

CPL operations in the form of battery charging and discharging are studied next. Three lead-acid batteries (*12V, 42Ah* each) in series are interfaced with a bi-directional converter operating in current control mode. The output of the bi-directional converter is connected to the DC link of VSC_3. The DC link of VSC is maintained at *48V*. Initially, battery discharges which reduce the power requirement (*571W* to *450W*) from the VSC set. The battery is made to charge next,

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Exeter. Downloaded on May 05,2020 at 15:09:36 UTC from IEEE Xplore. Restrictions apply.

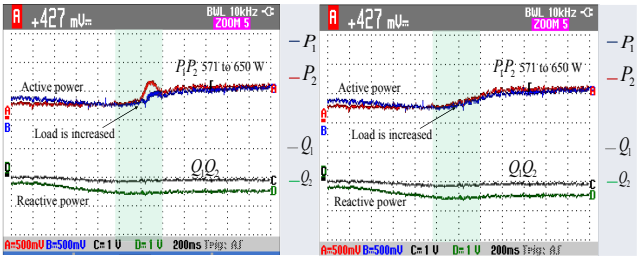
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

increasing power requirement from *450W* to *560W*. Fig. 11(b) shows the contribution of each VSC in the CPL operation. The current contribution drops from *3A* to *1.9A* when the battery starts to discharge. The current contribution starts to increase again from *1.9A* to *2.9A* when the battery starts charging. The sequence of events and the following current contributions are elaborated in Table IV. CPL dynamics has been presented in Fig. 12(a), 12(b). In steady-state, the battery is discharging at *1.8A*. A step-down transformer (*5:1*) is connected between distribution lines and VSC_3. The AC side voltage of VSC_3 is at *22V*.

TABLE IV: Events during real time CPL operation

Parameter	Time of operation	Value
$P_1 + P_2$ (W)	0.0 s to 1.5 s	571 W (285.5W each)(Fig. 9(b))
$P_1 + P_2$ (W)	1.5 s to 4.5 s	450 W (79W net change)(Fig. 9(b))
$P_1 + P_2$ (W)	4.5 s to 6 s	560 W (110W net change)(Fig. 9(b))
$i_{a1}\&i_{a2}$ (W)	0 s to 2.5 s	3A each (Fig. 9(b))
$i_{a1}\&i_{a2}$ (W)	2.5 s to 4.5 s	1.9A each (Fig. 9(b))
$i_{a1}\&i_{a2}$ (W)	4.5 s to 6 s	2.9A each (Fig. 9(b))

E. Comparision between PI based control and proposed control reponse (VSC_1 & VSC_2)



(a) Active power sharing (PI control) (b) Active power sharing (proposed control)

Fig. 13: Experimental Results (comparision between PI and proposed control)

Similar to the simulation results, a comparison between the PI-based control and the proposed control is carried out in real-time. The performance of power-sharing during a load transient (*571W* to *650W*) is studied next. Fig. 13(a) shows the proportional power-sharing profile of both VSCs for the load transient when PI control is applied for both VSCs. The loads are increased to make the loading dynamic, and it can be seen that the active power is increased from 571 W to 650 W. The same steps are carried out for the proposed adaptive droop based control to maintain the uniformity. The active power-sharing dynamics for the proposed control is shown in Fig. 13(b) The proposed adaptive backstepping control mitigates the transients better than PI control as can be seen in Fig. 13, as the peak overshoot for the proposed control, in case of VSC_2 is lesser than that for PI control.

VI. Conclusion

The work proposes an adaptive droop based control for active power-sharing between VSCs in an autonomous microgrid. The adaptive term introduced in the droop control acts as a feed-forward term and is responsible for the transient performance of the system. A Lyapunov based technique has been proposed, which ensures uniform ultimate boundedness (UUB) of the tracking errors. The focus of the work is on transient mitigation for different transients in the system. A thorough comparison of the proposed control with the conventional PI control is presented employing MATLAB/Simulink as the platform. AC load variations, Constant Power Load (CPL) switching, AC faults restoration transients are considered for the comparison. FPGA based hardware setup is employed to study the real-time performance of the proposed adaptive control.

References

[1] IEEE Guide for Design, Operation, and Integration of Distributed Resource Island Systems with Electric Power Systems," in *IEEE Std 1547.4-2011* , vol., no., pp.1-54, 20 July 2011.

[2] Q. Zhong, "Robust Droop Controller for Accurate Proportional Load Sharing Among Inverters Operated in Parallel," in *IEEE Transactions on Industrial Electronics,* vol. 60, no. 4, pp. 1281-1290, April 2013.

[3] J. He and Y. W. Li, "An Enhanced Microgrid Load Demand Sharing Strategy," in *IEEE Transactions on Power Electronics,* vol. 27, no. 9, pp. 3984-3995, Sept. 2012.

[4] R. Han, L. Meng, G. Ferrari-Trecate, E. A. A. Coelho, J. C. Vasquez and J. M. Guerrero, "Containment and Consensus-Based Distributed Coordination Control to Achieve Bounded Voltage and Precise Reactive Power Sharing in Islanded AC Microgrids," in *IEEE Transactions on Industry Applications,* vol. 53, no. 6, pp. 5187-5199, Nov.-Dec. 2017.

[5] T. Wu, Z. Liu, J. Liu, S. Wang and Z. You, "A Unified Virtual Power Decoupling Method for Droop-Controlled Parallel Inverters in Microgrids," in *IEEE Transactions on Power Electronics,* vol. 31, no. 8, pp. 5587-5603, Aug. 2016.

[6] T. Nguyen, H. Yoo and H. Kim, "A Droop Frequency Control for Maintaining Different Frequency Qualities in a Stand-Alone Multimicrogrid System," in *IEEE Transactions on Sustainable Energy,* vol. 9, no. 2, pp. 599-609, April 2018.

[7] G. Lou, W. Gu, Y. Xu, M. Cheng and W. Liu, "Distributed MPC-Based Secondary Voltage Control Scheme for Autonomous Droop-Controlled Microgrids," in *IEEE Transactions on Sustainable Energy,* vol. 8, no. 2, pp. 792-804, April 2017.

[8] X. Huang, K. Wang, J. Qiu, L. Hang, G. Li and X. Wang, "Decentralized Control of Multi-Parallel Grid-Forming DGs in Islanded Microgrids for Enhanced Transient Performance," in *IEEE Access,* vol. 7, pp. 17958-17968, 2019.

[9] D. E. Olivares et al., "Trends in Microgrid Control," in *IEEE Transactions on Smart Grid,* vol. 5, no. 4, pp. 1905-1919, July 2014.

[10] H. R. Baghaee, M. Mirsalim, and G. B. Gharehpetian, Real-time verification of new controller to improve small/large-signal stability and fault ridethrough capability of multi-DER microgrids, *IET Gener. Transmiss. Distrib.,* vol. 10, no. 12, pp. 30683084, Sep. 2016.

[11] H. R. Baghaee, M. Mirsalim, G. B. Gharehpetian, and H. A. Talebi, A generalized descriptor-system robust H control of autonomous microgrids to improve small and large signal stability considering communication delays and load nonlinearities, *Int. J. Elect. Power Energy Syst.,* vol. 92, no. 1, pp. 6382, Nov. 2017.

[12] H. R. Baghaee, M. Mirsalim, G. B. Gharehpetan and H. A. Talebi, "Nonlinear Load Sharing and Voltage Compensation of Microgrids Based on Harmonic Power-Flow Calculations Using Radial Basis Function Neural Networks," in *IEEE Systems Journal,* vol. 12, no. 3, pp. 2749-2759, Sept. 2018.

[13] H. R. Baghaee, M. Mirsalim and G. B. Gharehpetian, "Power Calculation Using RBF Neural Networks to Improve Power Sharing of Hierarchical Control Scheme in Multi-DER Microgrids," in *IEEE Journal of Emerging and Selected Topics in Power Electronics,* vol. 4, no. 4, pp. 1217-1225, Dec. 2016.

[14] H. R. Baghaee, M. Mirsalim, G. B. Gharehpetian and H. A. Talebi, "Unbalanced harmonic power sharing and voltage compensation of microgrids using radial basis function neural network-based harmonic power-flow calculations for distributed and decentralised control structures," in *IET Generation, Transmission & Distribution,* vol. 12, no. 7, pp. 1518-1530, 10 4 2018.

[15] M. B. Delghavi and A. Yazdani, "Sliding-Mode Control of AC Voltages and Currents of Dispatchable Distributed Energy Resources in Master-Slave-Organized Inverter-Based Microgrids," in *IEEE Transactions on Smart Grid,* vol. 10, no. 1, pp. 980-991, Jan. 2019.

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Exeter. Downloaded on May 05,2020 at 15:09:36 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2988712, IEEE Journal of Emerging and Selected Topics in Power Electronics

[16] M. Cucuzzella, G. P. Incremona and A. Ferrara, "Decentralized Sliding Mode Control of Islanded AC Microgrids With Arbitrary Topology," in *IEEE Transactions on Industrial Electronics,* vol. 64, no. 8, pp. 6706-6713, Aug. 2017.

[17] D. O. Amoateng, M. Al Hosani, M. S. Elmoursi, K. Turitsyn and J. L. Kirtley, "Adaptive Voltage and Frequency Control of Islanded Multi-Microgrids," in *IEEE Transactions on Power Systems,* vol. 33, no. 4, pp. 4454-4465, July 2018.

[18] T. Hosseinalizadeh, H. Kebriaei and F. R. Salmasi, "Decentralised robust T-S fuzzy controller for a parallel islanded AC microgrid," in *IET Generation, Transmission & Distribution,* vol. 13, no. 9, pp. 1589-1598, 7 5 2019.

[19] Graeme Heald "Issues with Reliability of Fuzzy Logic" Published in *International Journal of Trend in Scientific Research and Development (ijtsrd)*, ISSN: 2456-6470, Volume-2 — Issue-6, October 2018, pp.829-834,

[20] H. K. Khalil, Nonlinear Systems, 2nd ed., *Englewood Cliffs,* NJ, USA: Prentice-Hall, 1996.

[21] Yaolong Tan, Jie Chang, Hualin Tan and Jun Hu, "Integral backstepping control and experimental implementation for motion system," in Proceedings of the 2000 *IEEE International Conference on Control Applications,* Anchorage, AK, USA, 2000 pp. 367,368,369,370,371,372.

[22] D. Rajicic and A. Bose, "A modification to the fast decoupled power flow for networks with high R/X ratios," in *IEEE Transactions on Power Systems,* vol. 3, no. 2, pp. 743-746, May 1988.

[23] A. Emadi, A. Khaligh, C. H. Rivetta and G. A. Williamson, "Constant power loads and negative impedance instability in automotive systems: definition, modeling, stability, and control of power electronic converters and motor drives," in *IEEE Transactions on Vehicular Technology,* vol. 55, no. 4, pp. 1112-1125, July 2006.

[24] A. Firdaus and S. Mishra, "Mitigation of Power and Frequency Instability to Improve Load Sharing Among Distributed Inverters in Microgrid Systems," in *IEEE Systems Journal,* vol. 14, no. 1, pp. 1024-1033, March 2020.

[25] M. Yazdanian and A. Mehrizi-Sani, "Internal Model-Based Current Control of the RL Filter-Based Voltage-Sourced Converter," in *IEEE Transactions on Energy Conversion,* vol. 29, no. 4, pp. 873-881, Dec. 2014.

[26] P. D. Achlerkar, V. Nougain and B. K. Panigrahi, "Backstepping Technique based Customer-end Voltage Control Strategy of DC Distribution Network," in *IEEE Journal of Emerging and Selected Topics in Power Electronics.*



Surya Prakash received the B.Tech. degree in Electrical Engineering from School of Engineering, CUSAT, Kochi, India in 2009 and the M.Tech. degree in Power System from the NIT Trichy, Tamilnadu, India, in 2015. He has worked as Design Engineer at Hydro Machinery Development Station, BHEL, Bhopal India for 3 years. He is currently working toward the Ph.D. degree at Electrical Engineering Department, Indian Institute of Technology, Delhi, India.

His current research interests include control of grid connected inverter system under varying grid conditions.



Vaibhav Nougain Vaibhav Nougain received B.Tech degree in electrical engineering in 2017 from Delhi Technological University (Formerly Delhi College of Engineering), New Delhi, India. He is currently working toward the doctoral degree at the Department of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi, India. He was a Visiting Student with The Center for Electric Power and Energy (CEE), Technical University of Denmark in 2019. His research interests include protection and control of DC and AC microgrids, islanding detection, and protection of HVDC systems.



Sukumar Mishra (M'97-SM'04) received the B.Sc (Engg) degree in electrical engineering from University College of Engineering Burla, Odisha, India, in 1990, and the M.Sc (Engg) and Ph.D. degrees in electrical engineering from National Institute of Technology Rourkela, Odisha, India, in 1992 and 2000, respectively.

He a Professor with the Indian Institute of Technology Delhi, New Delhi, India, and has been a part of the institute for the past 15 years. He has authored or coauthored over 200 research articles (including papers in international journals, conferences, and book chapters). His research interests include power systems, power quality studies, renewable energy, and smart grid.

Prof. Mishra is a recipient of many accolades throughout his academic tenure of over 25 years. He has been a recipient of Indian National Science Academy Medal for Young Scientist (2002), Indian National Academy of Engineering (INAE) Young Engineer Award (2009), INAE Silver Jubilee Young Engineer Award (2012), and has recently won the Samanta Chandra Shekhar Award (2016). He has been granted fellowship from many prestigious technical societies like IET (U.K.), NASI (India), INAE (India), IETE (India), and IE (India) and is also recognized as the INAE Industry Academic Distinguish Professor. He is currently serving as an Editor for the IEEE TRANSACTIONS ON SUSTAINABLE ENERGY, IEEE TRANSACTIONS ON SMART GRID, and an Associate Editor for the IET Generation, Transmission, & Distribution journal. Currently, he is the Vice Chair of Intelligent System Subcommittee of Power and Energy society (PES) of IEEE. Apart from all research and academic collaborations, he is very actively involved in industrial collaborations. He is currently acting as INAE Chair Professor and was previously delegated as the NTPC and Power Grid Chair Professor. He is also serving as an Independent Director of Cross Border Power Transmission Company Ltd., Gurugram, India, and River Engineering Pvt. Ltd., Greater Noida, India.

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Exeter. Downloaded on May 05,2020 at 15:09:36 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

An Embedding-based Deep Learning Approach for Movie Recommendation

Ram Murti Rawat
Assistant Professor, Department of COE, Delhi Technological University, Delhi, India.

Vikrant Tomar, Vinay Kumar
Undergraduate Student, Department of COE, Delhi Technological University, Delhi, India.

Abstract - **Recommender systems are employed either as tools or algorithms, whose key task is to efficiently predict the ratings for items and recommend items, using the data generated by users. It assists users in finding items that they will like. Hence, recommendation systems are becoming an essential part of some applications, e-commerce websites, and online streaming services, etc. This paper emphasizes on recommendation system for movies whose main objective is to propose a movie recommendation system through a deep learning technique. The size and complication of websites have increased due to the rapid growth of the internet. On these websites, it has become time-consuming and extremely difficult for the users to find the information that they are searching for. Therefore, a collaborative filtering-based movie recommendation system using deep learning and embedding is proposed. The proposed system is evaluated by calculating the RMSE and MAE values. The proposed method is compared with other machine learning algorithms and some state-of-the-art methods. Our model gave an MAE score of 0.7305 and an RMSE score of 0.9311. It performs better than some previous approaches to collaborative-filtering based movie recommendation.**

Keywords – Recommendation system, artificial neural network, collaborative filtering, embedding, machine learning

I. INTRODUCTION

A recommendation system is now a central part of all e-commerce applications and online entertainment services. The recommendation system gains data about the user's preferences for a particular item (e.g. songs, movies, games, music, etc.) in two different ways. An implicit way of obtaining user data involves observing the user's past behavior such as order history, watched movies, and search queries. Another way of obtaining data is done by gathering the user's previous ratings and likes. The size and complication of websites have increased due to the exponential growth of the internet. On these websites, it has become time-consuming and extremely difficult for the users to find the information that they are searching for. Recommendation systems assist users in finding the information that is within their interests by some interaction with the user. A Recommendation system predicts the rating for items that the user has not rated yet based on his previous ratings and ratings by similar users. Recommendation systems have been a field of research for decades now, but the curiosity remains high because of the plenty of practical applications and the problem abundant domain. Recommender systems deliver tailored information by the familiarity of the user's interests by interaction or experience with that user. As online streaming services are becoming a part of our lives, the recommendation systems have gained importance. With this in mind, our objective is to work on a collaborative filtering-based deep learning strategy for the movie recommendation system. Collaborative filtering is an approach for recommendation systems that depends on the ratings given by a particular user along with the ratings for items given by other similar users. New items of interest can be recommended to the users if the ratings are precisely predicted for an item that the user has not yet rated. In this way, movies are recommended that are different than the user's previously watched movies. Whereas, a Content-Based Recommender only recommends movies based on the similarity between items. The basic idea behind content-based filtering is that if someone likes an item, then he or they will also like a "similar" item. It usually works well when it's easy to determine the properties of each item. Collaborative filtering helps users find diverse items if there are users with similar preferences. Different ways of performing collaborative filtering are present and have been implemented. One way is to create a user profile based on user's ratings and watch history and then find similar users using the k-nearest neighbor algorithm. Similarity can be defined in various ways between the users. The user profile can be defined as a vector of continuous values and then use some comparison measures like cosine similarity. Matrix factorization is one other known method for performing collaborative filtering. In this algorithm, a user-item matrix is composed, then it is factorized into two different matrices such that the inner dimension of the matrices denotes some latent factors. The representation of these factors in the above factorization helps recommend new unique items to the users. Now, deep learning is proving to be quite accurate in recommendation tasks. The popularity increase of deep learning in recommendation systems is due to its state-of-the-art performance and highly accurate predictions. Deep learning better represents the user's interest, features of the item, and past interactions between them as compared to other previous approaches to recommendation systems. A movie recommendation system using the deep learning technique and embedding layers are used.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:56:24 UTC from IEEE Xplore. Restrictions apply.

II. Related Work

Recommender systems are classified into three different techniques that are content-based [5], [6], collaborative filtering [7], [9]-[10], [20], hybrid [13], [14]. Also, clustering algorithms are used for recommendation purposes as it helps cluster similar users into groups. In a movie recommendation model, the system learns the previous user ratings then these ratings help the system to recommend news movies to users using different approaches. Collaborative filtering [7], [9]-[10], [20] is becoming extremely important as this technique is used by many of the recommender systems. Hybrid approaches use the combination of both the content and collaborative-based filtering techniques. Memory-based and model-based are the two main types of collaborative filtering techniques. Memory-based collaborative filtering [4], [16] discoveries similar users in the user space corresponding to a specific user and then recommend movies to that user dynamically. There are two problems related to this approach that are high computational complexity and sparsity of data. An item-based method for collaborative filtering was proposed by different authors to resolve these issues. They calculated the similarity amongst the nearby region items about a fixed entity. This usually uses the k-nearest neighbor algorithm to find groups of similar items. This technique reduced the time complexity to some extent and made accurate predictions. Model-based collaborative filtering [3], [8] approach prepares a model in advance based on the previous item ratings by the user and then the model delivers item recommendation. The model is prepared using different machine learning algorithms. This approach is scalable and solves data sparsity, but time complexity is higher than other techniques. The most general method of achieving collaborative filtering is to use a k-nearest-neighbor approach between users [7]. Clustering also helps in finding similar groups in userspace. K-Means clustering is the most common clustering algorithm. An effective collaborative movie recommendation system was proposed which uses a hybrid K-Means clustering with a Cuckoo search algorithm for optimization [1]. There is another popularly known technique matrix factorization for performing collaborative filtering [19]. Deep learning has reformed various fields of computer engineering, and now it is being extensively used in recommendation systems [2], [15], [20]. Wang et al. [20] built a collaborative model for recommendation using deep learning which works both on content data and rating matrix. A recommendation system is proposed by Elkahky et al. [15] using deep learning which uses search queries and web browsing history provided by users. But there is a limitation of this method, as it needs history and search queries which are not accessible all the time. Wei et al. [21] proposed a movie recommendation model using deep neural networks for prediction of movie ratings by mining the characteristics of items and is particularly used to cold start items. A deep learning technique for collaborative filtering based on autoencoders precisely predicts the movie ratings [2]. There is another extraordinary approach to applying collaborative filtering known as matrix completion (MC). In this technique, the rating matrix is approximated to a low-rank rating matrix. This is a mathematically complex approach. In 2014, Kalofolias et al [23] introduced a regularization of the MC model known as the geometric matrix completion (GMC) model by adding additional information in the form of user and item graphs. In [11], a more efficient graph-regularized method for matrix completion was proposed using alternating least squares optimization method (GRALS). More understandable and simpler approach with Embedding layers and deep neural networks is proposed.

III. Proposed Work

A. Dataset Description

MovieLens 100K data set is used in the proposed model. The MovieLens data sets were collected by the GroupLens Research Project at the University of Minnesota [22]. This data set contains 100,000 ratings in the range of 1 to 5 for 1682 movies from a user base of 943. The data set has been preprocessed and includes only those users who have rated more than 20 different movies. It also contains simple demographic information of users (age, gender, occupation, zip) and information of movie genres. 80% of the dataset will be used for training and rest 20% will be used for testing. The data of users who did not have complete information were removed. **u.data** file is the main file used in building the model. It contains the rating data from a user for a particular movie. It has 4 columns UserID, MovieID, Rating, Timestamp.

The distribution of ratings in range 1 to 5 is shown in Fig 1.

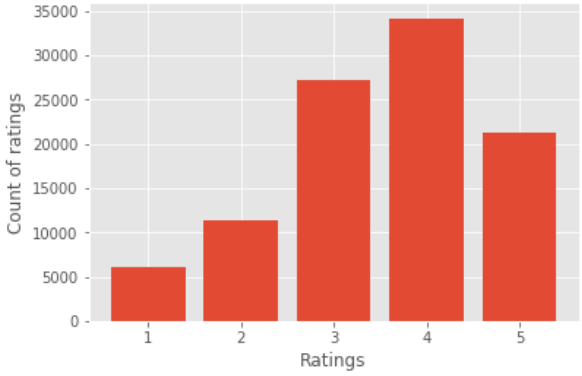


Fig. 1. Distribution of ratings

B. The architecture of the Model

After observing and pre-processing the dataset, 943 unique users and 1,682 unique movies is detected in the MovieLens dataset. However, the users and movie IDs are passed directly into our deep learning network because the model will try to find the relation between IDs that are not present. Our algorithm uses the embeddings to convert these IDs to a continuous vector of fixed size as machine learning algorithms

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:56:24 UTC from IEEE Xplore. Restrictions apply.

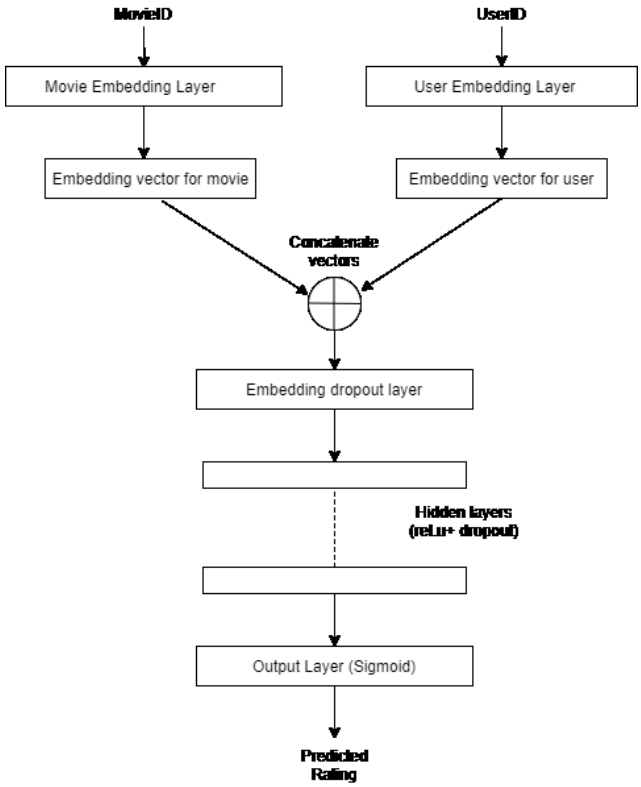


Fig. 2. Structure of network

expects numerical values as input. An embedding vector will be produced for a userID and one embedding vector for a movieID from the embedding space of 943 and 1682 respectively. Then, the vectors are concatenated into one and then passed to the dropout layer. The dropout layer is used to avoid the over fitting of a model. In each training epoch, dropout randomly sets outgoing edges of the layer to 0 with some predefined probability. Next, several fully-connected hidden layers are added with a dropout probability and reLU activation function. Each hidden layer can have a different value of output features and dropout probability. The input features are decided based on the output features of the previous layer. Finally, the predicted rating should be sent as an output for the input user and model IDs. For this, an output layer with a sigmoid activation function is employed and later rescale in the range of 1 to 5. The weights of the neurons in each layer are updated in each training epoch with a back propagation algorithm. The structure of our proposed network used as our model for predicting ratings is as shown in Fig.2.

C. Multi-Layer Perceptron (MLP)

Multi-Layer Perceptron (MLP) is a category of artificial neural networks (ANN). ANN a technique encouraged by the biological neuron network that makes up the human brain. This network learns to accomplish tasks by examining examples, without being explicitly programmed. The MLP is made up of several layers (an input layer, an output layer, and one or more hidden layers) that contains nodes with different activation functions. MLPs are fully connected, which means each node in a layer is connected to every node of the immediate next layer. The input layer receives the input data. No information processing is done on this layer. The received input is conveyed to the following layer, the hidden layer. The data directed from the input layer is processed in the hidden layers and passed to the next layer i.e. output layer. In MLP, many artificial neurons and hidden layers can be obtained. MLP can solve more complex tasks by adding a large number of neurons and layers for processing. However, it will increase the computation time. The output layer is the final layer, every node of the last hidden layer is connected to the final output layer, and hence, the result of the processed data from the hidden layer is produced at this layer. Fig. 3 shows a simple multi-layer perceptron. The advantages of MLP are:

1) It can solve highly complex and complicated problems.
2) It can handle missing data.
3) It learns the weights on its own using the backpropagation algorithm. Thus, it solves a wider range of problems than the algorithms.

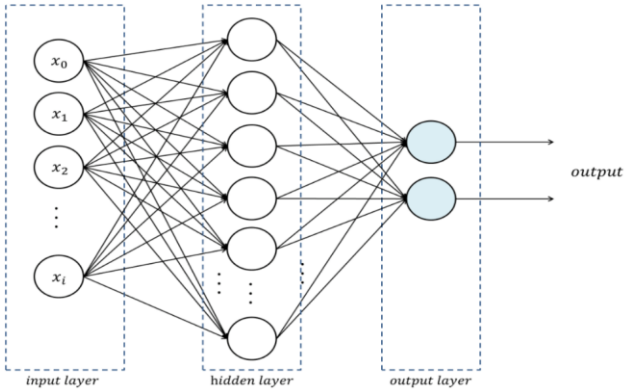


Fig. 3. Simple Multi-Layer Perceptron Model

D. Embeddings

An embedding is a conversion of discrete or categorical variables into a vector of continuous numerical values. In the view of deep learning, embeddings are multi-dimensional vector with learned continuous values that represent discrete or categorical variables. The embeddings are used in reducing the dimension of categorical variables and conveniently represent categories in a transformed space that can be used as input to neural networks and other machine learning algorithms. Embeddings have 3 basic purposes:

1. Finding nearest neighbors in the embedding space. It can be used in making recommendations based on item properties and user interests.
2. It can be used as input to various machine learning algorithms for a supervised learning problem.
3. Used in visualizing of relations between categorical variables and their characteristics.

Neural network embeddings overcome the problems of another simple method for transforming categorical variables, that is, one-hot encoding.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:56:24 UTC from IEEE Xplore. Restrictions apply.

E. Root Mena Square Error

Root Mean Square Error (RMSE) is a common evaluation metric used in calculating the accuracy of a system. RMSE is calculated by taking the square root of the mean squared errors. The mean value is calculated by dividing the sum of the squared difference between actual and predicted output by the number of samples. Equation (1) represents the RMSE value.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=0}^{N}(predicted_i - actual_i)^2}{N}} \quad (1)$$

N is the total number of instances. The smaller RMSE value means better performance for the system. Thus, a system which produces smaller RMSE value will provide better recommendations.

F. MAE

Mean Absolute error (MAE) is a similar evaluation metrics as RMSE. It is calculated by the mean of absolute error values. Absolute error is the positive value of the difference between the predicted output and the actual output. The equation for MAE is mentioned in (2).

$$\text{MAE} = \frac{\sum_{i=0}^{N}|predicted_i - actual_i|}{N} \quad (2)$$

IV. Experimental Result & Analysis

First, some pre-processing of the data is done on the dataset for the desired input for machine learning algorithms. The dataset is divided into two parts in which 80% is for training data and the rest 20% is test data for verification of the result. Our network has 4 basic parameters. To check the performance of our recommender system model, the RMSE and MAE is calculated corresponding to various values of these parameters.

Table 1. RMSE and MAE value for different parameter values.

k	hidden	emb_dropout	Dropouts	MAE	RMSE
100	[300, 200, 100]	0.02	[0.25, 0.25, 0.25]	0.7375	0.9371
100	[200, 100]	0.01	[0.15, 0.15]	0.7324	0.9321
75	[100, 200]	0.02	[0.25, 0.25]	0.7349	0.9353
50	[200, 300, 400]	0.02	[0.20, 0.20, 0.20]	0.7374	0.9355
50	[400, 300]	0.02	[0,25, 0.25, 0.25]	0.7319	0.9315
125	[100, 200]	0.02	[0.20, 0.25]	0.7337	0.9340
125	**[500, 400, 300]**	**0.01**	**[0.15, 0.15, 0.15]**	**0.7305**	**0.9311**
200	[300, 200, 300]	0.02	[0.25, 0.25, 0.20]	0.7353	0.9356
150	[200, 300, 200]	0.02	[0.25, 0.25, 0.25]	0.7402	0.9345

The size of embedding vectors for each unique user and movie is denoted by k. The hidden parameter is a list of values each corresponds to the number of output features of the hidden layers. The emp_dropout represents the dropout probability of the embedding dropout layer. The dropout parameter is a list of dropout probabilities for each hidden layer. Table 1 shows the RMSE and MAE for different parameter values.

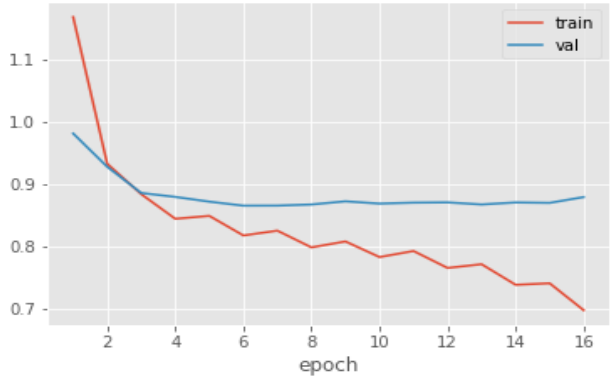


Fig 4. Training and Validation Error vs the number of Epoch

Fig. 4 shows how the training and validation errors are changing with increasing training epochs of the model. The training error is decreasing continuously but the validation error first decreases till epoch 6 but almost stays the same after that. Table 2 shows the corresponding RMSE and MAE values for a single parameter configuration but only changing k value, i.e. the size of output embedding from the embedding layer. The parameter configuration for the comparison is two hidden layers have [400, 300] output features, emp_dropout value is 0.02, dropout values are [0.25, 0.25].

Table 2. Evaluation metric values for different k.

k	*MAE*	*RMSE*
50	0.7359	0.9315
75	0.7330	0.9313
100	0.7370	0.9339
125	0.7315	0.9308
150	0.7338	0.9313
175	0.7328	0.9324
200	0.7331	0.9306

The proposed algorithm is compared with different algorithms for collaborative filtering. KNN is the K-nearest neighbor algorithm, which is a supervised machine learning algorithm for classification. PCA (Principal Component Analysis) is a technique for dimensionality reduction. A genetic algorithm (GA) is a nature-inspired algorithm. SOM refers to Self-organizing Map which maps a higher dimension input into a lower dimension. Clustering is an unsupervised learning algorithm that groups together similar items. GMC (geometric matrix completion) and GRALS (Graph regularized matrix completion method) are both state-of-the-art methods used for

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:56:24 UTC from IEEE Xplore. Restrictions apply.

collaborative filtering. KMEANS CUCKOO is a combination of K-Means clustering and Cuckoo search optimization technique. EMBEDDING+ANN is our proposed approach which uses Artificial Neural networks and Embedding. Table 3 shows the MAE and RMSE values for different algorithms. Our algorithm performs better than all of the algorithms in terms of RMSE value and has a larger value of MAE than just the KMEANS CUCKOO algorithm. Fig. 5 and 6 show the plot for different algorithms vs RMSE and MAE respectively.

Table 3. RMSE and MAE values for different algorithms

ALGORITHM	MAE	RMSE
KNN	-	1.124
PCA-GAKM	0.94	-
PCA-SOM	0.98	-
SOM-CLUSTER	0.75	-
GAKM-CLUSTER	0.76	-
KMEANS CUCKOO	0.68	1.2363
GMC	-	0.996
GRALS	-	0.945
EMBEDDING + ANN	**0.7305**	**0.9311**

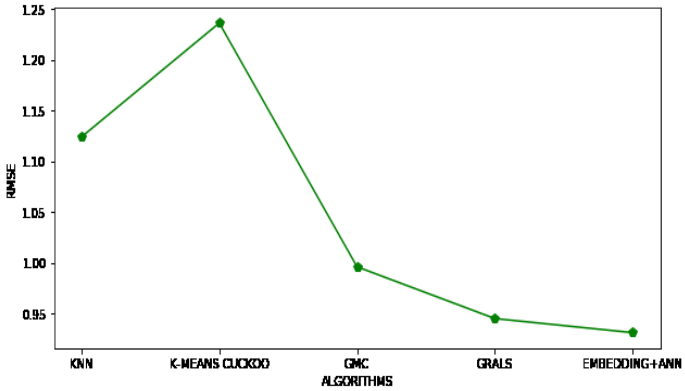


Fig 5. RMSE vs Different Algorithms

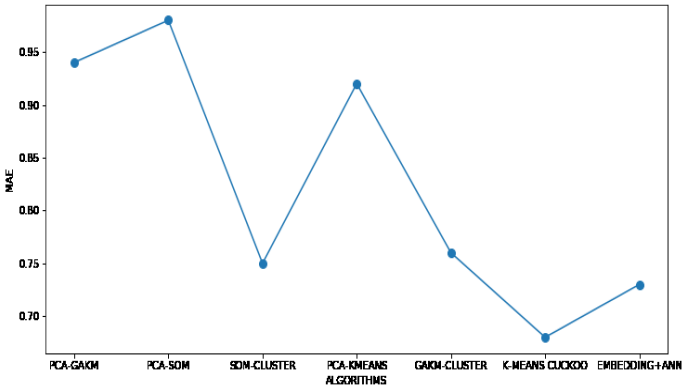


Fig 6. MAE vs Different Algorithms

V. CONCLUSION

In this paper, a collaborative recommender system using an embedding-based deep learning approach is implemented. MovieIDs and UserIDs are compared at the embedding layer to a vector of continuous values of a specific size. Then we add various layers with certain weights for processing of our input data. The hidden layers have a reLu activation function and have a dropout probability for better results. Finally, an output layer with sigmoid activation is added for the prediction of ratings. The movie ratings are predicted for the users in test data and calculated the evaluation metrics based on those predictions. Based on the evaluation metrics, our model gives better results than other machine learning algorithms in terms of MAE and RMSE values. Also, the proposed technique is compared with some state-of-the-art methods and provided more accurate results. While recommending new movies to a user, the rating based on the model is predicted and recommend the movie with higher ratings to the user. The deep learning technique with embedding layers has an additional advantage of scalability. In the future, a hybrid model will be built to give better recommendations. A combination of various machine learning algorithms can help make the recommendation system more efficient and accurate. Ensemble learning is also one of the techniques that will exploit in movie recommendation systems.

REFERENCES

[1] Katarya R, Verma OP. An effective collaborative movie recommender system with cuckoo Search. Egyptian Informatics Journal 18 (2017) 105–112.
[2] Jeffrey Lund, Yiu-Kai Ng. Movie Recommendation using deep learning approach. IEEE Computer Society, 2018.
[3] Lu J, Mao M, Wu D, Zhang G, Wang W. Recommender system application developments: a survey. Decision Support Sys 2015; 74:12–32.
[4] Gutiérrez a, Bobadilla J, Hernando a, Ortega F. Recommender systems survey. Knowledge Based System 2013; 46:109–32.
[5] Semeraro G, De Gemmis M, Lops P. Content-based recommender systems: state of the art and trends. Recommender System Handbook 2011:1–33.
[6] Billsus D, Pazzani MJ. Content-based recommendation systems 2007:325–41.
[7] W. Croft, T. Strohman, and D. Metzler. Search Engines: Information Retrieval in Practice. Addison Wesley, 2010.
[8] Zhang R, Wang Y, Bao H, Liu X, Sun H. Recommender systems based on ranking performance optimization. Front Computer Sci China 2015;1–11.
[9] Su X, Khoshgoftaar TM. A survey of collaborative filtering techniques. Adv Artificial Intelligence 2009; 2009:1–19.
[10] Zeng D, Huang Z, Chen H. A comparison of collaborative-filtering algorithms for e-commerce. IEEE Intelligent Systems 2007.
[11] Nikhil Rao, Inderjit S. Dhillon, Pradeep K. Ravikumar, Hsiang-Fu Yu. Collaborative filtering with graph information: Consistency and scalable methods. In C. Cortes, R. Garnett, N. D. Lawrence, M. Sugiyama, D. D. Lee, and editors, Advances in Neural Info Processing Syst 28, pages 2107–2115. Curran Associates, Inc., 2015.
[12] Ekstrand MD. Collaborative filtering recommender systems. Found Trends_Human-Comput Interact 2010; 4:81–173.
[13] Burke R. Hybrid web recommender systems. Adapt Web 2007.
[14] Kim BM. Clustering approach for hybrid recommender system. In: Proc IEEE/WIC Int Conf Web Intell (WI 2003). p. 33–8.
[15] A. Y. Song, Elkahky, and X. He. A Multi-View Deep Learning Approach for Cross Domain User Modeling in Recommendation Systems. In WWW, pages 278–288, 2015.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:56:24 UTC from IEEE Xplore. Restrictions apply.

[16] Chen G, Chen L, Wang F. Recommender systems based on user reviews: the state of the art. User Model. User-Adapt Interact 2015; 25:99–154.
[17] Jannach D, Nilashi M, Ithnin N, Bin Ibrahim O. Clustering and regression based multi-criteria collaborative filtering with incremental updates. Inf Sci(Ny) 2014;293:235–50.
[18] Tsapatsoulis N, Georgiou O. Improving the scalability of recommender systems by clustering using genetic algorithms. Lecture Notes Comput Sci 2010; 6352:442–9.
[19] C. Volinsky, R. Bell, Y. Koren. Matrix Factorization Techniques for Recommender Systems. Computer, 42(8):30–37, 2009.
[20] N. Wang, D.-Y. Yeung, H.Wang. Collaborative Deep Learning for Recommender Systems. In KDD, pages 1235–1244, 2015.
[21] J. Wei, K. Chen, J. He, Z. Tang, Y. Zhou. Collaborative Filtering and Deep Learning Based Recommendation System for Cold Start Items. Expert Systems with Applications, 69:29–39, 2017.
[22] F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets: History and Context. ACM Transactions on Interactive Intelligent Systems (TiiS).
[23] Vassilis Kalofolias, Pierre Vandergheynst, Michael Bronstein, Xavier Bresson. Matrix completion on graphs. arXiv preprint arXiv:1408.1717,2014.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:56:24 UTC from IEEE Xplore. Restrictions apply.

The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/1934-8835.htm

Analysis of demand risks for the Indian automotive sector in globally competitive environment

Girish Kumar
Department of Mechanical Engineering, Delhi Technological University, Delhi, India

Rajesh Kr. Singh
Management Development Institute, Gurgaon, India, and

Rishabh Jain, Raman Kain and Naveen
Department of Mechanical Engineering, Delhi Technological University, Delhi, India

Received 15 March 2020
Revised 23 April 2020
2 June 2020
11 June 2020
Accepted 11 June 2020

Abstract

Purpose – The purpose of this study is to understand the different types of risks affecting the demand for the automotive sector in India. The study is further trying to illustrate an approach for analyzing the relative intensities of these risks in the present uncertain business environment.

Design/methodology/approach – Risk on the overall demand is assessed by a combined Bayesian – multi-criteria decision-making approach. Data related to the different factors, affecting their product demand is collected from major automobile firms. Then, weights for these factors are evaluated by applying the analytic hierarchy process approach. Further, these weights are used in the Bayesian analysis network to evaluate the risk intensity for different subgroups, namely, political, economic, social, technological and environmental.

Findings – From the literature and experts' opinion, total 16 risk factors have been finalized and these are further grouped into 5 categories i.e. political, economic, social, technological and environmental. It is observed that the demand for organizations functioning in the automotive sector is more vulnerable to economic risk as compared to other risks considered in the study.

Practical implications – Managers and decision makers of associated organizations can use the proposed framework to assess the demand risks so as to pre-evaluate their demand corresponding to future changes. Factors can be added or removed and importance could be assigned to different risk factors according to the prevailing business environment for an organization or sector. This will also help the organizations to conduct a more effective risk management in an uncertain business environment.

Originality/value – The study will help in better understanding of the various demand risks prevalent in the Indian auto sector. The methodology used, provides a novel approach for assessing the macroeconomic demand risks and can be used by the firms working in the automotive sector. The proposed methodology could be used for assessing supply chain risk or any other business initiative risk. The suggested approach will help managers in devising flexible management techniques so as to mitigate the risk.

Keywords Bayesian analysis, Automotive industry, Demand, Risk modelling, MCDM, Macroeconomics.

Paper type Research paper

Authors would like to thank Editor of the journal and reviewers for valuable suggestions and comments to improve quality and content of this paper.

International Journal of Organizational Analysis
© Emerald Publishing Limited
1934-8835
DOI 10.1108/IJOA-03-2020-2076

1. Introduction

The automobile industry plays a significant role in the economic growth of any country (Ramey and Vine, 2004). According to data released by the Press Information Bureau (2019a, 2019b), Government of India, the automotive sector contributes around 7.1% to India's total gross domestic product (GDP) and is expected to grow to 12% by the year 2026. It also constitutes 49% of the manufacturing GDP and 27% of the industrial GDP. This industry is one of the largest employers and provides about 37 million direct and indirect jobs (Press Information Bureau, 2019a, 2019b). Thus, it could be inferred that the auto-component sector is one of the core contributors to the development of the Indian economy (Singh *et al.*, 2007). However, manufacturers in this industry are subjected to huge variations in the demand of their products, as, the demand is highly cyclical. Inaccurate presumption in the demands can lead to huge losses both directly and indirectly. Thus, manufacturers and dealers must properly assess and create a balance between the demand and supply and carry the optimal mix of products for each period in the business cycle. Therefore, it is very important to conduct a prior risk-based decision-making process so as to evaluate the forces that have an impact on the demand. Decision-making is a process involving interdependence in complex situations, which possess multiple (and generally opposing) objectives and uncertainty concerning the outcomes (Azis, 2005). According to Wu *et al.* (2015), risk-based decision-making has been always important for being competitive in global markets.

In India, the automotive demand began to experience a downfall from the last quarter of the calendar year 2018 and intensified with the passage of every month in 2019. According to the Society of Indian Automobile Manufacturers (SIAM), in 2018–2019, the overall vehicle sales fell by 8%, with total retail sales accounting for 16,82,656 units as compared to 18,21,538 units sold in March 2018. Two-wheeler sales fell by a sharp 17% in 2018-2019 from the previous year (SIAM). All categories saw a "de-growth" since March 2018, which was one of the biggest months for retail sales. This has also affected the automotive dealerships negatively, leading to the closure of about 300 of them across the country. Hundreds of thousands of jobs have also been cut citing slumping sales of cars and motorcycles, with many companies forced to shut down factories for days and axe shifts (ACMA, Automotive Components Manufacturers Association of India, 2020; SIAM, Society of Indian Automobile Manufacturers, 2020). The country's GDP too witnessed a slow-down, one of the main reasons for which is believed to be the decline in the automobile demand (Press Information Bureau, 2019a, 2019b). Through this study, we would like to analyze the factors that affect the auto demand globally and, thus, assess the risk proposed by each of those factors.

In the past, various studies have been done to relate the sales of automobiles to customer choices and vehicle attributes (Chand and Avikal, 2015; Aghdaie and Yousefi, 2011; Donatos and Kioulafas, 1990). Sales forecast of vehicles have also been studied using conventional models (Berkovec, 1985; Geurts and Kelly, 1986; Shahabuddin, 2009; Wang *et al.*, 2011). Some work has also been dedicated to automobile sales assessment using macroeconomic factors (Wetzel and Hoffer, 1982; Muhammad *et al.*, 2012; Zhang *et al.*, 2017). Gupta *et al.* (2019), Chirra and Kumar (2018), Zahraee *et al.* (2018), Mangla *et al.* (2015), are amongst a few researchers, who have carried out work on the supply chain in the auto sector. Their studies were limited to the management of production and the working efficiency of an organization by choosing the best supply chain network. However, they were unable to address the risks, which control the demand for its products, which are external to the direct control of an organization and sometimes even unknown to it. Nauhria *et al.* (2011), have identified the competitive priorities based on the manufacturing system and their relative importance for the Indian automobile industry. Chuang and Zhao (2019), have studied the effect of sales demand on the level of inventories for better dealership. Marx *et al.* (2020) have shown that

how governmental decision and policies have led to booming of the Brazilian automobile manufacturing industry. Study has also been done to improve the productivity of Indian automobile firms using knowledge management (Patil and Warkhedkar, 2016).

It has been observed from the above discussion and the literature review in Section 2, that enough research has not been carried out in the past to relate auto demand with the factors that lie outside the control of an organization (Chand and Avikal, 2015; Nauhria *et al.*, 2011, Patil and Warkhedkar, 2016). Auto sales are either attributed to vehicle features or to a few major macroeconomic factors or to any other practice followed by the organization explicitly, such as the supply chain network (Gupta *et al.*, 2019; Chirra and Kumar, 2018; Mangla *et al.*, 2015). Sales forecasting has been done using previously available data by applying conventional models (Wang *et al.*, 2011; Shahabuddin, 2009; Geurts and Kelly, 1986). In this study, it has been tried to understand and analyze all the factors and the risk emerging from these factors that may either explicitly or implicitly impact the demand for the entire auto-industry rather than just an organization.

This study is trying to answer the following research questions:

RQ1. What are the factors that affect the demand in automotive sector?

RQ2. How to determine the relative weights of these factors?

RQ3. How to assess the impact of overall risk in net demand?

The structure of this paper is as follows: Section 2 i.e. literature review deals with the detailed description of different risk factors affecting the demand in the auto sector and also examines the model to be used in the study. In Section 3, the research methodology is discussed. Section 4 deals with the results and numerical illustration. Section 5 discusses the obtained results, managerial implications and theoretical implications of the study. Finally, Section 6 deals with the limitations of the study and scope of future work.

2. Literature review

In this section, review analysis has been done to extract the risk factors affecting the demand and to identify the most appropriate tool for analyzing these risks. Risk factors have been finalized through literature review and experts' opinion. In the Section 2.1, each of these factors have been described briefly along with appropriate references and examples as applicable. Section 2.2 examines the model used for factor ranking. The significance of using multi-criteria decision-making (MCDM) is also discussed. Moreover, past studies on the use of the analytic hierarchy process (AHP) for ranking of criteria have been also reviewed. Section 2.3 describes the various terms regarding risk assessment and risk management. Consequently, the model to be used for risk assessment in the present study is derived.

2.1 Risk factors for automotive demand

The automotive demand is dependent on various factors. In this section, we have described these factors along with a brief discussion on the same. From the literature on automobile demand factors, and by experts' opinion, 16 risk factors have been finalized. These 16 factors have been grouped into 5 categories, namely, political, economic, social, technological and environmental. These risk factors are summarized in Table A1 given in the Appendix.

2.1.1 Political risk.

2.1.1.1 Tax policies. These policies refer to the tax rates and incentives that governments provide so as to take the control of variation in demands and consumption rates. Taxes generally include the corporate taxes and the indirect taxes that consumers have to pay

while purchasing (e.g. Goods and Services Tax). Tax policies may also include rebates and other forms of incentives that government might provide so as to increase or decrease the consumption rates. It is generally seen that demand rises as soon as the consumer starts to benefit from the policy and vice versa. Study by Pritchard and DeBoer (1995), Alberini and Bareit (2016) and Ott and Andrus (2000), show the change in consumer behaviour for car purchase with tax variation. Customers are always attracted towards the vehicles and ownership schemes that levy lower taxation rates.

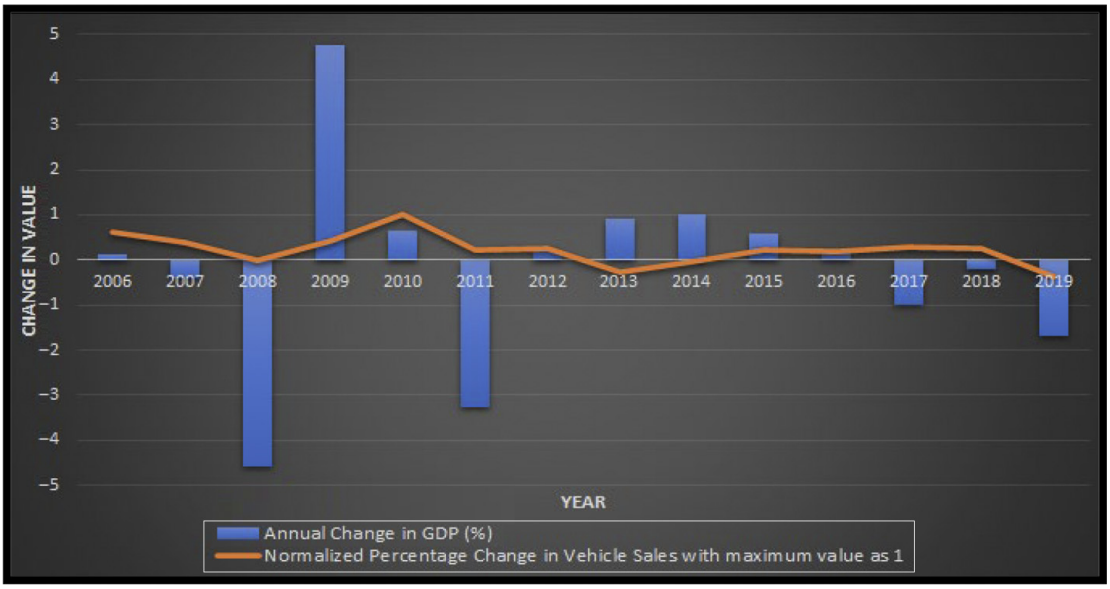
2.1.1.2 Environmental laws. These are the laws, which are implemented by the governments so as to address the problem of rapidly increasing environmental pollution. These may be different in different countries but generally, exist everywhere. These may include rules pertaining to use of vehicles that run on electricity or energy sources other than oil or shifting to modified grades of fuel. Various vehicle emission standards are also set up by the nations. According to the expert reviews demand of vehicles, generally has a negative correlation with the frequency of changes in the environmental laws as, the consumers are confused whether it is the right time to purchase or not. For example, in India one of the major reasons for stagnation in the demand of ic-engine vehicles is attributed to the government's policy to shift to Bharat Stage (BS)-VI (cleaner grade of fuel) from BS-IV by April, 2020 and to completely switch to electric vehicles by 2030 so as to reduce the air pollution.

2.1.1.3 Vehicle regulations. These are the rules and regulations set up by the governments for the manufacture and sale of new vehicles and to ply on the roads. These include some pre-requisites that the vehicle must satisfy such as the axle norms, safety standard norms and emission norms, and those which need to be followed while in use such as, the motor vehicle act. By expert reviews, it is known that sales drop with frequent changes in these rules. According to SIAM, the sales of commercial vehicles dropped by nearly 10% in May 2019, at 68,847 units as compared to 76,517 units in May 2018, which is mainly attributed to the new axle norms. Effect of government policies on the performance of automobile manufacturers was studies by Yadav and Sagar (2015).

2.1.2 Economic risk.

2.1.2.1 Economic growth and development. This imply to the rate at which the increase of per capita income takes place. It also includes, the inflation rate and the tax rates, which indirectly determine the net disposable income. New cars have a high-income elasticity of demand i.e. their demand is highly sensitive to a change in the real income of consumers, keeping all other things constant. When real incomes rise, the demand for new vehicles is expected to expand. Muhammad *et al.* (2012), Shahabuddin (2009) and Dargay (2001), had shown the impact of income rates on the sales of cars. Tambade *et al.* (2019), found that in case of Indian automobile industry business environment plays a very crucial role amongst the other factors. The graph shown in Figure 1 is plotted to simultaneously analyze the percentage change in vehicle sales with respect to the annual change in the GDP of India (Ministry of Road Transport and Highways, Government of India; Press Information Bureau, 2019a, 2019b; The World Bank). It could be clearly noticed that in most cases there exists a strong direct correlation between the two i.e. vehicle sales show a positive sign with an increase in GDP and vice versa.

2.1.2.2 Acquisition and running costs. These include the cost of the vehicle and the costs incurred while running it such as the fuel costs, insurance and maintenance. Hike in these costs lead to a decrease in the number of buyers. High fuel price, though, does not always bring a decline in demand, as a new vehicle would be technologically more advanced and fuel efficient than the older one, thus offering the buyer a chance to save money. Research by Hamilton (1983), Mork (1989), Lee and Ni (2002), Barber *et al.* (1999) and Duncan (1980), show that how significant increase in fuel price could impact car sales.



Source: Ministry of Road Transport and Highways, Government of India; PIB 2019; The World Bank

Figure 1.
Percentage change in vehicle sales w.r.t. annual change in the GDP

2.1.2.3 Inflation rate. Many economists, businessmen and other experts maintain that moderate inflation levels i.e. between 2%-3% are needed to drive the consumption, assuming that higher levels of spending are crucial for economic growth. Inflation, less than this might lead to under development and higher inflation rate may lead to significant rise in price of commodities. As auto industry makes up a good portion of a country's GDP it is necessary to maintain a low inflation rate (Chifurira *et al.*, 2014).

2.1.2.4 Availability of credit or finance. Some of the new customers may purchase a vehicle in one single payment but many mainly depend on securing a finance agreement. Even though interest rates may decrease according to government monetary policy, it still remains hard to get a new loan for a car and also, the rate of interest on unsecured credit keeps on rising. Thus, the high real rate of interest on car loans does not allow new buyers to purchase. Ludvigson (1998), had found that an increase in the rate of interest on loans poses a significant negative impact on car sales. A similar observation was also made by Muhammad *et al.* (2012), in their research on the analyzing the impact of economic variables on automobile sales.

2.1.2.5 Employment rate. Muhammad *et al.* (2012) and Shahabuddin (2009) have studied the effect of rate of employment on the sales of new vehicles. Better rate of employment means, more money in the hands of consumers leading to better consumption rates, and thus, increased demand. This process is cyclical in itself, i.e. decline in employment rate leads to lower demands and lower demand leads to again decline in the employment. Thus, the reason becomes the consequence of the problem. This is valid for the auto sector as well.

2.1.3 Social risk.

2.1.3.1 Mass transit systems. According to a few experts, less availability of good public transport system leads to increased demand of personal vehicles. Similarly, a good transport facility leads to less demand of automobiles. Improvement of public transport and reduction in distance to the nearest railway station leads to less dependence on private vehicles (Rith *et al.*, 2019). A well-planned metro network leads to decrease in private vehicle ownership by 2% ∼ 3% (Mulalic *et al.*, 2016).

2.1.3.2 Cab ridership services. Similar to the mass transit system, better cab facility leads to decreased demand of private vehicles. Consumers, generally prefer riding a cab, rather than driving themselves. Cab services are growing better and better by providing comfortable rides at competitive prices. This has led to a risk for the automobile industry. Rawal and Sisodia (2017), have shown that how the acceptance of cab services have increased in India.

2.1.3.3 Vicinity of roads and availability of parking. Ample of driving and parking space leads to comfort and helps in saving a lot of time. Thus, it encourages consumers to purchase new vehicles. On the other hand, congestion and lack of parking, forces users to avail mass transit and cab services. With rapidly increasing population and speedy urbanization, a well-developed transport system is required as individual vehicles cannot be copped up in such a large volume (Barbero and Galeota, 2019).

2.1.4 Technological risk.

2.1.4.1 Introduction of new vehicles. With introduction of vehicles with new designs and features, it is generally expected that the sales would also rise. Consumers who are pioneer innovators and early adopters are attracted towards new and modular looks. If the vehicles have an upper edge above the past running models, it would generate good reviews and demands would shoot up. Automobile industry is fast changing dynamic market, the growth of which depends upon the competitive technology (Nag and De, 2020; Yadav and Sagar, 2015).

2.1.4.2 Automation. Other than the looks and design, consumers also look for sophisticated and user-friendly functions and automatic features. According to Andrea and Smith (1994) and Aghdaie and Yousefi (2011), customers look for specifications, performance and automation while selecting amongst vehicles.

2.1.4.3 Engine efficiencies. It has been seen in the past that with the use of research and development, new technologies to enhance the fuel efficiency (such as modified fuel supply systems, integrated starter generator, start-stop technology and modified gear ratios) have evolved. Such features act as a huge mode of attraction for the consumers. Customers are easily encouraged to purchase new vehicles if they have better efficiencies than their older ones. Andrea and Smith (1994) and Aghdaie and Yousefi (2011), also showed that customers have higher demands for vehicles with higher fuel efficiencies.

2.1.5 Environmental risk.

2.1.5.1 Pollution levels. With the rapidly changing climate patterns all over the world whose root cause is pollution, it becomes a big risk factor for the automotive business. An automobile, from the start of its manufacturing, throughout its use and till the end of its scrapping, leaves a giant footprint of metals, carbon and other chemically toxic wastes. It leads to rise in all forms of pollution. Manufacturers need to think of ways to mitigate this risk by pre-planning the alternative risk barriers. More movement of people and goods lead to more pollution and emission of greenhouse gases (Barbero and Galeota, 2019).

2.1.5.2 Season and climate. During a certain period of the year the demand may remain very low (such as the rainy season) while during some other time the demand suddenly shoots up (like during the festive season). Vehicle manufactures should be ready to bear such fluctuations in the demand by prior mitigation strategies. Busse *et al.* (2015), found that the purchase of an automobile is highly dependent on the weather, which is controlled by psychological effects, rather than by classical utility theory. Bloesch and Gourio (2015), too suggest that though weather has a temporary, but a significant impact on the economic activities.

2.2 Ranking of the risk factors

Demand pattern of automobiles, as observed from discussion in the previous sub-section, depends upon many different factors. Sorting of these factors according to their relative importance requires the consideration of various evaluation notions. Therefore, it is known as MCDM

problem i.e. decision-making in the presence of multiple and generally conflicting criteria (Rao, 2007). MCDM techniques have been widely used in the studies of risk analysis and management (Chand *et al.*, 2015a, 2015b, 2017). The AHP can be used for the assessment of relative weights for the given factors in a complex problem involving many criteria (Singh, 2013; Chauhan, 2016). The essence of the AHP is that human judgements, and not just the underlying information, can be used in performing the evaluations. The AHP converts these evaluations to numerical values that can be processed and compared over the entire range of the problem. A numerical weight or priority is derived for each element of the hierarchy, allowing diverse and often incommensurable elements to be compared to one another in a rational and consistent way. This capability distinguishes the AHP from other decision-making techniques (Saaty, 1980). Chhabra *et al.* (2017) have used it for analyzing green logistics alternatives in context to an automotive organization. Apak *et al.* (2012) have used AHP to research preferences of a customer for features of luxury cars. Chand and Avikal (2015), have also used AHP in their research to select medium range cars on the basis of different criteria such as cost, performance and fuel consumption Kumar and Singh (2012) have used fuzzy AHP for evaluation of third-party logistic provider. AHP has been used to analyze productivity and wastage reduction for an Indian manufacturing firm (Chauhan, 2016). Very recently, AHP has been used for factor weight calculation of internet of things and big data in customer relationship management (Abu Ghazaleh and Zabadi, 2020). Al Hammadi and Hussain (2019), have used AHP to help decision makers in public health care sector to sort out the factors that affect sustainable organizational performance. It has also been used to prioritize the indicators of managerial success (Ghosh, 2013).

In the classical AHP, a decision maker is asked to supply exact pairwise comparison ratios for each attribute in each level of the hierarchy (Saaty, 1980). However, these decisions might be imprecise due to uncertainty in human opinions in statistical context (Saaty and Vargas, 1987; Moreno-Jimenez and Vargas, 1993; Basak and Saaty, 1993; MacKay *et al.*, 1996). Outcomes of real-life problems are usually not bivalent i.e. there is no clear-cut solution such as "yes" or "no" (Tong and Bonissone, 1980). Zadeh (1965), had introduced the concept of fuzzy set theory (FST) in decision-making, to show the linguistic variables on a numbered scale. The extended form of AHP method was proposed by Chang (1996). It combines the use of FST with the AHP method, i.e. the data used for AHP is a fuzzy number. Thus, when uncertainty exists, FST could be used (Leung and Cao, 2000; Taylan, 2014). Accuracy in assigning weights to some given alternatives could be increased by using fuzzy-MCDM methodology (Bellman and Zadeh, 1970). As the preferences in AHP are usually judgements of human beings based their perception, Leung and Cao (2000) believe that the fuzzy approach allows a more accurate description of the decision-making process. Torfi *et al.* (2010), in their research based on calculating weight using fuzzy AHP have found that AHP is one of the most outstanding MCDM approaches, used to obtain the relative weights. Fuzzy AHP, has also been used in studies regarding supply chain problems (Gupta *et al.*, 2019; Kumar and Rahman, 2017; Yadav and Sharma, 2016; Mangla *et al.*, 2015; Singh and Sharma, 2014). Recently, it has also been used to rank the criteria that determine the stress amongst hospitality and tourism employees (Vatankhah *et al.*, 2020). Thus, in comparison with other MCDM methods, the AHP method has widely been used in similar scenarios and the fuzzy AHP technique, could provide with more accurate results in many practical decision-making problems.

2.3 Risk assessment

Risk implies any form of consequence or the probability of a major loss (Tipu, 2017). Risk analysis is the "systematic use of available information to identify hazards and to estimate the risk to individuals, property and the environment" (IEC 60300-3-9, 1995). Risk evaluation is the:

> [. . .] process in which judgments are made on the tolerability of the risk on the basis of a risk analysis and taking into account, the factors such as, socioeconomic and environmental aspects (IEC 60300-3-9, 1995).

Risk assessment is the "overall process of risk analysis and risk evaluation" (IEC 60300-3-9, 1995). Risk management is a continuous management process with the objective to identify, analyze and assess the potential uncertainty related to an activity and to identify and introduce risk control measures to reduce or completely eliminate the potential harms to the system, as well as the environment (Rausand, 2013; Chand *et al.*, 2014). Framework for risk analysis, evaluation and assessment, given by Rausand (2013), is shown in Figure 2.

Risk analysis has three main parts, namely, to identify the possible undesired events, the likelihood of these events and the consequences of these events (Kaplan and Garrick, 1981). Singh *et al.* (2008) have observed that organizations should develop competencies to evaluate their risks for being competitive in global market. Quantitative risk assessment requires two quantities, namely, the probability of the occurrence of a loss and the magnitude of the potential loss (Ayyub, 2014). According to Rausand (2013), there are three main approaches to probability:

(1) The classical approach;

(2) The frequentist approach; and

(3) The Bayesian or subjective approach.

The first two approaches i.e. the classical and the frequentist approach are used when there is either a fixed sample space of outcomes or when the experiment can be repeated again and again. The Bayesian or the subjective approach is used when these conditions do not exist. In this, the probability is considered to be subjective i.e. there is a numerical value in the interval [0, 1] representing an individuals' degree of belief about the probability of an event to occur.

Thus, the Bayesian network analysis is a good risk assessment methodology that may be used for analyzing qualitatively, quantitatively or both, depending on the scope of the analysis. One of the earliest researches for risks assessment by using Bayesian analysis was done by Pai *et al.* (2003). Bayesian analysis has also been used for optimization of the cost (Narayanan *et al.*, 2005), to conduct diagnostics (Kao *et al.*, 2005; Kauffmann *et al.*, 2002) and to analyze the flexibility in supply chains (Milner and Kouvelis, 2005; Wu, 2005). Nielsen and Jensen (2009) and Alexander (2003), have used the Bayesian network analysis in the business field, for multi-variate and integrated analysis of the risks, for monitoring and evaluating intervention strategies. Bayesian network analysis has also been used to

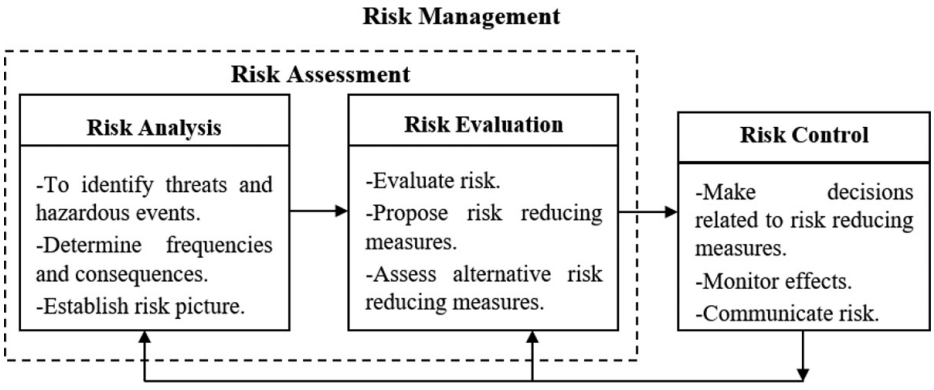


Figure 2.
Risk analysis, evaluation, assessment and management

Source: Rausand (2013)

examine the fluctuations in exchange rates caused by monetary fundamentals (Charfi *et al.*, 2019). Dalla Valle and Giudici (2008), used the Bayesian methods to integrate the scarce and inaccurate quantitative data with prior information provided by the experts to estimate operational loss distributions. In a different research, to assess the enterprise risk, Bonafede and Giudici (2007), had shown that the Bayesian networks are a useful tool to integrate different information and to study risk distribution by using the data collected from the experts. Salehi *et al.* (2020), have used the Bayesian approach to analyze the predictability of the firm level data for determining macroeconomic factors, and suggest that it is a good method, which provides results similar to the classical theory.

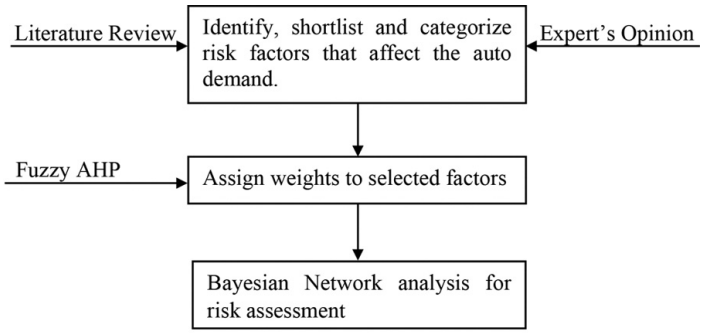
Through the above literature analysis, three important aspects regarding the significance of the present study have been drawn. Firstly, various studies have been presented to illustrate the different factors on which demand of the automotive sector depends. Different authors have worked upon different risk factors as presented in Table A2. Secondly, the application of fuzzy AHP technique has been also reviewed. Finally, to assess the probability of the risk factors under such a scenario, the Bayesian network analysis application has been studied. In an industry working under practical situations, the risk for demand may not always depend only on a few factors in a selected field, rather it is often diversified and interconnected over a vast network of factors (Kumar *et al.*, 2019). In this study, the risk through all the factors has been assessed by a combined Bayesian- MCDM model. Factors have been collected through Section 2.1, factor weights have assigned by using the fuzzy AHP technique and risk probability has been assessed using the Bayesian network analysis.

3. Research methodology

The factors affecting the sales of automotive have been identified and discussed in the final section. The methodology includes, application of fuzzy AHP and Bayesian analysis. For applying AHP, level of importance for each of the demand factor is required. This was achieved by expert survey of the decision makers of four different firms. Profile of experts is given in Table A3 in the Appendix. Thus, the obtained level is based on personal opinions and thoughts supported with some historical data, which makes the use of FST necessary. Relative weights for each of the factor is obtained by the importance level in fuzzy AHP, further the weights obtained by fuzzy AHP are used to assess the relative risk for each factor and its subgroup through the Bayesian analysis. Figure 3 shows the flowchart of the applied research methodology in this study.

3.1 Fuzzy set theory

Definition 1. A fuzzy number ($\tilde{\tilde{n}}$), which is triangular in nature, is denoted by triplet Tr (j,k,l), where [j-lower, k-middle and l-upper]. In terms of the membership function ($\mu_{\tilde{\tilde{n}}}$), the graphical representation is shown in Figure 4.



Source: Self-created

Figure 3.
Flowchart of the applied methodology

$$\mu_{\tilde{\text{n}}} = \begin{cases} \dfrac{x-j}{k-j} & ;\text{for } \text{j} \le \text{x} \le \text{k} \\ \dfrac{x-l}{l-k} & ;\text{for } \text{k} \le \text{x} \le \text{l} \\ 0 & ;\text{otherwise} \end{cases} \tag{1}$$

Definition 2. Let two fuzzy-triangular numbers be, $\tilde{\tilde{\text{x}}}_1 = (\text{j}_1, \text{k}_1, \text{l}_1)$ and $\tilde{\tilde{\text{x}}}_2 = (\text{j}_2, \text{k}_2, \text{l}_2)$, then the mathematical operations associated with these fuzzy numbers are given as following:

- "adding" two Fuzzy Tr. Numbers:

$$\begin{aligned} \tilde{\tilde{\text{x}}}_1 \oplus \tilde{\tilde{\text{x}}}_2 &= (\text{j}_1, \text{k}_1, \text{l}_1) \oplus (\text{j}_2, \text{k}_2, \text{l}_2) \\ &= \left[(\text{j}_1 + \text{j}_2), (\text{k}_1 + \text{k}_2), (\text{l}_1 + \text{l}_2)\right. \end{aligned} \tag{2}$$

- "subtracting" two Fuzzy Tr. Numbers:

$$\begin{aligned} \tilde{\tilde{\text{x}}}_1 \ominus \tilde{\tilde{\text{x}}}_2 &= (\text{j}_1, \text{k}_1, \text{l}_1) \ominus (\text{j}_2, \text{k}_2, \text{l}_2) \\ &= \left[(\text{j}_1 - \text{j}_2), (\text{k}_1 - \text{k}_2), (\text{l}_1 - \text{l}_2)\right. \end{aligned} \tag{3}$$

- "multiplication" of two Fuzzy Tr. Numbers:

$$\begin{aligned} \tilde{\tilde{\text{x}}}_1 \otimes \tilde{\tilde{\text{x}}}_2 &= (\text{j}_1, \text{k}_1, \text{l}_1) \otimes (\text{j}_2, \text{k}_2, \text{l}_2) \\ &= \left[(\text{j}_1 \times \text{j}_2), (\text{k}_1 \times \text{k}_2), (\text{l}_1 \times \text{l}_2)\right. \end{aligned} \tag{4}$$

- "dividing" two Fuzzy Tr. Numbers:

$$\begin{aligned} \tilde{\tilde{\text{x}}}_1 / \tilde{\tilde{\text{x}}}_2 &= (\text{j}_1, \text{k}_1, \text{l}_1) / (\text{j}_2, \text{k}_2, \text{l}_2) \\ &= \left[(\text{j}_1 / \text{j}_2), (\text{k}_1 / \text{k}_2), (\text{l}_1 / \text{l}_2)\right. \end{aligned} \tag{5}$$

Definition 3. Distance between two fuzzy numbers:

Chen (2000), gave the vertex method for finding the distance between to fuzzy Tr. Numbers.

$$\text{d}\left(\tilde{\tilde{\text{x}}}_1, \tilde{\tilde{\text{x}}}_2\right) = \left\{\frac{1}{3}\left[(\text{j}_1 - \text{j}_2)^2 + (\text{k}_1 - \text{k}_2)^2 + (\text{l}_1 - \text{l}_2)^2\right]\right\}^{1/2} \tag{6}$$

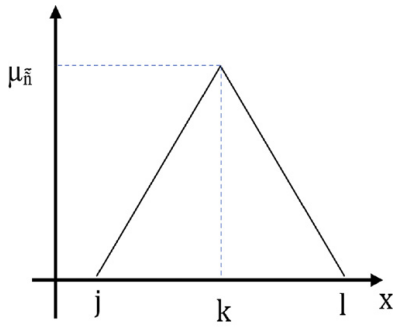


Figure 4.
Depiction of Fuzzy-Triangular number $\tilde{\tilde{x}} = (\text{j}, \text{k}, \text{l})$

Source: Rao (2007)

Definition 4. De-fuzzifying the fuzzified values:

Let $\tilde{\tilde{x}} = (j, k, l)$ represent a fuzzy number. The fuzzy number can defuzzied by the centre of gravity method given in equation (6) Jager *et al.* (1992).

$$\text{De fuzzify}\left(\tilde{\tilde{y}}_i\right) = \frac{1}{3}[j + k + l] \tag{7}$$

3.2 Fuzzy analytic hierarchy process method

As discussed in the literature review, in this study, the AHP technique is applied in its extended form, to determine the weights of the concerned factors. Thus, the geometric mean method of fuzzy AHP is used and the detailed procedure for the same is given in this section. The conventional basic AHP method is not capable to handle the imprecision of human judgments. While, fuzzy AHP, an improved form of AHP is able to handle this issue. The scale used for this analysis is given in Table 1.

In the AHP method, factor importance is first obtained in linguistic terms from decision makers from different firms or organizations based on their level of importance. Then, an aggregated pairwise comparison matrix is constructed amongst the considered factors for their weight calculation. The level of importance obtained from the expert review are converted into triangular fuzzy numbers (TFNs) as per the Table 1.

The complete process of fuzzy AHP, used in this study is explained in five steps:

Let the TFN is represented by:

$$\tilde{\tilde{x}}_i^r = (j_r, k_r, l_r), \text{ where, } r = (1, 2, 3, \ldots., m)$$

Step 1: Calculate the fuzzy geometric mean value, $\tilde{\tilde{r}}$,

$$\tilde{\tilde{r}}_i = \left((j_{i1} \times j_{i2} \times \ldots. \times j_{im})^{1/m}, (k_{i1} \times k_{i2} \times \ldots. \times k_{im})^{1/m}, (l_{i1} \times l_{i2} \times \ldots. \times l_{im})^{1/m}\right. \tag{8}$$

Step 2: Calculate the sum of all the fuzzy geometric mean values,

$$\sum_{i=1}^{m} \tilde{\tilde{r}}_I = [r_1 \oplus r_2 \oplus r_b \oplus \ldots\ldots. \oplus r_m] \tag{9}$$

Step 3: Calculate the fuzzy weights, $\tilde{\tilde{w}}_I$, for all the fuzzy geometric mean values,

$$\tilde{\tilde{w}}_i = r_i \otimes [r_1 \oplus r_2 \oplus r_b \oplus \ldots\ldots. \oplus r_m]^{-1} \tag{10}$$

Step 4: Defuzzify the fuzzy weights to obtain weights, w_i,

$$w_i = \frac{1}{3}[j_{w1} + k_{w2} + l_{w3}] \tag{11}$$

Level of importance	Triangular fuzzy number
1	(1, 1, 1)
2	(1, 2, 3)
3	(2, 3, 4)
4	(3, 4, 5)
5	(4, 5, 6)

Source: Kannan *et al.* (2013)

Table 1. Triangular fuzzy number of a given level of importance

Step 5: Obtain the normalized weights, (w_{ni}) by following,

$$w_{ni} = w_i / \sum w \tag{12}$$

3.3 Bayesian analysis

A Bayesian network graphically represents the probability distribution for a large set of variables (Pai *et al.*, 2003). In a Bayesian network, nodes depict random variables. Nodes that are not connected imply that those variables are independent of each other (Heckerman, 1996). Each node, associated with a probability function takes a particular set of values as input and gives the probability of the variable represented by the node as output (Heckerman, 1996; Nielsen and Jensen, 2009).

The Bayes' Theorem was given by Thomas Bayes in the late 1700*s* who worked on the concept of conditional probability. It states that:

$$P(H|E,r) = \frac{P(H\ |r) \times P(E|H,r)}{P(E|r)} \tag{13}$$

The equation, i.e. P(H|E,*r*) represents the probability of any hypothesis H after considering the occurrence of event E on any past event r. The term P (H|*r*) is the probability of occurrence of H given that only *r* occurs. The likelihood is represented by the term P (E|H,*r*), which gives the probability of the E assuming H and *r* are true. The term, P(E|*r*) is independent of H and is regarded as a normalizing or scaling factor (Niedermayer, 2008). Thus, we see that the Bayesian network analysis could be used for combining subjective a hypothesis with the available evidence. A combined diagram for the AHP and Bayesian Network Analysis is presented in Figure 5.

4. Results and analysis

Results are analyzed in two parts. First, we obtain the relative weights for each of the factors by applying the fuzzy AHP MCDM technique on the data received from industry experts.

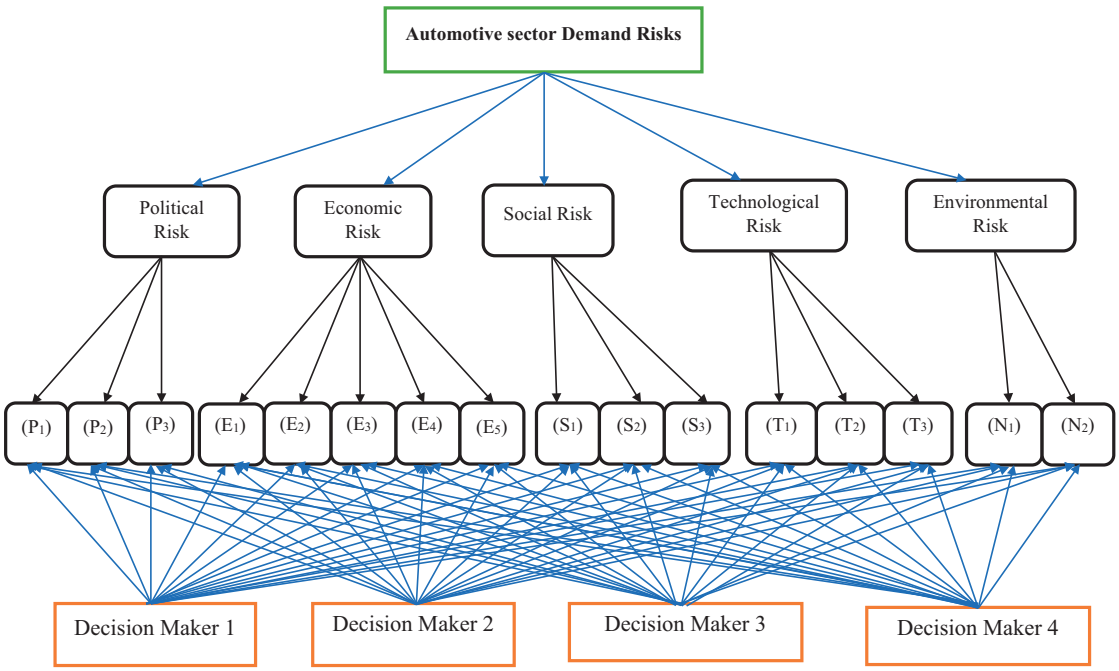


Figure 5.
Integrated AHP and Bayesian Network for automotive risk assessment

Secondly, these weights are used in the Bayesian network to assess the risk for each of the node, and then on each category of risk. The numerical illustration for the same is shown in the following section.

4.1 Numerical illustration

First, the weights for all the risk factors based on their relative importance were collected from four decision makers. Importance of different factors as obtained from the decision makers is given in Table A4 in the Appendix.

4.1.1 Analysis based on fuzzy analytic hierarchy process. Steps to obtain the final ranking of factors affecting the demand of automotive have been illustrated:

Step 1: Following the procedure explained in the research methodology, an integrated pairwise comparison matrix for all the factors (based on expert reviews) is obtained and is presented in Table A5 in the Appendix.

Step 2: By using equation (8), the Fuzzy Geometric Mean ($\tilde{r}$) for each fuzzy number with respect to each of the factors are obtained. For example:

$$\tilde{r}_i = (1 \times 1 \times \ldots \times 1)^{1/16}, (1 \times 1 \times \ldots \times 2)^{1/16}, (1 \times 1 \times \ldots \times 3)^{1/16}$$
$$\Rightarrow \tilde{r}_i = (1.000, 1.500, 2.059)$$

Step 3: Using equation (10), the fuzzy weights, ($\tilde{\tilde{w}}_i$), for all the fuzzy geometric mean values are calculated. For example:

$$\tilde{\tilde{w}}_1 = (1.000, 1.500, 2.059) \otimes [(1.000, 1.500, 2.059) \otimes (1.000, 1.500, 2.059) \otimes \ldots\ldots \ldots \otimes (0.642, 0.760, 1.000)]^{-1}$$
$$\Rightarrow \tilde{\tilde{w}}_1 = (0.04317, 0.08823, 0.16268)$$

Step 4: Using equation (11), each of the fuzzy weights ($\tilde{\tilde{w}}_i$), are defuzzified, to obtain the crisp weights (w_i). For example,

$$w_1 = \frac{1}{3}[0.04317 + 0.08823 + 0.16268]$$
$$\Rightarrow w_1 = 0.0980266$$

Step 5: Final value of normalized weights, (w_{ni}) are obtained using equation (12),

$$w_{n1} = 0.0980266/(0.0980266 + 0.0980266 + \ldots\ldots + 0.0504733)$$
$$\Rightarrow w_{n1} = 0.0870986$$

Calculated values of normalized weights are given in Table A6 in the Appendix.

4.1.2 Analysis based on bayesian network. Normalized weights obtained by fuzzy AHP are given in Table A6 in Appendix. These are used to calculate their respective probability of impact on the demand by solving them through Bayesian network analysis.

Using the formula given in equation (13), the risk for each of the category is calculated,

$$P(\text{Political Risk}) = \frac{\Sigma(\textit{Probability of political risk factors}) \times (\textit{Probability of event occurence})}{\Sigma(\textit{Probability of event occurence})}$$

$$\Rightarrow \mathrm{P(Political\,Risk)} = (P_1 + P_2 + P_3)/(1 + 1 + 1)$$
$$= (0.0732081 + 0.0448465 + 0.0732081)/3$$
$$= 0.06375423$$

Similarly, the risk for all the categories is calculated. The results have been tabulated in Table A7 in the Appendix.

Using the results obtained after applying the Bayesian analysis, different categories of risk are arranged in decreasing order of their magnitude in Table 2. It is observed that economic risk is the most important risk category followed by political, social, technological and environmental risks, respectively.

5. Discussion of findings

This study identifies the different types of risk factors impacting the demand in the automotive industry and presents an approach for modelling and evaluating risk factors for the demand in this sector. By the help of literature review and experts' opinion, total 16 risk factors are identified and these are further grouped into 5 categories i.e. political, economic, social, technological and environmental. Risk profile on the overall demand was assessed by a combined technique of fuzzy AHP and Bayesian network analysis approach. Relative weights for these factors are evaluated by applying the fuzzy AHP approach on the basis of the inputs received from expert interviews. Further, these weights are used in the Bayesian analysis to evaluate the risk for the different subgroups and its impact on the final demand is assessed. After the application of Bayesian analysis, it is observed that organizations are more vulnerable to economic risk. It means, organizations should try to control major dimensions of economic risks such as acquisition and running cost, inflation rate and availability of credit or finance carefully. Thus, from this study, it can be inferred that the changes in demand are more driven by economic factors.

Economic risks include- the economic growth and development, acquisition and running costs, inflation rate, availability of credit or finance and the employment rate. Figure 1 validates the finding that vehicle sales largely depend on the GDP. Study by Muhammad *et al.* (2012) also shows that in the Association of Southeast Asian Nations countries, these economic factors have a noteworthy deep-rooted correlation with the automobile sales. Further, Dargay and Gately (1999), have found that in India, being a developing country, the percentage change in vehicle sales was even more than twice the percentage change in GDP, in between 1966 and 1992. Results by Patra and Rao (2017), in their study regarding economic factors using co-integration and vector models suggest that GDP has a direct relation with the automobile demand, while, the prices and lending rates have an inverse relationship.

Economic risk is followed by political risk, which contain the tax policies, environmental laws and vehicle regulations. This type of risk is again outside the direct control of the firms and possess a huge amount of risk for the demand. Studies suggest that these factors too

Table 2. Risk categories in descending order of risk

Rank	Risk category
1	Economic risk
2	Political risk
3	Social risk
4	Technological risk
5	Environmental risk

contribute greatly to the demand of the automotive (Fridstrøm and Østli, 2017; Clibborn *et al.*, 2016; Alberini and Bareit, 2016; Ott and Andrus, 2000; Pritchard and DeBoer, 1995). Further, from the studies itself, it could be seen that though social and technological factors also impact the overall demand, the intent of these risks is not that high and often lies between 2%-3% (Mulalic *et al.*, 2016; Rith *et al.*, 2019; Barbero and Galeota, 2019; Rawal and Sisodia, 2017; Aghdaie and Yousefi, 2011; Andrea and Smith, 1994). Amongst these factors, the technological ones could be directly or indirectly controlled by an organization itself. Chatterjee *et al.* (2020), suggest that competency in these factors is important for an organization to expand its customer base. At last, the seasonal factors, may have a control on the demand only for a short range of time. These could easily be anticipated by the organizations and its effects could simply be mitigated. Team of experts considered for this study were again consulted with final results of the study. All the experts agreed with final findings and its significance for the industry.

5.1 Managerial implications

The analysis of different types of risks and assessment of the overall risk to the demand in automotive sector will help the industries in this sector in devising their strategies for better gains. The model used in this study would help the associated practitioners to easily assess the risk through each factor at any moment by using appropriate support data. This will, in turn, assist them in a better preparation for facing the future circumstances (for e.g. a sharp surge or decline in the demand) and to deal with the risk situation with accurate mitigation plans by the anticipation of the future demand. For a better efficiency and increased profits, the factors and their categories having higher risk should be given priority to be reduced earlier. The organizations need to devise and implement their policies accordingly. This study reveals that economic risk has the highest magnitude, therefore organizations need to modify their plans more often so as to move hand in hand with the economic changes. These changes majorly include- the economic growth, costs, inflation rate, liquidity and employment. As discussed earlier, the demand increases with economic growth and decreases with an increase in costs, inflation and unemployment. Thus, the rate of production should go hand in hand with the data regarding these aspects. Deviation from the appropriate rate of production may eventually lead to huge losses for the organization. This is valid for all other categories of risk as well. Hence, prioritizing risk factors becomes an essentially important step for risk assessment and its management.

5.2 Theoretical implications

This study has a threefold theoretical contribution to the literature regarding automobile demand and risk assessment. The research contributes:

- to understand the risk factors affecting demand of automotive;
- to study a model that prioritizes these factors by assigning relative weights to them; and
- to assess the risk from each of these factors and correspondingly the overall risk on automotive demand.

Till present there is no single study available where all the external factors affecting the automotive demand have been discussed. In the current study, these factors have been collected through thorough literature and expert reviews. Relative weights have been assigned by applying the fuzzy AHP MCDM technique by using human judgements supported with historical data, and eventually risk probability of each factor has been assessed using the Bayesian network analysis. Thus, it lays the foundation for development

of a new Bayesian MCDM hybrid model for calculation of probability of occurrence of different phenomenon in other areas of study as well. In any sector or industry, it is often required to analyze different factors affecting the outcome of a phenomena, to prioritize the factors and to manipulate their depth of occurrence on the event. The applied methodology provides immense help to the user in assessing and prioritizing factors that lie outside the limits of his direct control. This model would also be extremely useful when an experiment cannot be conducted again and again or when a large sample space of data cannot be gathered. It would also be helpful, when the list of factors varies for organizations or when an organization wants to vary its own importance for the different factors. A successful combination of the fuzzy AHP and Bayesian network is, hence, found to be a very good tool for risk or probability evaluation. Other researchers can use this integrated approach in context to other sectors for analyzing the risks.

6. Limitations and future scope

While the study is very helpful in analyzing the demand risk for the automobile industry, it has a few limitations too. One of the limitations of this study is that the risk factors taken into consideration for assessing are based on literature review and experts' opinions. Therefore, these factors or their extent, could change in the future depending upon the sector and business environment. If a new factor, identical to a non-optimal factor, is introduced, then the ranking of the existing factor may change. Belton and Gear (1983), suggest that the reason for this inconsistency in ranking is that the relative performance measures of all alternatives in terms of each attribute is summed to one. Instead of having the relative performance values sum up to one, dividing each relative performance value by the maximum value in the corresponding vector of relative values was proposed. This avoids the rank reversals when a new non-optimal factor is added. Sometimes it has also been observed that the AHP is not able to include the effects resulting from inter-reliant factors (Gurumurthy *et al.*, 2013). So, there is a necessity to present the interdependencies amongst the factors. In this study, a large sample space of data is not available, as only a limited number of experts could be contacted to gather the data. Though, it has been tried to be resolved by using FST, but there may still be some scope of error, which could be easily removed if it is applied to an industry practically. Other limitation is that it does not provide the mitigation plans for the assessed risks. In the future these mitigation plans can also be devised so as to negate the risk from each of the factors. The suggested approach may also help managers in devising risk mitigation plans. Continuing with this, processes for lean and agile systems can also be implemented to overcome these risks. Future studies may be carried out for finding new risks and possible mitigation strategies in emerging business environment of circular economy and Industry 4.0. Findings may be further validated through case and empirical studies.

The applied methodology can be used to evaluate the factors that affect the demand and the overall risk imposed by them on the auto industry. Organizations need to maintain balance between the production and change in forces of factors affecting demand. Managers and decision makers of associated organizations can use the proposed framework to pre-evaluate their demands corresponding to the future changes. Anticipated demands will help significantly in cost-saving and also in maximizing the profits and to increase the efficiency. This study provides a novel approach for assessing the demand risks. Similar approach can be also applied for other manufacturing sectors. This study can also be extended to assess the physical factors that affect the production rate in an industry. Subjectivity caused due to limited number of experts can be solved by empirical or case-based studies.

References

Abu Ghazaleh, M. and Zabadi, A. (2020), "Promoting a revamped CRM through Internet of Things and Big Data: an AHP-based evaluation", *International Journal of Organizational Analysis*, Vol. 28 No. 1, pp. 66-91.

ACMA, Automotive Components Manufacturers Association of India (2020), "Message", available at: www.acma.in/message.php?id=262# (accessed 11 April 2020).

Aghdaie, S.F.A. and Yousefi, E. (2011), "The comparative analysis of affecting factors on purchasing domestic and imported cars in Iran market-using AHP technique", *International Journal of Marketing Studies*, Vol. 3 No. 2, pp. 142.

Akbari, F., Salehi, M. and Vlashani, M.A.B. (2018), "The relationship between tax avoidance and firm value with income smoothing – a comparison between classical and Bayesian econometric in multilevel models", *International Journal of Organizational Analysis*, Vol. 27 No. 1, pp. 125-148.

Al Hammadi, F. and Hussain, M. (2019), "Sustainable organizational performance: a study of health-care organizations in the United Arab Emirates", *International Journal of Organizational Analysis*, Vol. 27 No. 1, pp. 169-186.

Alberini, A. and Bareit, M. (2016), "The Effect of Registration Taxes on New Car Sales and Emissions: Evidence from Switzerland. CER-ETH Economics Working Paper, Vol. 16, p. 245.

Alexander, C. (Ed.). (2003), *Operational Risk: regulation, Analysis and Management*, Pearson Education, Great Britain.

Andrea, D.J. and Smith, B.C. (1994), "Automotive demand, markets, and material selection processes", *SAE Transactions*, pp. 435-443.

Apak, S., Göğüş, G.G. and Karakadılar, İS. (2012), "An analytic hierarchy process approach with a novel framework for luxury car selection", *Procedia – Social and Behavioral Sciences*, Vol. 58, pp. 1301-1308.

Ayyub, B.M. (2014), *Risk Analysis in Engineering and Economics*, Chapman and Hall/CRC, Taylor and Francis Group, FL.

Azis, I.J. (2005), "A regional cooperation to support financial crisis management and prevention: an application of a model with uncertainty and feedback influences", *International Journal of Organizational Analysis*, Vol. 13 No. 3, pp. 196-215.

Barber, B.M., Click, R.W. and Darrough, M.N. (1999), "The impact of shocks to exchange rates and oil prices on US sales of American and Japanese automakers", *Japan and the World Economy*, Vol. 11 No. 1, pp. 57-93.

Barbero, J.A. and Galeota, C. (2019), "Urban transportation", *The Wiley Blackwell Encyclopedia of Urban and Regional Studies*, pp. 1-12.

Basak, I. and Saaty, T. (1993), "Group decision making using the analytic hierarchy process", *Mathematical and Computer Modelling*, Vol. 17 Nos 4/5, pp. 101-109.

Bellman, R.E. and Zadeh, L.A. (1970), "Decision-making in a fuzzy environment", *Management Science*, Vol. 17 No. 4, pp. B-141.

Belton, V. and Gear, T. (1983), "On a short-coming of saaty's method of analytic hierarchies", *Omega*, Vol. 11 No. 3, pp. 228-230.

Berkovec, J. (1985), "Forecasting automobile demand using disaggregate choice models", *Transportation Research Part B: Methodological*, Vol. 19 No. 4, pp. 315-329.

Bloesch, J. and Gourio, F. (2015), "The effect of winter weather on US economic activity", *Economic Perspectives*, Vol. 39 No. 1.

Bonafede, C.E. and Giudici, P. (2007), "Bayesian networks for enterprise risk assessment", *Physica A: Statistical Mechanics and Its Applications*, Vol. 382 No. 1, pp. 22-28.

Busse, M.R., Pope, D.G., Pope, J.C. and Silva-Risso, J. (2015), "The psychological effect of weather on car purchases", *The Quarterly Journal of Economics*, Vol. 130 No. 1, pp. 371-414.

Chand, M. and Avikal, S. (2015), "An MCDM based approach for purchasing a car from indian car market", in *2015 IEEE Students Conference on Engineering and Systems (SCES)*, IEEE, pp. 1-4.

Chand, M., Raj, T. and Shankar, R. (2014), "Analysing the operational risks in supply chain by using weighted interpretive structure modelling technique", *International Journal of Services and Operations Management*, Vol. 18 No. 4, pp. 378-403.

Chand, M., Raj, T. and Shankar, R. (2015a), "A comparative study of multi criteria decision making approaches for risks assessment in supply chain", *International Journal of Business Information Systems*, Vol. 18 No. 1, pp. 67-84.

Chand, M., Raj, T. and Shankar, R. (2015b), "Risk mitigations strategy in supply chain planning and control: an ANP approach", *International Journal of Productivity and Quality Management*, Vol. 16 No. 1, pp. 92-113.

Chand, M., Raj, T., Shankar, R. and Agarwal, A. (2017), "Select the best supply chain by risk analysis for Indian industries environment using MCDM approaches", *Benchmarking: An International Journal*, Vol. 24 No. 5, pp. 1400-1413.

Chang, D.Y. (1996), "Applications of the extent analysis method on fuzzy AHP", *European Journal of Operational Research*, Vol. 95 No. 3, pp. 649-655.

Charfi, S., BenHamad, S. and Masmoudi, A. (2019), "Assessing the impact of monetary fundamentals on exchange rate fluctuations a Bayesian network approach", *Journal of Modelling in Management*, Vol. 15 No. 1, pp. 166-181.

Chatterjee, S., Chaudhuri, R., Vrontis, D., Thrassou, A., Ghosh, S.K. and Chaudhuri, S. (2020), "Social customer relationship management factors and business benefits", *International Journal of Organizational Analysis*, doi: 10.1108/IJOA-11-2019-1933.

Chauhan, G. (2016), "An analysis of the status of resource flexibility and lean manufacturing in a textile machinery manufacturing company", *International Journal of Organizational Analysis*, Vol. 24 No. 1, pp. 107-122.

Chen, C.T. (2000), "Extensions of the TOPSIS for group decision-making under fuzzy environment", *Fuzzy Sets and Systems*, Vol. 114 No. 1, pp. 1-9.

Chhabra, D., Garg, S.K. and Singh, R.K. (2017), "Analysing alternatives for green logistics in an Indian automotive organization: a case study", *Journal of Cleaner Production*, Vol. 167 No. 20, pp. 962-969.

Chifurira, R., Mudhombo, I., Chikobvu, M. and Dubihlela, D. (2014), "The impact of inflation on the automobile sales in South Africa", *Mediterranean Journal of Social Sciences*, Vol. 5 No. 7, pp. 200-200.

Chirra, S. and Kumar, D. (2018), "Evaluation of supply chain flexibility in automobile industry with fuzzy dematel approach", *Global Journal of Flexible Systems Management*, Vol. 19 No. 4, pp. 305-319.

Chuang, C.H. and Zhao, Y. (2019), "Demand stimulation in finished-goods inventory management: Empirical evidence from general motors dealerships", *International Journal of Production Economics*, Vol. 208, pp. 208-220.

Clibborn, S., Lansbury, R.D. and Wright, C.F. (2016), "Who killed the Australian automotive industry: the employers, government or trade unions?", *Economic Papers: A Journal of Applied Economics and Policy*, Vol. 35 No. 1, pp. 2-15.

Dalla Valle, L. and Giudici, P. (2008), "A Bayesian approach to estimate the marginal loss distributions in operational risk management", *Computational Statistics and Data Analysis*, Vol. 52 No. 6, pp. 3107-3127.

Dargay, J.M. (2001), "The effect of income on car ownership: evidence of asymmetry", *Transportation Research Part A: Policy and Practice*, Vol. 35 No. 9, pp. 807-821.

Dargay, J. and Gately, D. (1999), "Income's effect on car and vehicle ownership, worldwide: 1960-2015", *Transportation Research Part A: Policy and Practice*, Vol. 33 No. 2, pp. 101-138. No.

Donatos, G.S. and Kioulafas, K.E. (1990), "A quantitative analysis of new car sales and advertising in Greece", *European Journal of Operational Research*, Vol. 48 No. 3, pp. 311-317.

Duncan, R.S. (1980), "The effect of gasoline prices on automobile sales", *The American Economist*, Vol. 24 No. 1, pp. 62-66.

Fridstrøm, L. and Østli, V. (2017), "The vehicle purchase tax as a climate policy instrument", *Transportation Research Part A: Policy and Practice*, Vol. 96, pp. 168-189.

Geurts, M.D. and Kelly, J.P. (1986), "Forecasting retail sales using alternative models", *International Journal of Forecasting*, Vol. 2 No. 3, pp. 261-272.

Ghosh, K. (2013), "Prioritization of individual success indicators for managerial success", *International Journal of Organizational Analysis*, Vol. 21 No. 3, pp. 260-287.

Gupta, S., Soni, U. and Kumar, G. (2019), "Green supplier selection using multi-criterion decision making under fuzzy environment: a case study in automotive industry", *Computers and Industrial Engineering*, Vol. 136, pp. 663-680.

Gurumurthy, A., Mazumdar, P. and Muthusubramanian, S. (2013), "Graph theoretic approach for analysing the readiness of an organisation for adapting lean thinking: a case study", *International Journal of Organizational Analysis*, Vol. 21 No. 3, pp. 396-427.

Hamilton, J.D. (1983), "Oil and the macroeconomy since world war II", *Journal of Political Economy*, Vol. 91 No. 2, pp. 228-248.

Heckerman, D. (1996), "A tutorial on learning with bayesian networks (tech. Rep. No. MSR-TR-95-06)", *One Microsoft Way, Redmond, WA, 98052*.

IEC 60300-3-9 (1995), "IEC 60300-3-9, international electrotechnical commission. (1995), "IEC 60300-3-9: 1995 risk management: Part. 3 guide to risk analysis of technological systems", IEC, Geneva.

Jager, R., Verbruggen, H.B. and Bruijn, P.M. (1992), "The role of defuzzification methods in the application of fuzzy control", *IFAC Proceedings Volumes*, Vol. 25 No. 6, pp. 75-80.

Kannan, D., Khodaverdi, R., Olfat, L., Jafarian, A. and Diabat, A. (2013), "Integrated fuzzy multi criteria decision making method and multi-objective programming approach for supplier selection and order allocation in a green supply chain", *Journal of Cleaner Production*, Vol. 47, pp. 355-367.

Kao, H.Y., Huang, C.H. and Li, H.L. (2005), "Supply chain diagnostics with dynamic Bayesian networks", *Computers and Industrial Engineering*, Vol. 49 No. 2, pp. 339-347.

Kaplan, S. and Garrick, B.J. (1981), "On the quantitative definition of risk", *Risk Analysis*, Vol. 1 No. 1, pp. 11-27.

Kauffmann, P.J., Jacobs, D.A. and Fernandez, A.A. (2002), "Use of bayesian probabilities to identify and improve distribution center error rates", *Production and Inventory Management Journal*, pp. 1-6.

Kumar, D. and Rahman, Z. (2017), "Analyzing enablers of sustainable supply chain: ISM and fuzzy AHP approach", *Journal of Modelling in Management*, Vol. 12 No. 3, pp. 498-524.

Kumar, P. and Singh, R.K. (2012), "A fuzzy AHP and TOPSIS methodology to evaluate global 3PL", *Journal of Modelling in Management*, Vol. 7 No. 3, pp. 287-303. VolNo.

Kumar, P., Kumar, D. and Chand, M. (2019), "Analysing the interactions of operational issues in sustainable supply chain: a framework", *International Journal of Services and Operations Management*, Vol. 32 No. 4, pp. 507-521.

Lee, K. and Ni, S. (2002), "On the dynamic effects of oil price shocks: a study using industry level data", *Journal of Monetary Economics*, Vol. 49 No. 4, pp. 823-852.

Leung, L.C. and Cao, D. (2000), "On consistency and ranking of alternatives in fuzzy AHP", *European Journal of Operational Research*, Vol. 124 No. 1, pp. 102-113.

Ludvigson, S. (1998), "The channel of monetary transmission to demand: evidence from the market for automobile credit", *Journal of Money, Credit and Banking*, Vol. 30 No. 3, pp. 365-383.

MacKay, D.B., Bowen, W.M. and Zinnes, J.L. (1996), "A thurstonian view of the analytic hierarchy process", *European Journal of Operational Research*, Vol. 89 No. 2, pp. 427-444.

Mangla, S.K., Kumar, P. and Barua, M.K. (2015), "Flexible decision modeling for evaluating the risks in green supply chain using fuzzy AHP and IRP methodologies", *Global Journal of Flexible Systems Management*, Vol. 16 No. 1, pp. 19-35.

Marx, R., de Mello, A.M. and de Lara, F.F. (2020), "The new geography of the automobile industry: trends and challenges in Brazil", *In New Frontiers of the Automobile Industry*, Palgrave Macmillan, Cham, pp. 349-375.

Milner, J.M. and Kouvelis, P. (2005), "Order quantity and timing flexibility in supply chains: the role of demand characteristics", *Management Science*, Vol. 51 No. 6, pp. 970-985.

Ministry of Road Transport and Highways, Government of India (2020), "Annual report 2018-2019", "Total Number of Registered Motor Vehicles in India: 2003-2017", p. 115, available at: https://morth.nic.in/sites/default/files/Annual_Report_English_2018-19.pdf" (accessed 11 April 2020).

Moreno-Jimenez, J.M. and Vargas, L.G. (1993), "A probabilistic study of preference structures in the analytic hierarchy process with interval judgments", *Mathematical and Computer Modelling*, Vol. 17 Nos 4/5, pp. 73-81.

Mork, K.A. (1989), "Oil and the macroeconomy when prices go up and down: an extension of hamilton's results", *Journal of Political Economy*, Vol. 97 No. 3, pp. 740-744.

Muhammad, F., Hussin, M.Y.M. and Razak, A.A. (2012), "Automobile sales and macroeconomic variables: a pooled mean group analysis for asean countries", *IOSR Journal of Business and Management*, Vol. 2 No. 1, pp. 15-21.

Mulalic, I., Pilegaard, N. and and Rouwendal, J. (2016), "Does improving public transport decrease car ownership? Evidence from the copenhagen metropolitan area", Tinbergen Institute Discussion Paper 15-139/VIII. 10.2139/ssrn.2710547.

Nag, B. and De, D. (2020), "The Indian automobile industry: technology enablers preparing for the future", *In New Frontiers of the Automobile Industry*, Palgrave Macmillan, Cham, pp. 301-321.

Narayanan, V.G., Raman, A. and Singh, J. (2005), "Agency costs in a supply chain with demand uncertainty and price competition", *Management Science*, Vol. 51 No. 1, pp. 120-132.

Nauhria, Y., Pandey, S. and Kulkarni, M.S. (2011), "Competitive priorities for Indian car manufacturing industry (2011-2020) for global competitiveness", *Global Journal of Flexible Systems Management*, Vol. 12 Nos 3/4, pp. 9-20.

Niedermayer, D. (2008), "An introduction to bayesian networks and their contemporary applications", *In Innovations in Bayesian Networks*, Springer, Berlin, Heidelberg, pp. 117-130.

Nielsen, T.D. and Jensen, F.V. (2009), *Bayesian Networks and Decision Graphs*, Springer Science and Business Media, New York, NY.

Ott, R.L. and Andrus, D.M. (2000), "The effect of personal property taxes on consumer vehicle-purchasing decisions: a partitioned price/mental accounting theory analysis", *Public Finance Review*, Vol. 28 No. 2, pp. 134-152.

Pai, R.R., Kallepalli, V.R., Caudill, R.J. and Zhou, M. (2003), "Methods toward supply chain risk analysis", in *SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme-System Security and Assurance (Cat. No. 03CH37483)*, IEEE, Vol. 5, pp. 4560-4565.

Patil, N. and Warkhedkar, R. (2016), "Knowledge management implementation in Indian automobile ancillary industries: an interpretive structural model for productivity", *Journal of Modelling in Management*, Vol. 11 No. 3, pp. 802-810.

Patra, T. and Rao, M.J. (2017), "Impact of macroeconomic factors on automobile demand in India", *Journal of International Economics*, Vol. 8 No. 1, pp. 97-113.

Press Information Bureau (2019a), "Press information bureau, government of India, ministry of heavy industries and public enterprises, (2019 November 26), "slow down in automobile sector", available at: https://pib.gov.in/newsite/PrintRelease.aspx?relid=194999

Press Information Bureau (2019b), "Press information bureau, government of India, ministry of heavy industries and public enterprises, (2019 December 20), "year ender 2019 ministry of heavy industry", available at: https://pib.gov.in/newsite/PrintRelease.aspx?relid=196019

Pritchard, T. and DeBoer, L. (1995), "The effect of taxes and insurance costs on automobile registrations in the United States", *Public Finance Quarterly*, Vol. 23 No. 3, pp. 283-304.

Ramey, V.A. and Vine, D.J. (2004), "Tracking the source of the decline in GDP volatility: an analysis of the automobile industry", (No. w10384), National Bureau of Economic Research.

Rao, R.V. (2007), *Decision Making in the Manufacturing Environment: using Graph Theory and Fuzzy Multiple Attribute Decision Making Methods*, Springer Science and Business Media, London.

Rausand, M. (2013), *Risk Assessment: theory, Methods, and Applications*, John Wiley and Sons, NJ, Vol. 115.

Rawal, P. and Sisodia, D.R. (2017), "Growing acceptance of cab services in urban India", *International Journal in Management and Social Science*, Vol. 5 No. 6, pp. 270-280.

Rith, M., Fillone, A. and Biona, J.B.M. (2019), "The impact of socioeconomic characteristics and land use patterns on household vehicle ownership and energy consumption in an urban area with insufficient public transport service – a case study of metro manila", *Journal of Transport Geography*, Vol. 79, pp. 102484.

Saaty, T.L. (1980), *The Analytic Hierarchy Process*, Mcgraw Hill, New York, NY.

Saaty, T.L. and Vargas, L.G. (1987), "Uncertainty and rank order in the analytic hierarchy process", *European Journal of Operational Research*, Vol. 32 No. 1, pp. 107-117.

Salehi, M., Gah, A.D., Akbari, F. and Naghshbandi, N. (2020), "Does accounting details play an allocative role in predicting macroeconomic indicators? Evidence of Bayesian and classical econometrics in Iran", *International Journal of Organizational Analysis*, doi: 10.1108/IJOA-10-2019-1902.

Shahabuddin, S. (2009), "Forecasting automobile sales", *Management Research News*, Vol. 32 No. 7.

SIAM, Society of Indian Automobile Manufacturers (2020), "Domestic sales trends", available at: www.siamindia.com/statistics.aspx?mpgid=8&pgidtrail=14/ (assessed 10 January 2020).

Singh, R.K. and Sharma, M.K. (2014), "Selecting competitive supply chain using fuzzy AHP and extent analysis", *Journal of Industrial and Production Engineering*, Vol. 31 No. 8, pp. 524-538.

Singh, R.K., Garg, S.K. and Deshmukh, S.G. (2007), "Strategy development for competitiveness: a study on indian auto component sector", *International Journal of Productivity and Performance Management*, Vol. 56 No. 4, pp. 285-304.

Singh, R.K. (2013), "Prioritizing the factors for coordinated supply chain using analytic hierarchy process (AHP)", *Measuring Business Excellence*, Vol. 17 No. 1, pp. 80-98.

Singh, R.K., Garg, S.K. and Deshmukh, S.G. (2008), "Competency and performance analysis of indian SMEs and large organizations: an exploratory study", *Competitiveness Review*, Vol. 18 No. 4, pp. 308-321.

Tambade, H., Singh, R. and Modgil, S. (2019), "Identification and evaluation of determinants of competitiveness in the indian auto-component industry", *Benchmarking: An International Journal*, Vol. 26 No. 3, pp. 922-950.

Taylan, O. (2014), "IT project risk assessment of learning organizations by fuzzy set and systems", *International Journal of Organizational Analysis*, Vol. 22 No. 2, pp. 161-180.

Tipu, S.A.A. (2017), "Entrepreneurial risk taking: themes from the literature and pointers for future research", *International Journal of Organizational Analysis*, Vol. 25 No. 3, pp. 432-455.

Tong, R.M. and Bonissone, P.P. (1980), "A linguistic approach to decision making with fuzzy sets", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 10 No. 11, pp. 716-723.

Torfi, F., Farahani, R.Z. and Rezapour, S. (2010), "Fuzzy AHP to determine the relative weights of evaluation criteria and fuzzy TOPSIS to rank the alternatives", *Applied Soft Computing*, Vol. 10 No. 2, pp. 520-528.

Vatankhah, S., Bouzari, M. and Safavi, H.P. (2020), "Unraveling the fuzzy predictors of stress at work", *International Journal of Organizational Analysis*, doi: 10.1108/IJOA-01-2020-2005.

Wang, F.K., Chang, K.K. and Tzeng, C.W. (2011), "Using adaptive network-based fuzzy inference system to forecast automobile sales", *Expert Systems with Applications*, Vol. 38 No. 8, pp. 10587-10593.

Wetzel, J. and Hoffer, G. (1982), "Consumer demand for automobiles: a disaggregated market approach", *Journal of Consumer Research*, Vol. 9 No. 2, pp. 195-199.

Wu, D., Olson, D.L. and Dolgui, A. (2015), *Decision Making in Enterprise Risk Management: A Review and Introduction to Special Issue*, Omega, Elsevier Ltd.

Wu, J. (2005), "Quantity flexibility contracts under Bayesian updating", *Computers and Operations Research*, Vol. 32 No. 5, pp. 1267-1288.

Yadav, N.S. and Sagar, M. (2015), "Modeling strategic performance management of automobile manufacturing enterprises: an Indian context", *Journal of Modelling in Management*, Vol. 10 No. 2, pp. 198-225.

Yadav, V. and Sharma, M. (2016), "Multi-criteria supplier selection model using the analytic hierarchy process approach", *Journal of Modelling in Management*, Vol. 11 No. 1, pp. 326-354.

Zadeh, L.A. (1965), "Fuzzy sets", *Information and Control*, Vol. 8 No. 3, pp. 338-353.

Zahraee, S.M., Mamizadeh, F. and Vafaei, S.A. (2018), "Greening assessment of suppliers in automotive supply chain: an empirical survey of the automotive industry in Iran", *Global Journal of Flexible Systems Management*, Vol. 19 No. 3, pp. 225-238.

Zhang, Q., Zhan, H. and Yu, J. (2017), "Car sales analysis based on the application of big data", *Procedia Computer Science*, Vol. 107, pp. 436-441.

Further reading

Bhawan, V. and Marg, S.J.S. (2005), "Interpretive matrix: a tool to aid interpretation of management and social research", *Global Journal of Flexible Systems Management*, Vol. 6 No. 2, pp. 27-30.

Dargay, J. (2007), "The effect of prices and income on car travel in the UK", *Transportation Research Part A: Policy and Practice*, Vol. 41 No. 10, pp. 949-960.

Jagathy Raj, V.P. and Balakrishnan, M. (2012), "Dominant partial least square factors of consumer purchase behaviour of passenger cars", *Asian Journal of Management Research*, Vol. 3 No. 1, pp. 83-99.

Press Information Bureau (2017), "Press information bureau, government of India, ministry of heavy industries and public enterprises, (2017, august 2), "automobile sector's contribution to GDP", available at: https://pib.gov.in/newsite/PrintRelease.aspx?relid=169436

Shalender, K. and Yadav, R.K. (2019), "Strategic flexibility, manager personality, and firm performance: the case of Indian automobile industry", *Global Journal of Flexible Systems Management*, Vol. 20 No. 1, pp. 77-90.

The World Bank, Data (2020), "GDP growth rate (annual %), India", available at: https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?locations=IN (accessed 11 April 2020).

Appendix

Political risk	Economic risk	Social risk	Technological risk	Environmental risk
Tax policies (P_1) Environmental laws(P_2) Vehicle regulations (P_3)	Economic growth and development (E_1) Acquisition and running cost (E_2) Inflation rate (E_3) Availability of credit or finance (E_4) Employment rate (E_5)	Mass transit system (S_1) Cab ridership services (S_2) Vicinity of Roads and availability of parking (S_3)	Introduction of new vehicles (T_1) Automation in vehicles (T_2) Engine efficiencies (T_3)	Pollution levels (N_1) Season and climate (N_2)

Table A1. Classification of factors affecting the risk of automotive

Risk	Risk category	References
Economic growth and development	Economic	Muhammad *et al.* (2012), Shahabuddin (2009), Dargay (2001), Tambade *et al.* (2019)
Introduction of new vehicles	Technological	Nag and De (2020), Yadav and Sagar (2015)
Acquisition and running costs	Economic	Hamilton (1983), Mork (1989), Lee and Ni (2002), Barber *et al.* (1999), Duncan (1980), Chand and Avikal (2015)
Inflation rate	Economic	Chifurira *et al.* (2014)
Availability of credit or finance	Economic	Ludvigson (1998), Muhammad *et al.* (2012)
Tax policies	Political	Pritchard and DeBoer (1995), Alberini and Bareit (2016), Ott and Andrus (2000), Clibborn *et al.* (2016)
Employment rate	Economic	Muhammad *et al.* (2012), Shahabuddin (2009)
Mass transit systems	Social	Rith *et al.* (2019), Mulalic *et al.* (2016)
Cab ridership services	Social	Rawal and Sisodia (2017), Akbari *et al.* (2018)
Vicinity of roads and availability of parking	Social	Barbero and Galeota (2019)
Automation	Technological	Andrea and Smith (1994), Aghdaie and Yousefi (2011)
Engine efficiencies	Technological	Andrea and Smith (1994), Aghdaie and Yousefi (2011), Chand and Avikal (2015)
Pollution levels	Environmental	Barbero and Galeota (2019)
Season and climate	Environmental	Busse *et al.* (2015), Bloesch and Gourio (2015)

Table A2. Risk factors affecting product demand

Experts	Designation	Education qualification	Company where employed
Expert 1	Senior manager	MBA	Omax Auto
Expert 2	Manager	MBA	Maruti Suzuki
Expert 3	Assistant manager	MBA	Honda
Expert 4	Technical expert	PGDBA, M.Tech.	Mercedes

Table A3. Experts profile

Table A4. Level of factor importance according to decision makers

S. no.	Factors	Level of importance DM1	DM2	DM3	DM4
1.	Economic growth and Dev. (e.g. disposable income, etc), E_1	4	4	4	2
2.	Introduction of new vehicles with newer technology, T_1	5	4	4	3
3.	Acquisition and running costs, E_2	2	4	4	4
4.	Inflation rate, E_3	2	2	5	5
5.	Availability of credit/ finance, E_4	1	3	3	5
6.	Tax policies (taxes on purchase, incentives, etc), P_1	4	2	3	4
7.	Environmental laws (EVs, cleaner fuels, etc), P_2	1	2	2	4
8.	Vehicle regulations (Axle, safety, emission, etc), P_3	3	1	3	3
9.	Employment rate, E_5	2	1	1	4
10.	Public transport, S_1	2	1	3	2
11.	Cab services, S_2	3	4	4	3
12.	Vicinity of roads/ availability of parking spots, S_3	2	2	2	3
13.	Automation in vehicles, T_2	2	4	1	2
14.	Engine efficiencies, T_3	4	1	2	1
15.	Pollution levels, N_1	1	3	3	2
16.	Season and climate, N_2	1	2	3	1

Table A5.
Fuzzy integrated pairwise comparison matrix

Factors	E_1	T_1	E_2	E_3	E_4	P_1	P_2	P_3
E_1	(1, 1, 1)	(1, 1, 1)	(1,1.34,2)	(1, 1, 1)	(1,1.34,2)	(1,1.34,2)	(1, 2, 3)	(1,1.34,2)
T_1	(1, 1, 1)	(1, 1, 1)	(1,1.34,2)	(1, 1, 1)	(1,1.34,2)	(1,1.34,2)	(1, 2, 3)	(1,1.34,2)
E_2	(0.6,0.75,1)	(0.6,0.75,1)	(1, 1, 1)	(0.6,0.75,1)	(1, 1, 1)	(1, 1, 1)	(1,1.5,3)	(1, 1, 1)
E_3	(1, 1, 1)	(1, 1, 1)	(1,1.34,2)	(1, 1, 1)	(1,1.34,2)	(1,1.34,2)	(1, 2, 3)	(1,1.34,2)
E_4	(0.6,0.75,1)	(0.6,0.75,1)	(1, 1, 1)	(0.6,0.75,1)	(1, 1, 1)	(1, 1, 1)	(1,1.5,3)	(1, 1, 1)
P_1	(0.6,0.75,1)	(0.6,0.75,1)	(1, 1, 1)	(0.6,0.75,1)	(1, 1, 1)	(1, 1, 1)	(1,1.5,3)	(1, 1, 1)
P_2	(0.34,0.5,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.5,0.67,1)	(1, 1, 1)	(0.5,0.67,1)
P_3	(0.6,0.75,1)	(0.6,0.75,1)	(1, 1, 1)	(0.6,0.75,1)	(1, 1, 1)	(1, 1, 1)	(1,1.5,3)	(1, 1, 1)
E_5	(0.34,0.5,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.5,0.67,1)	(1, 1, 1)	(0.5,0.67,1)
S_1	(0.34,0.5,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.5,0.67,1)	(1, 1, 1)	(0.5,0.67,1)
S_2	(1, 1, 1)	(1, 1, 1)	(1,1.34,2)	(1, 1, 1)	(1,1.34,2)	(1,1.34,2)	(1, 2, 3)	(1,1.34,2)
S_3	(0.34,0.5,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.5,0.67,1)	(1, 1, 1)	(0.5,0.67,1)
T_2	(0.34,0.5,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.5,0.67,1)	(1, 1, 1)	(0.5,0.67,1)
T_3	(0.34,0.5,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.5,0.67,1)	(1, 1, 1)	(0.5,0.67,1)
N_1	(0.34,0.5,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.5,0.67,1)	(1, 1, 1)	(0.5,0.67,1)
N_2	(0.34,0.5,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.34,0.5,1)	(0.5,0.67,1)	(0.5,0.67,1)	(1, 1, 1)	(0.5,0.67,1)

(*continued*)

Table A5.

Factors	E_5	S_1	S_2	S_3	T_2	T_3	N_1	N_2
E_1	(1, 2, 3)	(1, 2, 3)	(1, 1, 1)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)
T_1	(1, 2, 3)	(1, 2, 3)	(1, 1, 1)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)
E_2	(1,1.5,3)	(1,1.5,3)	(0.6,0.75,1)	(1,1.5,3)	(1,1.5,3)	(1,1.5,3)	(1,1.5,3)	(1,1.5,3)
E_3	(1, 2, 3)	(1, 2, 3)	(1, 1, 1)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)
E_4	(1,1.5,3)	(1,1.5,3)	(0.6,0.75,1)	(1,1.5,3)	(1,1.5,3)	(1,1.5,3)	(1,1.5,3)	(1,1.5,3)
P_1	(1,1.5,3)	(1,1.5,3)	(0.6,0.75,1)	(1,1.5,3)	(1,1.5,3)	(1,1.5,3)	(1,1.5,3)	(1,1.5,3)
P_2	(1, 1, 1)	(1, 1, 1)	(0.34,0.5,1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)
P_3	(1,1.5,3)	(1,1.5,3)	(0.6,0.75,1)	(1,1.5,3)	(1,1.5,3)	(1,1.5,3)	(1,1.5,3)	(1,1.5,3)
E_5	(1, 1, 1)	(1, 1, 1)	(0.34,0.5,1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)
S_1	(1, 1, 1)	(1, 1, 1)	(0.34,0.5,1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)
S_2	(1, 2, 3)	(1, 2, 3)	(1, 1, 1)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)	(1, 2, 3)
S_3	(1, 1, 1)	(1, 1, 1)	(0.34,0.5,1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)
T_2	(1, 1, 1)	(1, 1, 1)	(0.34,0.5,1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)
T_3	(1, 1, 1)	(1, 1, 1)	(0.34,0.5,1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)
N_1	(1, 1, 1)	(1, 1, 1)	(0.34,0.5,1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)
N_2	(1, 1, 1)	(1, 1, 1)	(0.34,0.5,1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)

Factors	Fuzzy geometric mean ($\tilde{r}$)	Fuzzy weights ($\tilde{w}_i$)	Weights (w_i)	Normalized weights(w_{ni})
E_1	(1.000, 1.500, 2.059)	(0.04317, 0.08823, 0.16268)	0.0980266	0.0870986
T_1	(1.000, 1.500, 2.059)	(0.04317, 0.08823, 0.16268)	0.0980266	0.0870986
E_2	(0.880, 1.230, 1.732)	(0.03798, 0.07235, 0.13685)	0.0823933	0.0732081
E_3	(1.000, 1.500, 2.059)	(0.04317, 0.08823, 0.16268)	0.0980266	0.0870986
E_4	(0.880, 1.230, 1.732)	(0.03798, 0.07235, 0.13685)	0.0823933	0.0732081
P_1	(0.880, 1.230, 1.732)	(0.03798, 0.07235, 0.13685)	0.0823933	0.0732081
P_2	(0.642, 0.760, 1.000)	(0.02771, 0.04470, 0.07901)	0.0504733	0.0448465
P_3	(0.880, 1.230, 1.732)	(0.03798, 0.07235, 0.13685)	0.0823933	0.0732081
E_5	(0.642, 0.760, 1.000)	(0.02771, 0.04470, 0.07901)	0.0504733	0.0448465
S_1	(0.642, 0.760, 1.000)	(0.02771, 0.04470, 0.07901)	0.0504733	0.0448465
S_2	(1.000, 1.500, 2.059)	(0.04317, 0.08823, 0.16268)	0.0980266	0.0870986
S_3	(0.642, 0.760, 1.000)	(0.02771, 0.04470, 0.07901)	0.0504733	0.0448465
T_2	(0.642, 0.760, 1.000)	(0.02771, 0.04470, 0.07901)	0.0504733	0.0448465
T_3	(0.642, 0.760, 1.000)	(0.02771, 0.04470, 0.07901)	0.0504733	0.0448465
N_1	(0.642, 0.760, 1.000)	(0.02771, 0.04470, 0.07901)	0.0504733	0.0448465
N_2	(0.642, 0.760, 1.000)	(0.02771, 0.04470, 0.07901)	0.0504733	0.0448465
Σ	(12.656, 17.000, 23.164)		1.125466	1.000000

Table A6. Calculated values of normalized weights

Risk categories	Risk factors	Factor probability	Risk
Political risk	Tax policies	0.0732081	0.06375423
	Environmental laws	0.0448465	
	Vehicle regulations	0.0732081	
Economic risk	Economic growth and development	0.0870986	0.07309198
	Acquisition and running cost	0.0732081	
	Inflation rate	0.0870986	
	Availability of credit or finance	0.0732081	
	Employment rate	0.0448465	
Social risk	Mass transit system	0.0448465	0.05893053
	Cab ridership services	0.0870986	
	Vicinity of roads and availability of parking	0.0448465	
Technological risk	Introduction of new vehicles	0.0870986	0.05893053
	Automation in vehicles	0.0448465	
	Engine efficiencies	0.0448465	
Environmental risk	Pollution levels	0.0448465	0.04484650
	Season and climate	0.0448465	

Table A7. Risk for each of the risk category

About the authors

Dr Girish Kumar is Associate Professor in Mechanical Engineering Department at Delhi Technological University, Delhi, India. He has more than 22 years of experience in industry, teaching and research. His teaching and research areas include Reliability Engineering, Maintenance Management, Quality engineering, Stochastic Modeling, etc. He holds a PhD degree in Reliability Engineering from Indian Institute of Technology Delhi, India. Dr Kumar has published 25 papers in International Journals and conference proceedings.

Dr Rajesh Kr. Singh is a Professor at the Management Development Institute, Gurgaon, India. His areas of interest include supply chain management, Industry 4.0, quality management and circular economy. He has about 170 research papers published in reputed international/national journals and

conferences. He has published papers in journals such as *Resources, Conservation and Recycling, Journal of Cleaner Production, Industrial Management and Data Systems, Production planning and Control, Singapore Management Review, International Journal of Productivity and Performance Management, Benchmarking: An International Journal, Journal of Modelling in Management, Competitiveness Review: An International Business Journal, Global Journal of Flexible Systems and Management, International Journal of Productivity and Quality Management, International Journal of Services and Operations Management, International Journal of Automotive Industry and Management, International Journal of Logistic Systems Management, IIMB Management Review and Productivity Promotion.* Rajesh Kr. Singh is the corresponding author and can be contacted at: rksdce@yahoo.com

Mr Rishabh Jain is an undergraduate student at Delhi Technological University, (formerly known as Delhi College of Engineering), New Delhi, India. He is pursuing his Bachelor in Technology with Mechanical Engineering as major. His areas of interest include- decision-making, risk-modelling and other concepts of industrial engineering management.

Mr Raman Kain is a final year student at Delhi Technological University, New Delhi, India. He is pursuing his B.Tech with Mechanical Engineering as major. His areas of interest include- creative problem solving, risk analysis and concepts of quality management.

Mr Naveen is a final year student at Delhi Technological University, New Delhi, India. He is pursuing his Bachelor in Technology with Mechanical Engineering as major. His specialization is automotive engineering.

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/336668692

Analysis of Total Quality Management (TQM) Barriers on its Performance using TOC

Article · August 2019

DOI: 10.22214/ijraset.2019.8097

CITATIONS
0

READS
32

1 author:



Rakesh Kumar
Delhi Technological University
7 PUBLICATIONS **0** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Performance Measurement View project

Population Statistics View project

All content following this page was uploaded by Rakesh Kumar on 12 July 2020.

The user has requested enhancement of the downloaded file.

iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** VIII **Month of publication:** August 2019

DOI: http://doi.org/10.22214/ijraset.2019.8097

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VIII, Aug 2019- Available at www.ijraset.com

Analysis of Total Quality Management (TQM) Barriers on its Performance using TOC

Rakesh Kumar[1], Prof. R.S. Mishra[2]
[1]*Ph.D Scholar,* [2]*Professor, Department of Mechanical Engineering, Delhi Technological University, Delhi, India-110042*

Abstract: ***This study examines and analyzes the TQM factors which act as barrier and affects its performance. TQM view the organization as a collection of processes that must be continuously improved through utilization of the knowledge and experience of associates in all functions and at all levels. The industries must be more focused on understanding their own structure in terms of production/ service processes. Theory of Constraints (TOC) focuses on the weakest ring(s) in the chain to improve the performance of process systems. In this situation, TOC becomes an important problem structuring and solving methodology which changes the way of thinking of decision makers. This paper includes the extensive literature review on Theory of Constraints and suggested that how the same approach to analyze the performance of the Total Quality Management system.***

Keywords: Performance measurement, Theory of constraints (TOC), TQM.

I. INTRODUCTION

In today scenario of the industrial increasing competitiveness and cutthroat competition to capture wide market has led to the introduction of unconventional manufacturing management strategies such as just-in-time (JIT), material resources planning (MRP), Six-sigma quality control (SSQC), Lean Manufacturing and the theory of constraints (TOC). The basic concept of Theory of Constraints, shortly known as TOC assumes that the performance of the organization or say organizational system is adversely affected by constraints present in the system. Within the organizational system process, various functions are performed that are part of aiding the increase within production. The TOC then develops an explicit approach to manage these constraints to maximize throughput with continuous improvement. That's why it was also called that TOC is a multi-faceted systems methodology that has been progressively developed to assist people and organizations to think about problems, develop breakthrough solutions and implement those solutions successfully. The TOC defines a constraint as, *"Anything that prevents the organisation from achieving higher performance versus its goal"*. So, in the case of a profit-making organisation this centers on what limits. The TOC claim is that there are few constraints to any system preventing it from achieving its goal. Although TOC originally focused on physical resource constraints the underlying constraints are commonly policy, or deeper paradigm constraints. TOC uses cognitive mapping to verify this assumption where necessary, but there are many other authors who acknowledge the importance of underlying core problems that constrain system improvement. The main objective of this study is to identify the critical factors which behaves like bottleneck is required for TQM implementation in the organization, and these factors act as the critical chain. TOC is a complex methodology requiring skill and cooperation to implement. The past literature about the link between TQM practices and organization performance gives contradictory results. So the purpose of this study is to develop a conceptual framework and a research model showing the relationship between TQM implementation and performance measures of organization. Most of the previous works show that TQM has significant relationship with firm's performance. However, the examining of moderators is less given in previous work, which mediators are known generally as general tools and techniques without specific focus on types of improvement. And also to determine the effects of moderators TPM, SPC, Lean such as which improve the relationship between TQM practices and firm's performance.

II. LITERATURE REVIEW

In past TOC has developed rapidly in terms of both methodology and area of applications. Dettmer (1995) in his literature compared the system approach taken by TQM and TOC. He argued that TQM views the system in terms of discrete processes and then optimizes the quality in each process, whereas TOC improves performance by concentrating on the weakest link in the total system. It is apparent, however, that without subsequent identification of the weakest links, and application of the philosophy to these groups or processes within the organization, major improvements cannot be achieved. Dettmer further described the CRT technique of TOC is as functional rather than organizational and as such is blind to internal and external system boundaries.

©IJRASET: All Rights are Reserved

The TOC is providing a method of focusing improvement on the critical areas, and thus has the potential for much faster improvement towards a more global goal. The TOC can also contribute by emphasizing methods of dealing with change. The main obstacle for broader implementation of TQM in many industrial organizations, the need to win over those who do not believe in the method and who find it difficult to wait for TQM payoffs when they are battling daily problems.

The TOC provides some answers to these problems by contributing the Socratic thinking process for dealing with change, and potentially creating ownership and commitment throughout the organization. The focus on the weakest links helps also in hastening the improvement and generating more apparent results at a quicker pace. To address the policy constraints and effectively implement the process of on-going improvement, Goldratt (1990, 1994) develop a generic approach called the "thinking process" (TP).

Literature concerning TOC in different manufacturing system, like, JIT and TOC buffering philosophies are compared and suggested that improved system performance stems from the strategic placement of buffers in DBR(Drum-Buffer-Rope), which maximizes protection of the constraint from variation rather than attempting to protect each individual station is given by Watson, K.J.et.al (2008). Spencer, M.S. et.al (1995) also supported through his paper that DBR technique of TOC can be used to improve current production operations.

Gupta, S., (1997), Perez, J.L. (1997) and Watson, K.J. et.al (2003) advocated for the same technique of TOC for managing the supply chains systems. Boyd and Gupta (2004) established TOC as a theory by identifying the underlying construct "throughput orientation" along with its three dimensions: Mindset, Measures and Methodology. The thinking process tools of TOC provide techniques for identifying and solving root problems – which is a stated aim of TQM practices. Cox et al. (2003) provided some real-life examples of the applications of the TOC thinking process tools to quality problems.

A. TOC Method to Address TQM

A number of authors and quality experts consider that TQM principles and practices are universally applicable to any organization (Deming, 1982) regardless of the organisational contextual factors (e.g. size, technology, culture or external environment) (Zhao et al., 2004). Despite the claimed benefits of TQM, the relevant literature points to many research studies that indicate a high rate of failures in the process of implementing TQM practices, due to barriers that hinder its implementation (Sila, 2007; Soltani and Wilkinson, 2010).

TQM, Six-sigma, Lean Manufacturing, TOC all these strategic tools aiming similar manner, but with distinct flavor. These quality improvement tools first create its own virtual system in the existing system where it has to be implemented with active participation of the employee. In our further analysis of the relationship between TQM and TOC we will use the following definition of TQM (Dahlgaard et al., (1998): TQM is a company culture characterized by increased customer satisfaction through continuous improvements, in which all employees actively participate.

Researchers exploited these strategic tools far away from its actual usages. Like Six-sigma was initially developed by Motorola in later 80's for the product centric defect reduction and quality improvement tool, later it was experimented and successfully implemented in various areas.

Schonberger (2008) points out that the objective of Six Sigma programs is to create a higher perceived value of the company's products and services in the eyes of the customer. Six Sigma is a well-established approach that seeks to identify and eliminate defects, mistakes or failures in business processes or systems by focusing on those process performance characteristics that are of critical importance to customers (Snee, 2004).

Ehie and Sheu (2005) were able to integrate the theory of constraints with the DMAIC concept for achieving continuous quality improvement. Similarly, Lean manufacturing is designed for the process improvement as well as reduction/elimination of the existing waste, and further it also applied in the field which were substantial.

TQM is a wide concept since it embraces the whole organization and its processes instead of focusing on the product. It is also considered by many to be a holistic approach, which seeks to convert the culture and structure of the organization into a total commitment to quality (Barad, 1996). TQM is generally considered to be based on a number of core values such as customer focus, decisions based on facts, process orientation, continuous improvement, everybody's commitment and leadership; see Hellsten and Klefsjo (2000).

©IJRASET: All Rights are Reserved

Table I: Linkage of Barriers to TQM

Factors	Barriers
Support Quality	• Lack of Leadership commitment • High turnover of management level • Company work culture is not synergetic with company goal and objectives • Difficult to change employee mindset about quality • Lack of information about quality
Support Internal Communication	Lack of coordination amongst departments
Provide quality resources	Lack of sufficient funds to mobilize TQM driven activities
Use competent personal	• Poor education levels of workers • Low morale (absenteeism, industrial action etc.) • Lack of skill of workers • High worker turnover • Indiscipline (non-conformances with procedures)
Support Competence	• Insufficient Training about quality • Management related training is not achieving organizational training targets
Provide quality Infrastructure	• Poor condition of machine • Ineffective maintenance programs • Poor condition of equipment spare part procurement
Control Purchasing Process	• Raw material does not confirm with the specification • Spontaneous delivery of raw materials • Difficulty in procurement of raw material

The core idea of TOC is that every system has at least one constraint that prevents from achieving the goal to a larger degree. Constraints can be physical resources or policies. TOC develops a set of procedures and methodologies to identify and optimize such constraints. In this paper TQM is considered as a system under the organizational system and the barriers are considered as the constraints, which hinders the quality goal. And the well-known fact is that the success of the system depends on everybody's participation (one of the core principles of TQM).

The Theory of Constraints is based on five steps.

1) Identify the system's constraints(s)
2) Decide how to exploit the system's constraint(s)
3) Subordinate everything else to the above decision
4) Elevate the system's constraint(s)
5) If, in the previous steps, a constraint has been broken, go back to step 1, and do not allow inertia to cause a system's constraint.

The system's constraint is that part of the system that constrains the objective or goal of the system. The issues with TOC is that the constraints must be kept operating at its full capacity, if not then entire process slows further. The organization where TQM is implemented, and the barriers/hurdles/constraints even if identified must be kept operating to maintain the pace of the production. Upstream operations must provide only what the constraint can handle and downstream operations will only receive what the constraint can put out. Total Quality Management System is system of any organizational systems. It is known that System of systems is a collection of task-oriented or dedicated systems that pool their resources and capabilities together to create a new, more complex system which offers more functionality and performance than simply the sum of the constituent systems.

The failures in implementing TQM are typically attributed to organisational contingent, situational or contextual factors such as unsupportive organisational culture, resistance to change, lack of resources, or uncommitted leadership (Zhao et al., 2004; Sadikoqlu and Zehir, 2008). It is also said about TQM that it is system that starts with consumers and ends with them. So successful implementation of TQM improves organizational performance and leads to high degree of satisfaction to customers and employees.

©IJRASET: All Rights are Reserved

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VIII, Aug 2019- Available at www.ijraset.com

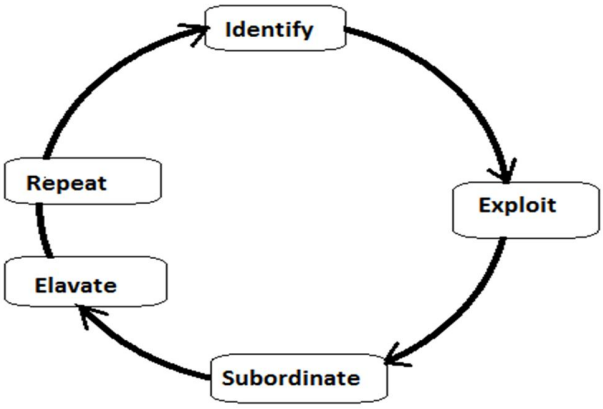


Figure1. Five aiming steps of the TOC

TQM systems include the procedures for quality planning strategy and operations, for setting capital and operating budgets, for measuring and rewarding performance, and for reporting progress and conducting meetings. many companies introduced total quality management as a new management system. But while TQM enabled firms to focus more effectively on process improvements, the ability to implement strategy across organizational units remained elusive. Companies' management systems were still tactical and operational, not strategic. In terms of the TQM system's constraint is the existing barriers. The whole performance of TQM system depends on successful run of the TQM drivers and enablers and the TQM driver depends on the TQM capability. Consequently, it is essential to keeping in view the quality goals. In Step 1 this needs to be identified. By identifying these barriers, company management can increase their understanding of implementation failure and the subsequent removal of these barriers can lead to the possibility of successful implementation. Thus, any company that was rated as high performing was evaluated as moving toward successful implementation of TQM practices. Activities to remove barriers for consensus encompasses a wide range of activities including leadership, decision-making tools, and other tools or tactics to restructure culture and/or to re-motivate employees to the TQM paradigm. As it is a system, then the maximum possible utilization needs to be achieved. This may mean to eliminate those factors which were identified as obstacle for the accomplish the quality improvement procedure as soon as possible or reducing the number of barriers. It will mean ensuring that there is always analysis of the system to do. Thus the system's constraint is exploited (Step 2). If this is the constraint, then there is no point running the enablers: so every other TQM tools needs to be subordinated to remove the bottleneck if still present (Step 3). To achieve the goal/ objective, the system's constraint may need to be `eliminated' or minimized. The enablers, the capacity of the system, has been increased, or `elevated' (Step 4). The application of Step 4 may have changed the system's constraint. With its increased drivers, the original bottleneck may no longer be constraining the system, so the new bottleneck needs to be identified, and the process repeated (Step 5). Thus, this is a process of continual improvement.

B. TOC Strategies

The core premise of TOC is that any organization or system's performance is limited most by a "constraint" — the thing that most blocks throughput. Identifying the nature and location of the system's constraint is the key to continuous system improvement, stated by Srikanth et al. (1997). TOC is a data-driven approach like other continuous improvement processes uses Identify the system's constraints, decide how to exploit the constraint(s), Subordinate everything else to the above decisions, Elevate the constraint(s) and Don't let inertia become the new constraint - go back to step 1, but do not allow previous decisions made in steps 2 to 4 to become constraints because TOC views constraints as positive, not negative. Because constraints determine the performance of a system, a gradual elevation of the system's constraints will improve its performance. According to Dettmer (1995), he compared the system approach taken by TQM and TOC. He argued that TQM views the system in terms of discrete processes and then optimizes the quality in each process, whereas TOC improves performance by concentrating on the weakest link in the total system. Whether organization manages stand-alone or multiple projects, whether those projects are small or large, whether customers are internal or external, or whether the nature of the work performed is product development, construction, design, IT, or service; most projects are difficult to manage because of two things:

©IJRASET: All Rights are Reserved

1) They involve uncertainty, and
2) They involve three different and opposing commitments: Due date, budget, and content

The main strategies of TOC along with tasks and its descriptions summarized in Table II.

Table II

TASK	Description
Identifying "What to Change?"	Organization must effectively address the underlying root causes that lead to the problems.
Identifying "To What to Change To?"	To address these root causes and coping mechanisms, a comprehensive solution must be provided. The solution includes: 1) a robust planning process, 2) a more effective scheduling process, 3) a methodology for introducing work that actually leads to increased capacity, 4) execution processes that provide excellent project control, visibility and decision support, and 5) work behaviors that are more conducive to good project performance.
Identifying "How to cause the change?"	The right people must be brought into the picture at the right time, in just the right way. The process must move slowly enough to permit identification of essential changes, yet fast enough that it does not lose the momentum that is necessary to sustain continued progress.

III. CONCLUSION AND FUTURE RESEARCH

The performance of the Total Quality Management (TQM) System is the responsibility of every person involved in all activities related to the company. As in this work TQM considered as a management system for a customer-focused organization that involves all employees in continual improvement. This paper combines the powerful and practical management theories – theory of constraints over TQM. When applied this concept enable managers to analyse and control the most important factors in their firm – the constraints – in new ways.

TOC may be also applied for the lean manufacturing, six-sigma, Business process reengineering (BPR), supply chain management (SCM) and logistic system by considering them as a system inside the system. Cellular manufacturing system and group technology is also one of the untouched field where TOC strategic management tool may be implemented in the similar fashion for the better result. All the above mentioned the defined process or system functions as intended and produces intended results under normal operating conditions. So, as a future work, the barriers may be ranked and weighted to get the precise result is recommended. Choosing the right TOC tool for the desired solution is also the one of the most challenging job always.

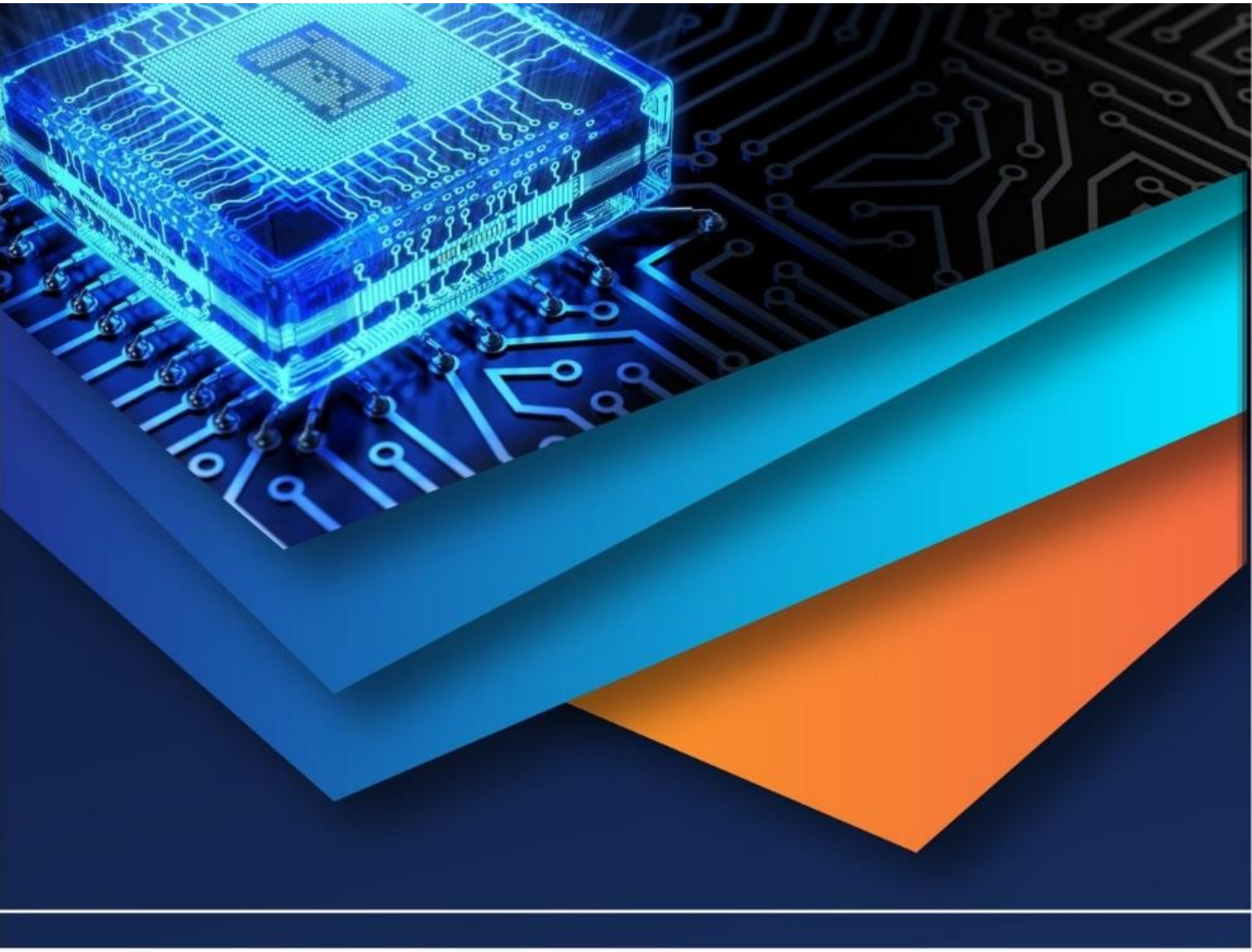
REFERENCES

[1] Boyd, L.H., Gupta, M.C. (2004), "Constraints management: what is the theory?", International Journal of Operations & Production Management, Vol. 24 No. 4, pp. 350-371.
[2] BARAD M,. (1996) Total quality management, International Encyclopedia of Business and Management, 5.
[3] Baladhandayutham, R., Devadasan, S. R., Selladurai, V., & Senthil, V. (2001). Integration of BPR and TQM: Past, present and future trends. Production Planning Control: The Management of Operations, 12(7), 680–688.
[4] Catalin, S. H., Bogdan, B., & Dimitrie, G. R. (2014). The existing barriers in implementing total quality management. Annals of the University of Oradea, Economic Science Series, 23(1), 1234–1240.
[5] Cox, J.F., Blackstone, J.H. and Schleier, J.G. (2003), Managing Operations: A Focus on Excellence, North River Press, Great Barrington, MA.
[6] C. Carl Pegels, Craig Watrous, (2005) "Application of the theory of constraints to a bottleneck operation in a manufacturing plant", Journal of Manufacturing Technology Management, Vol. 16 Issue: 3, pp.302-311,
[7] Chakravorty, S.S., Atwater, J.B., (2006) Bottleneck management: theory and practice. Production & Planning Control 17 (5), 441–447.
[8] Dahlgaard, J.J., Kristensen, K. and Kanji, G.K. (1998), Fundamentals of TQM, Carfax, London.
[9] Deming, W.E., (1982) "Quality Productivity and Competitive Position" Massachusetts Institute of Technology, Cambridge.
[10] Dettmer, H.W., (1995) "Quality and the theory of constraints", Quality Progress, Vol. 28 No. 4, pp. 77-81.
[11] Ehie, I. and Sheu, C. (2005), "Integrating Six-Sigma and theory of constraints for continuous improvement: a case study", Journal of Manufacturing Technology Management, Vol. 16 No. 5, pp. 542-53.

©IJRASET: All Rights are Reserved

[12] Goldratt, E.M., (1990), "What is This Thing Called Theory of Constraints and How Should it be Implemented?", North River Press, New York, NY.
[13] Goldratt, E.M., (1994), "It's Not Luck", Gower, England.
[14] Gupta, S., (1997) "Supply Chain Management in Complex Manufacturing", IIE Solutions, vol. 29(3), 18-21.
[15] Hellstenu & Klefsjob,. (2000) TQh.1 as a management system consisting of values, techniques and tools, The TQM Magazine, 12, No. 4, pp. 238-244.
[16] Hakes, C. (1991), Total Quality Management: The Key to Business Improvement, Chapman and Hall, London, UK.
[17] Ho, S.K.M. and Fung, C.K.H. (1994), "Developing a TQM excellence model", The TQM Magazine, Vol. 6 No. 6, pp. 24-30.
[18] Issac, G., Rajendran, C. and Anantharamanan, R.N. (2004), "A conceptual framework for total quality management in software organizations", Total Quality Management, Vol. 15 No. 3, pp. 307-44.
[19] Jun, M., Shaohan, C., & Peterson, R. (2004). Obstacles to TQM implementation in Mexico's Maquiladora industry. Total Quality Management & Business Excellence, 15(1), 59–72.
[20] Koh, S. C. L., Demirbag, M., Bayraktar, E., Tatoglu, E., & Zaim, S. (2007). The impact of supply chain management practices on performance of SMEs. Industrial Management & Data Systems, 107(1), 103–124.
[21] Mabin, V.J. and Balderstone, S.J., (2000) The World of the Theory of Constraints: A Review of the International Literature, (CRC Press LLC: Boca Raton, FL).
[22] Moorthy, M. K., Tan, A., Choo, C., Wei, C. S., Tan, J., Ping, Y., & Leong, T. K. (2012). A Study on Factors Affecting the Performance of SMEs in Malaysia. Academic Research in Business and Social Sciences, 2(4), 224–239.
[23] Meredith, J., Mantel, S., 2003. Project Management: A Managerial Approach. Wiley, New York, NY.
[24] Perez, J.L. (1997), "TOC for World Class Global Supply Chain Management", Computers and Industrial Engineering, vol 33(1-2), 289-293.
[25] Nair A (2006). Meta-analysis of the relationship between quality management practices and firm performance – implications for quality management theory development. Journal. Operations. Management. 24: 948-975.
[26] Pyzdek, T., (2003), The Six Sigma Handbook: A Complete Guide for Green Belts, Black Belts, and Managers at All Levels. McGraw-Hill, New York, NY.
[27] Rahman, S. (1998), "Theory of constraints: a review of the philosophy and its applications", International Journal of Operations & Production Management, Vol. 18 No. 4, pp. 336-56.
[28] Snee, R.D. (2004), "Six Sigma: the evolution of 100 years of business improvement methodology", International Journal of Six Sigma and Competitive Advantage, Vol. 1 No. 1, pp. 4-20.
[29] Sila, I., (2007) "Examining the effects of contextual factors on TQM and performance through the lens of organizational theories: an empirical study", J. Oper. Manag. 25, 83–109.
[30] Salenga, G. and Fazel, F, (2000), "Obstacles to implementing TQM", Quality Progress, Vol. 33 No. 7, Milwaukee, July,pp 53-64
[31] Salegna, G., & Fazel, F. (2000). Obstacles to implementing quality. Quality Progress, 33(7), 53–57.
[32] Schonberger, R.J., (2008). Best Practices in Lean and Six Sigma Process Improvement. John Wiley & Sons, Hobokon, NJ.
[33] Soltani, E., Wilkinson, A., (2010) "Stuck in the middle with you: the effects of incongruency of senior and middle managers' orientations on TQM programmes". Int. J.Oper.Prod.Manag, Vol. 30,365–397.
[34] Srikanth, M.L., and Umble, M.M., (1997). Synchronous Management: Profit Based Manufacturing for the 21st Century Volume 1, The Spectrum Publishing Company: Guilford.
[35] Salaheldin, S. I. (2009). Critical success factors for TQM implementation and their impact on performance of SMEs. International Journal of Productivity and Performance Management, 58(3), 215–237.
[36] Spencer, M.S. & Cox, III, J.F., (1995) "Optimum Production Technology (OPT) and the Theory of Constraints (TOC): Analysis and Genealogy", International Journal of Production Research, vol 33(6), 1495- 1504.
[37] Sadikoqlu, E., Zehir, C., (2008) "The impact of contextual factors on total quality management practices". J.Glob.Strateg.Manag.4,88–97.
[38] Watson, K.J. & Polito, T., (2003) "Comparison of DRP and TOC Financial Performance within a Multi-Product, Multi-Echelon Physical Distribution Environment", International Journal of Production Research, vol 41(4), 741-765.
[39] Watson, K.J. & Patti, A., (2008) "A comparison of JIT and TOC buffering philosophies on system performance with unplanned M/C downtime", International Journal of Production Research, vol. 46(7), 1869- 1885.
[40] Zhao, X., Yeung, A.C.L., Lee,T.S.,(2004) "Quality management and organisational context in selected service industries of China.", J. Oper. Manag. 22,575–587.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)

View publication stats

Anomalous Object Tracking in Distributed Camera Network

MD Shahbaz Khan, Dr Indu Sreedevi
Delhi Technological University
New Delhi, India
mdshahbaz_bt2k16@dtu.ac.in, s.indu@dtu.ac.in

***Abstract*—The paper introduces a novel framework for real-time tracking of an object with higher precision in a Pan-Tilt-Zoom (PTZ) camera network. In the above-mentioned framework, the object which behaves anomalously is picked for tracking. An SVM classifier has been used to pick anomaly behaving object among all the objects present in a frame. It is a self-initializing algorithm as it does not require user intervention for object detection. Target detection is followed by its autonomous tracking in the distributed camera network. Hence it does not make the system bulky and takes less time to execute. The paper implements multiple object tracking using Discriminative Correlation Filtering with Channel and Spatial Reliability(CSR-DCF) tracking algorithm. Existing works are based on Particle filtering or Kalman filtering algorithms which are computationally complex and not self-initializing.**

I. INTRODUCTION

An efficient and fast algorithm is required for tracking a specified object from a frame having many objects using distributed PTZ camera based network. For each frame, some objects are at a higher priority for tracking than the rest. Although it is possible that none of the objects needs to be tracked in a particular sequence of frames. In such cases, tracking is not initialized till an anomaly behaving object is detected. Act of running, bending or picking was considered as anomalous behaviour. For detecting an anomaly behaving object, SVM based classifying model is trained. The distance of centroid of an object from its contour was used as a feature for developing the SVM model. This leads to the enhancement of the computational efficiency of the proposed algorithm. Sink frame of a particular camera in the network serves as the source frame for the other. For the mentioned decision-making task, the paper proposes a real-time, computationally efficient algorithm in which the camera having sink frame generates a histogram and is transmitted to all the adjacent cameras. Other cameras generate a histogram for their respective source frame and compare it with the received one. Camera having the same histogram starts tracking the object. Use of histogram formation and their comparison results in a smooth tracking at the time of object handover from one field of view (FOV) to another. Real-time and computationally fast algorithms are required for surveillance of densely populated public places using distributed PTZ cameras. The proposed approach assumes overlapping FOV between adjacent cameras in the network with the same resolution of the sink and the source frame of the neighbouring cameras.

II. LITERATURE LAYOUT

Several algorithms have come up to make object tracking autonomous with higher precision and real-time. However, on increasing the number of cameras in the network, their computational efficiency decreases manifold. Algorithms like [1],[2] do not consider the formation of the shadow of objects while few other algorithms like [3] are not suited for public places where the number of objects is more than an ordinary scenario. Current works in this field [3] attempts to track the object using particle filtering algorithm along with user intervention. As far as places with larger crowds are concerned, particle filtering algorithms are not efficient and do not yield real-time processing of data. CSR-DCF based tracking is a multi-object tracking algorithm and hence it can handle a large number of objects in crowd places. Paper [2] deals in Kalman filtering based object tracking, which is not efficient for real-time object tracking. In the case of Kalman filtering, noise is measured by modelling it as additive white Gaussian noise (AWGN). This makes the noise measurement in each frame as a joint Gaussian distribution. This, in turn, makes the object tracking vague when there is a bifurcation in the path of the object. This paper implements CSR-DCF tracking algorithm which is one of the best available real-time object tracking algorithm as discussed in table II. In papers like [1], object tracking is initialized by detecting the object in the very first frame by using the user interface. This, in turn, binds the algorithm from being fully autonomous. To eliminate this bottleneck, our approach proposes the concept of self-initialization of the object. Self initialization can be defined as the detection of an object of interest by the system itself. This paper implements self-initialization by choosing the object of interest as per its anomalous behaviours among all the objects present in that frame. The anomaly behaving object is picked using the SVM model. The model has been trained by using the distance of centroid of the object from its contour. No sooner did the object is detected, then its tracking gets started in all the cameras present in that distributed camera network. This algorithm proves to be efficient for the crowded region as it can track the movement of a suspicious object by itself. Each camera in the network tracks the object and communicates with adjacent cameras when the object of interest leaves its FOV. This novel method of object handover eliminates the chance of object missing during tracking in a

multi-camera system and hence yields a higher accurate tracking.

III. Problem Statement

The paper aims to find the solution for the detection of anomaly behaving object followed by its tracking in a distributed camera network. The intent is to improve the tracking accuracy and precision in comparison with probability and Gaussian distribution-based tracking algorithms[3] and [2] like particle filtering and Kalman filtering. The proposed methodology also eliminates random id generation technique[3] used to track objects in multiple cameras. Moreover, the proposed algorithm of object handover reduces the probability of object missing during its tracking in a multi-camera network system.

IV. Anomalous Behaviour Detection

The prerequisite for this algorithm is the minimum number of objects to be present in the camera frame which is achieved by classifying objects as normal and abnormal behaving groups. Activities like running, bending, or the act of picking are classified as anomalous behaviour in this paper as shown in figure 5 & 8. The paper proposes Principal Component Analysis (PCA) based feature extraction followed by its classification as anomalous behaviour using Support Vector Machine (SVM) classifier as discussed in [4] and the result has been collected in the table I. To prepare an SVM model, the maximum distance of any point on the contour from the centroid of the object was calculated and was used as features for SVM classier.

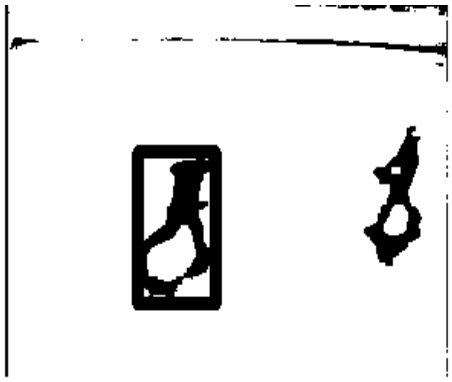


Fig. 1. A person running classified as anomalous behaviour

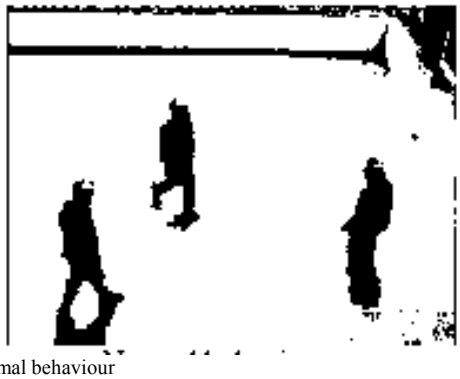


Fig. 2. Normal behaviour



Fig. 3. A person bending in the pathway – classified as anomalous behaviour.

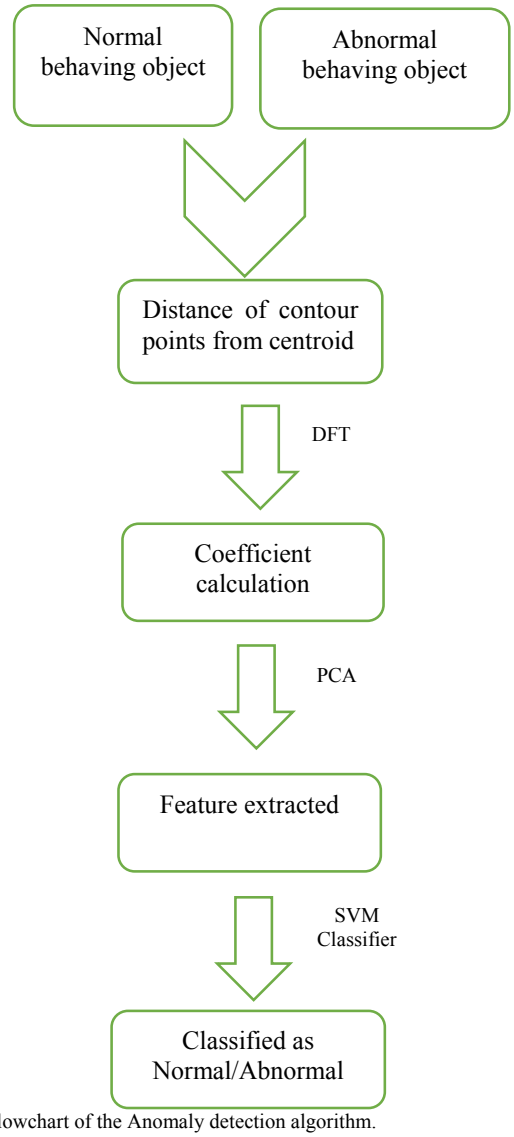


Fig. 4. Flowchart of the Anomaly detection algorithm.

First of all, the contour of every object was formed. Then using the below-mentioned formula, the centroid of that object was obtained.

$$x_c = \frac{1}{N}\sum_1^N x_i \quad (1)$$

$$y_c = \frac{1}{N}\sum_1^N y_i \quad (2)$$

where x_c, y_c are the coordinates of the centroid of contour.

For each point on contour, Euclidean distance was calculated and was transformed to coefficients by using DFT (Discrete Fourier Transform), which was later used for feature extraction using PCA (Principal Component Analysis).

A. PCA Based Feature Extraction

To extract features from each frame of camera input, PCA algorithm was implemented. The training dataset was divided into two groups namely G1, G2 each having 'n' training samples. Eigenvector of features was generated as follows: -

1) For each group of the dataset, a matrix '*X*' was formed having 'n' rows of samples and 'c' columns of features.

2) From the feature matrix '*X*', covariance matrix 'C' was computed as

$$C = (X - X')^T (X - X') \quad (3)$$

3) For the computed covariance matrix, eigenvector and eigenvalues were determined using the following relation

$$CV = V\Lambda \quad (4)$$

Where V is eigenvector and Λ is an eigenvalue.
Out of 'c' columns of Λ, we picked 'k' number of columns thus reducing the feature vector size to n x k.

B. SVM Classifier

As per the statistical learning theory [5],[6] & [7], SVM is one of the most precise classifying algorithms. The paper implements SVM for classifying the extracted features into normal and abnormal behaviour by using separating hyperplane. The constructed hyperplane was localized by maximizing the separating margins and consequently, it led to the formation of nonlinear decision boundary in the training dataset.
By implementing SVM algorithm, an optimal hyperplane was found such that it divides the dataset into two groups completely. In the course of building the SVM model, a hyperplane is adjusted by maximizing the margin between data values and hyperplane. Margin maximizes by minimizing the cost function.
Where cost function $J(x, y, f(x))$ is defined as,

$$J(x, y, f(x)) = \begin{cases} 0, \text{if } y * f(x) \geq 1 \\ 1 - y * f(x), \text{else} \end{cases} \quad (5)$$

On building the SVM model for training data, it was tested on the testing dataset whose labels was known. The hyperplane was further adjusted to get the maximum accuracy of classification. At the end of testing, the SVM model was used for unknown data.

Anomaly Behaviour Detection Algorithm

Prerequisites: Continuous video dataset, the dataset must contain anomaly behaviour, the dataset should contain video of the different timing of day having varying brightness.

Behaviour Detection:
1. Dividing the dataset into training and testing data group in a ratio of 1:4.
2. Extracting features from training dataset.
3. Optimizing number of features required to detect behaviour using PCA.
4. Adjusting the location of hyperplane by minimizing the cost function using SVM.
5. Applying the trained model to the testing dataset and evaluating the precision percentage.
6. If the precision is not more than 85% (as depicted in table I), then the number of data example in training dataset was increased and step 6 was repeated.
7. Else proceed to apply the trained model to an unknown dataset.

V. CSR-DCF ALGORITHM

In this paper, we have applied background subtraction to our input data before detecting anomaly behaviour. For the abovementioned task, background segmentation algorithm (based on Gaussian Mixture) was used, as it flexible for diurnal changes and for shadow detection as well. This algorithm is based on two research papers [8],[9] given in the year 2004 and 2006 respectively. On detecting an abnormal behaving object in a frame, the tracking algorithm starts executing. A rectangle bounding box is formed around that region of interest with a different colour for different objects in each frame. CSR-DCF algorithm was implemented for tracking objects which is a real-time processing algorithm with higher precision.
As discussed in [10], CSR-DCF is a tracking algorithm which provides learning with higher efficiency and better integration of filter updates and tracking. In this algorithm, spatial reliability map is used which helps to adjust filter support to enhance the accuracy of tracking. Similarly, reliability scores are calculated to check the quality of filters and thus helps in evaluating coefficients in localization.
This paper proposes a novel framework of object tracking based on discriminative correlation filtering followed by channel and spatial reliability scores[11]. It is a proximity-based algorithm which learns a set of filters to represent the object and uses them to find the highest correlated point in the adjacency of tracking an object from the previous frame.

A. Spatially Correlation Filters

With the help of the following approach, we comprehend the importance of learning correlation filters. For a given set of 'n' channels with features f = $\{f_k\}_{k=1:n}$ and target templates h = $\{h_k\}_{k=1:n}$, and the position of the object is determined by locating the highest frequency of ĝ(h),

$$\hat{g}(h) = \sum_{k=1}^{n} f_k \star h_k \quad (6)$$

where $\star$ is circular convolution operator & $\hat{g}(h)$ is correlation response.

Similarly, to find optimum correlation filter h, we have to minimize

$$\mathcal{E}(h) = \|\hat{g}(h) - g\|^2 + \lambda\|h\|^2 \quad (7)$$

where g is the expected output.

To find the efficiency of tracking, we transformed it into Frequency Domain as

$$\mathcal{E}(h) = \| \sum_{k=1}^{n} diag(\hat{f}_k)\hat{h}'_k - \hat{g}' \|^2 + \lambda\|h\|^2 \quad (8)$$

And on minimizing the above equation we get,

$$\hat{h}_k = (diag(\hat{f}_k)\hat{g}') \odot^{-1} (\sum_{k=1}^{n} diag(\hat{f}_k)\hat{f}'_k + \lambda) \quad (9)$$

where $\odot^{-1}$ is element-wise division operator and the above result can be used to design correlation filter trackers.

On considering each channel to be independent, we get the cost function as

$$\mathcal{E}(h) = \sum_{k=1}^{n} \|f_k \star h_k - g\|^2 + \lambda\|h\|^2 \quad (10)$$

To scale, we introduce a factor depending upon the discriminative power of feature channels which is termed as channel weights $w = \{\hat{w}_k\}_{k=1:n}$ and whose correlation response is given by,

$$\hat{g}(h) = \sum_{k=1}^{n} f_k \star h_k \cdot \hat{w}_k \quad (11)$$

B. Correlation Filter Learning

As mentioned above, it has been assumed that the correlation filter is independent of channels, hence it allows to consider the case of one filter only. The constraints are formalized as following

$$h_c \; m \odot h \equiv 0 \quad (12)$$

where h_c is defined as the dual variable with the above constraints.

On using Lagrangian augmentation [9], we find the following outcome

$$s\mathcal{L}(\hat{h}_c, h, \hat{I}|m) = \|diag(\hat{f}_k)\hat{h}'_k - \hat{g}\|^2 + \lambda\|h_m\|^2 + [\hat{I}^H(\hat{h}_c - \hat{h}_m) + \hat{I}^H(\hat{h}_c - \hat{h}_m)'] + \mu\|\hat{h}_c - \hat{h}_m\|^2 \quad (13)$$

Here in the above equation, we have introduced a new term μ which is the constraint penalty value and its value is updated by the following equation.

$$\mu^{i+1} = \beta\mu^I \quad (14)$$

C. Spatial Reliability Map

On locating the target, the region of interest is extracted from the frame and is used for updating the filter. The constrained filter uses spatial reliability map m_p, which is the collection of pixels of training model matching with that of the testing mode. To determine m_p, we have designed the following algorithm.

While the algorithm is getting executed for tracking, f oregr ound, and background of the object is stored in colour histogram c. Let $v_i = [y_i^c, y_i^x]$ is the observation i.e., it denotes the colour of y_i^c along with the y_i^x position for each pixel and m_p = random variable indicating the unknown foreground/background. For y_i to be observed on output, its joint probability was found as follows

$$\begin{aligned} p(y_i) &= \sum_{j=0}^{1} p(y_i | m_i = j)p(m_i = j) \\ &= \sum_{j=0}^{1} p(y_i^c | m_i = j)p(y_i^x | m_i = j)p(m_i = j) \end{aligned} \quad (15)$$

Algorithm 1 CSR-DCF

Prerequisites: Image I_i, the position of an object in previous frame pos_{i-1}, scale k_{i-1}, filter h_{i-1}, image histograms c_{i-1}, channel reliability w_{i-1}.

Determination and approximation of scale:-

1. Locating target at pos_i : finding the highest frequency in a correlation between filter (h_{i-1}) and extracted features (f) at pos_{i-1} position and adjusted by reliability scores (w).
2. Calculation of $\hat{w}$(det) (detection validation) with the help of per-channel responses.
3. Scale k_i was calculated, with the help of pos_i.
4. Colour histograms for foreground as well as the background were calculated and updated using $\hat{w}$(det) (detection validation).
5. Reliability map m_p was calculated using the above-mentioned relation.
6. Filter $\hat{h}$ was calculated and updated using m_p.
7. Using h, learning channel reliability was determined.
8. Calculate channel reliability from learning channel reliability & detection reliability.
9. Each filter hi was updated using the relation $h_i = (1-\eta)h_{i-1} + \eta\hat{h}$.
10. Channel reliability w_i was updated using the expression $w_i = (1-\eta)w_{i-1} + \eta\hat{w}$.

VI. Object Handover

In a distributed camera network, each camera is capable of processing the data and extracting information out of it. Cameras in a network can communicate with each other wirelessly in a range of 100m. The exchange of data between two cameras was done using OFDM modulation (Orthogonal

Frequency-Division Multiplexing). The advantage of using OFDM modulation is that it provides versatility in the presence of multiple channels.

The problem arises in handing over the object from one camera to the other. To eliminate the object handover problem, this paper implements the following algorithm:-

OBJECT HANDOVER ALGORITHM

Case 1: If the system has not yet detected any anomaly behaviour.
- i. In this case, each camera will keep on processing the input frames for anomaly detection.
- ii. If any anomaly behaviour is detected, then that camera will switch to case 2 else will continue to execute case1.

Case 2: if any anomaly behaviour is detected.
- i. Object tracking algorithm will be applied to the camera input in which anomaly action has been detected.
- ii. Camera tracking the anomaly behaving object will stop executing the anomaly detection algorithm for new objects.
- iii. As soon as the object becomes the sink for a camera, the camera generates its histogram from the last frame (say h1).

Case 3: when the object moves out of FOV of one camera
- i. In this case, the anomaly detection algorithm is paused and for each camera histogram of the new entering object is calculated (say h2).
- ii. Then both colour histograms (h1 and h2) are compared and the camera (say C2) whose histogram matches with h1, is now used for tracking.
- iii. The object tracking algorithm is applied to the camera (C2) input frame and case 2 is followed.

*Figures 6,7 and 9,10 illustrate the above-mentioned algorithm.

VII. EXPERIMENTAL RESULTS

The above experiment was performed on a dataset collected from Computer Vision Laboratory CVLAB ('https://cvlab.epfl.ch/data/data-pom-index-php/'). The mentioned dataset consists of inputs of 4 cameras located at different worldly points but with the same intrinsic parameters. On implementing the above-proposed algorithm, real-time object detection was done using anomaly behaviour detection technique with the following results

TABLE I. RESULT OF SVM CLASSIFIER

RESULT OF SVM CLASSIFIER		
Behaviour	***Success rate***	***No. of testing samples***
Normal	97%	297
Anaomaly	86%	328

The above-proposed algorithm was tested on the dataset at VOT2016 [12] benchmark which is comprised of 14 videos (7 training and 7 testing). On rigorous testing and comparison with various tracking algorithms, we got the following result (Table II).

TABLE II. COMPARISONS OF TRACKING ALGORITHMS

Tracking Algorithm	EAO	Aav	Fav	Total
Proposed method	0.34	0.510	0.851	-0.001
CCO Tracking	0.33	0.520	0.856	-0.006
TCN Net.	0.33	0.521	0.961	-0.110
SSA Tracking	0.32	0.531	1.044	-0.193
MLD Filter Based	0.31	0.482	0.833	-0.041
Staple	0.30	0.512	1.351	-0.539
DOC	0.29	0.513	1.233	-0.430
EB Tracking	0.29	0.441	0.902	-0.171
SRB Tracking	0.29	0.504	1.251	-0.547
STAPLEp	0.29	0.533	1.321	-0.498

*EAO – Expected Average Overlap
*A_{av} – Average Accuracy
*F_{av} – Average Failure
*Total = EAO + A_{av} - F_{av}

The above-mentioned comparison of various tracking algorithms was carried on the ground of expected average overlap, average accuracy and average failure. For an ideal tracking algorithm expected average overlap and average accuracy should be higher while lower average failure is preferred. In the case of real tracking algorithms, we need to compromise between these three parameters. To bring uniformity in comparison, a total usefulness function formed whose formula has been mentioned above. The higher total usefulness value for a tracking algorithm is, better that algorithm is to use. From the above table, we conclude that our tracking algorithm is better among the rest as its total usefulness value is higher than rest tracking algorithms.



Fig. 5. A person bending classified as anomalous behaviour by the camera C1.





Fig. 6. C1 detected object being tracked by the camera C2.





Fig. 7. Tracking object moving out of C2 camera being tracked by the camera C3



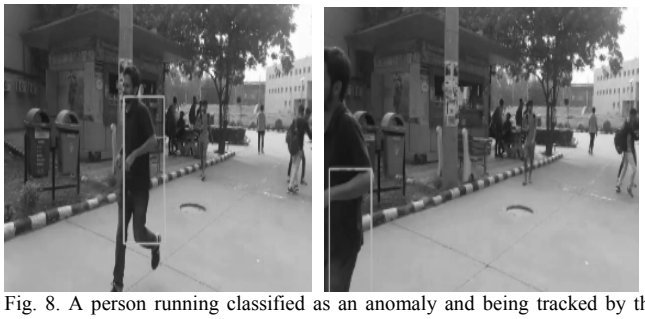


Fig. 8. A person running classified as an anomaly and being tracked by the camera C1.

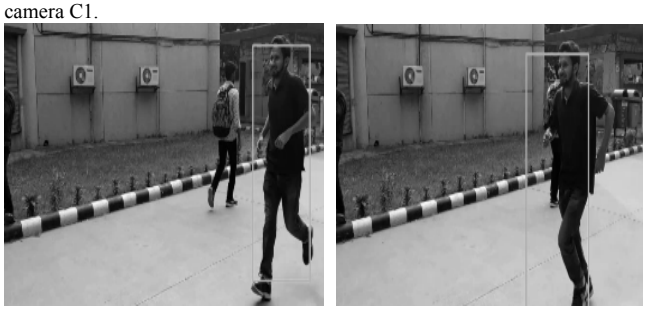




Fig. 9. C1 detected object being tracked by the camera C2





Fig. 10. Tracking object being moved out of C2 and tracked by camera C3

*Fig. 6-10 depicts the results of packing anomaly behaviour (running, bending and picking) using 3 cameras C1,C2 and C3.

VIII. Conclusion and future scope

This novel framework proposes an accurate and precise method for object tracking in distributed cameras. It is a real-time based approach with higher computational efficiency for tracking in a camera network. One of the important feature of this algorithm is self-initialization, which do not require user interventions. The proposed algorithm deals with the tracking of anomaly behaving objects using CSR-DCF based algorithm. Running, bending or picking has been classified as anomaly behaviour and based on this classification, objects are picked for tracking. For tracking the object, CSR-DCF algorithm has been implemented which is faster and more precise. Histogram comparison-based algorithm has been implemented to handover the object from one camera to the other in the network system.

Acknowledgement

I express my deepest gratitude to all those who helped me make this research complete and meaningful in all sense. First of all, I would like to thank our faculty in-charge, Prof. S. Indu & Dr Ayesha Chowdhary, whose contribution to this research work is incredible. They helped me to get through this research work by providing their valuable suggestions and sparing their precious time. Furthermore, I acknowledge the contribution of other staffs of the ECE department (DTU), who helped me by providing their valuable suggestions related to my topic. Last but not least, I am highly thankful to my parent and friends, who provided me moral support in times of distress and failure and played a vital role to make this work complete in all sense.

References

[1] G. R. Bradski, “Real time face & object tracking as a component of a perceptual user interface,” IEEE workshop on Application of Computer Vision, Princeton, NJ USA, 1998.

[2] S. Weng, C. Kuo and S. Tu, “Video Object Tracking using Adaptive Kalman Filter”, *Journal of Visual Communication & Image Representation*, 17(2006), 1190–1208.

[3] A. Choudhary, S. Sharma, I. Sreedevi and S. Choudhary, “Real-Time Distributed Multi-object Tracking in a PTZ Camera Network”, *Springer*, vol. 9127, 2015, pp. 183-192.

[4] X. Wu, Y.Ou, H. Qian, Y. Xu, “A detection system for Human Abnormal Behavior” RSJ International Conference on Intelligent Robots and Systems, Canada, 2005.

[5] K. Lee and Y. Xu, “Modeling human actions from learning” RSJ International Conference on Intelligent Robots and Systems, Japan, 2004.

[6] B. Scholkopf and A.J Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. Massachusetts: MIT Press, 2002.

[7] Y. Ou and Y. Xu, “Learning human control strategy for dynamically stable robots: Support Vector Machine approach” International Conference on Robotics and Automation, Taiwan, 2003.

[8] Z. Zivkovic, “Improved Adaptive Gaussian Mixture Model for Background Subtraction” ICPR, UK, 2004.

[9] Z. Zivkovic and F. Hejiden, “Efficient adaptive density estimation per image pixel for the task of background subtraction”, *Pattern Recognition Letters*, vol. 27, No. 7, 2006, pp. 773-780.

[10] A. Lukezic, T. Vojir, L. Zajc, J. Matasa and M. Kristan, “Discriminative Correlation Filter with Channel and Spatial Reliability”, *Internaltional Journal of Computer Vision*, 126(2017), 6309–6318.

[11] M. Kristan, R. Pflugfelder, A. Leonardis, J. Matas, F. Porikli, L. Cehovi, G., G. Fernandez and T. Vojir “The Visual Object Tracking, The Visual Object Tracking Challenge Results”, 2013, pp. 98-111.

[12] M. Kristan, et al, “The Visual Object Tracking VOT2016 Challenge Results”, Proc. European Conference Computer Vision, 2016.

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Arduino GSM based Power Theft Detection and Energy Metering System

Vivek Kumar Jaiswal
Department Of Electrical Engineering
Delhi Technological University
vivekjais01@gmail.com

Hritik Kumar Singh
Department Of Electrical Engineering
Delhi Technological University
singhhritik.singh5@gmail.com

Kuldeep Singh
Assistant Professor
Delhi Technological University
kuldeepsingh@dtu.ac.in

Abstract: Electrical energy has been considered as an essential form of energy. Whether it is Industrial need or household need, electricity is a must. Nowadays, with the increasing need for electricity, power theft is also increasing. This power theft is a severe problem like load scheduling will just become a hoax because of this Theft. Also, in developing countries like India, this power theft is 30%- 40% of total generated power and the electricity board has to provide this from their funds this results in major losses to electricity boards. This paper presents a way to detect power theft, alert the consumer, and cut off the supply accordingly. When unauthorized actions are detected, and a separate message is sent back to the microcontroller to remove the illegal supply, an SMS is sent automatically to the user via the GSM module. The GSM is interspersed with a unique system energy meter with a relay to handle the non-technical losses, problems with billing, and variations in voltage fluctuations

Keyword **– GSM, Atmega328, LM2596, Arduino GSM**

Introduction

Over the last few decades, communications technology has improved through connectivity and boundaries. The use of the "Embedded Program in Communication" provides various interesting applications. Electricity is now more than necessary. The demand for electricity is increasing continuously. With increasing electricity demand, power theft is also on the rise. Power theft is a concerning issue even in the most efficient power systems like in USA and a moderately efficient system like in Malaysia [1]. It has become important to improve the system performance and to avoid Theft. The purpose of this project is to design a system to monitor any power consumed before the power meter, and that is power pilferage. If some power theft is detected, that results in penalty or disconnection of energy supply to the respective consumer, and simultaneously an SMS is sent to the control unit. In the past, numerous non-technical and scientific methods for detecting electricity pilfering have been suggested. Non-technical procedures may involve customer inspection with a suspect load profile. Although regular inspections can help reduce theft, such action calls for the massive labour force and enormous effort. In most situations, this attempt always fails due to employee dishonesty. “The use of the main observer meter at the secondary distribution terminals are some of the technological ways to detect pilferage [3], harmonic generator [6], genetic support vector machines [7], extreme learning machine, power line impedance technique [8] [3].

However, these technical approaches can be effectively implemented only if proper communication is ensured between the central control station and the appropriate test points. Recently, energy meters based on GSM network have been proposed. These meters provide the energy metering and remote load control system facility [3]. In this paper, an energy metering scheme based on GSM is introduced to deal with various aspects of electricity theft. The proposed system avoids meter abuse and bypassing irregularities.

Existing Technology

The meter is manipulated to display low readability, bypass meter, and billing irregularities. Advanced distribution automation and dynamic price models relying on automated metering (AMR) and advanced measuring infrastructure (AMI) can overcome the billing irregularities and requirements for more employees [3]. However, this system has failed to benefit the atmosphere for energy theft and requires frequent inspection by NTL. Electrical cables, coaxial cables, optical fiber, infrared, Bluetooth are also used for current communication technologies but do not have reliability, communication speed, high installation costs, and maintenance are expensive and risky.

Methodology

For successful completion of this project, the following steps were carried out in a sequential manner

- Literature Reviewing: This includes exploring various documents and books regarding this topic.
- Problem identification: identifying the problem and its consequences.
- System modelling: Combining different elements to describe the system.
- System analysis and simulation: This indicates how the system behaves using proteus.
- Proposing Solution: designing of power theft control system to use power wisely.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:55:26 UTC from IEEE Xplore. Restrictions apply.

Proposed System

In this paper, the energy stealing system is proposed to send the SMS using a GSM Network. The most common methods of electricity theft are to bypass the phase line, cut off the neutral line, bypass the entire meter, and distort the energy meter. The current transformer is blinded to zero current, bypassing the phase line. The voltage transformer shows zero potential by disconnecting the neutral line and giving the local ground. The entire meter is bypassed, and the energy meter is measured at zero. Faulty readings are triggered by tampering the energy meter with unauthorized modification. The main contributions of this paper as follows

- The proposed system uses only a very small number of components and a simple electricity theft algorithm.
- The customer is warned initially after several warnings, the connection of the customer is being cut, and data is simultaneously sent to the control center
- Compared to other techniques discussed, total system costs are relatively low.
- It can be integrated with existing meters and new meters of production processes.

Many named and unnamed manners for theft of electricity are also available. The use of a single household meter cannot detect and control these types of theft. An observation meter is used in the proposed system to identify such unrecounted theft. Figure 1 shows the observer meter. Measure the overall household energy consumption at agreed time intervals. For example, outside an apartment building, this type of meter can be installed. In this case, all the apartments in this apartment building measure the energy consumed. The household meters communicate their measured units via SMS to the central meter. The central meter detects the theft of electricity when both values are significantly different. Since the observer meter can be read individually, households in which theft occurred can be calculated based on zero or low power consumption. The central meter informs the server of potential theft locations via SMS. The authority may send technical personnel to check the households reported and detect the dishonest consumer [3].

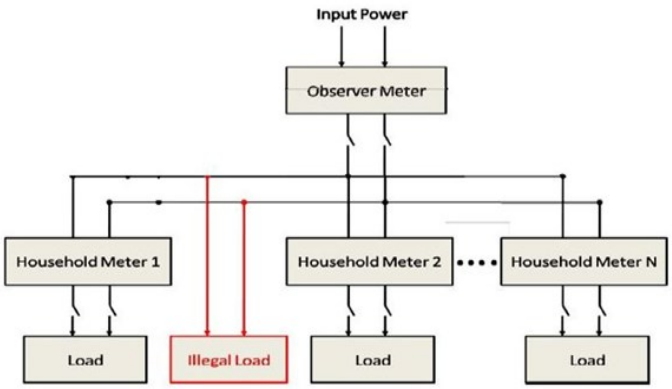


Figure 1: Block Diagram

Block Diagram

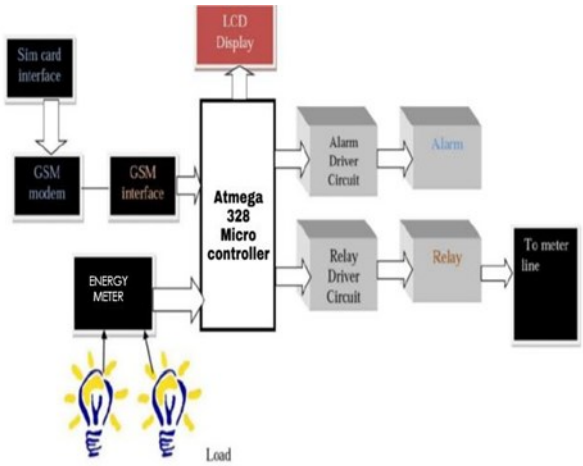


Figure 2: Proposed Architecture

The aims of the proposed program are mainly to monitor electricity theft through enhanced GSM communication features. Figure 2 shows the block diagram of the prototype. The most important components included power meter, GSM module [10], LM2596 voltage regulator, Arduino UNO [9], PC817 optocoupler, and LCD. The prototype model is a device that uses energy to link a variable load. The PC817 optocoupler and the CAL (calibration) energy meter LED is attached for the detection of consumed units. At 3200 imp/kWh the pulse rate of the power meter is. The LED blink rate is, therefore, directly proportionate to the load. To record consumed units, the output of the optocoupler is forwarded to the microcontroller.

Proposed Circuit

In the circuit first, a power supply block i.e. IC LM2596, which is made by TEXAS INSTRUMENT is introduced. This is used to get a constant 5v DC output, that is used to power 16*2 LCD, Microcontroller, Relay Module.

Data pins of the LCD display are connected to digital pins of the microcontroller (pin-6,7,8 & 9) which is used as output pins. GSM module is also interfaced with the microcontroller through digital pins 2 and 3(in the software part this is defined as transmitter and receiver pins). A buzzer is also connected to pin 11 of the microcontroller which functions in case of theft. Also, a relay is used to trip the load in case of theft, this is connected to pin 10 of the microcontroller.

Two energy meters are used one is named as master energy meter and other as a house energy meter. These meters are connected to pins A0 and A1(Analog input pins) of microcontroller through PC817(optocoupler). Optocoupler's input terminals are connected to calibration LED of Energy meter and output is connected to pins A0 & A1

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:55:26 UTC from IEEE Xplore. Restrictions apply.

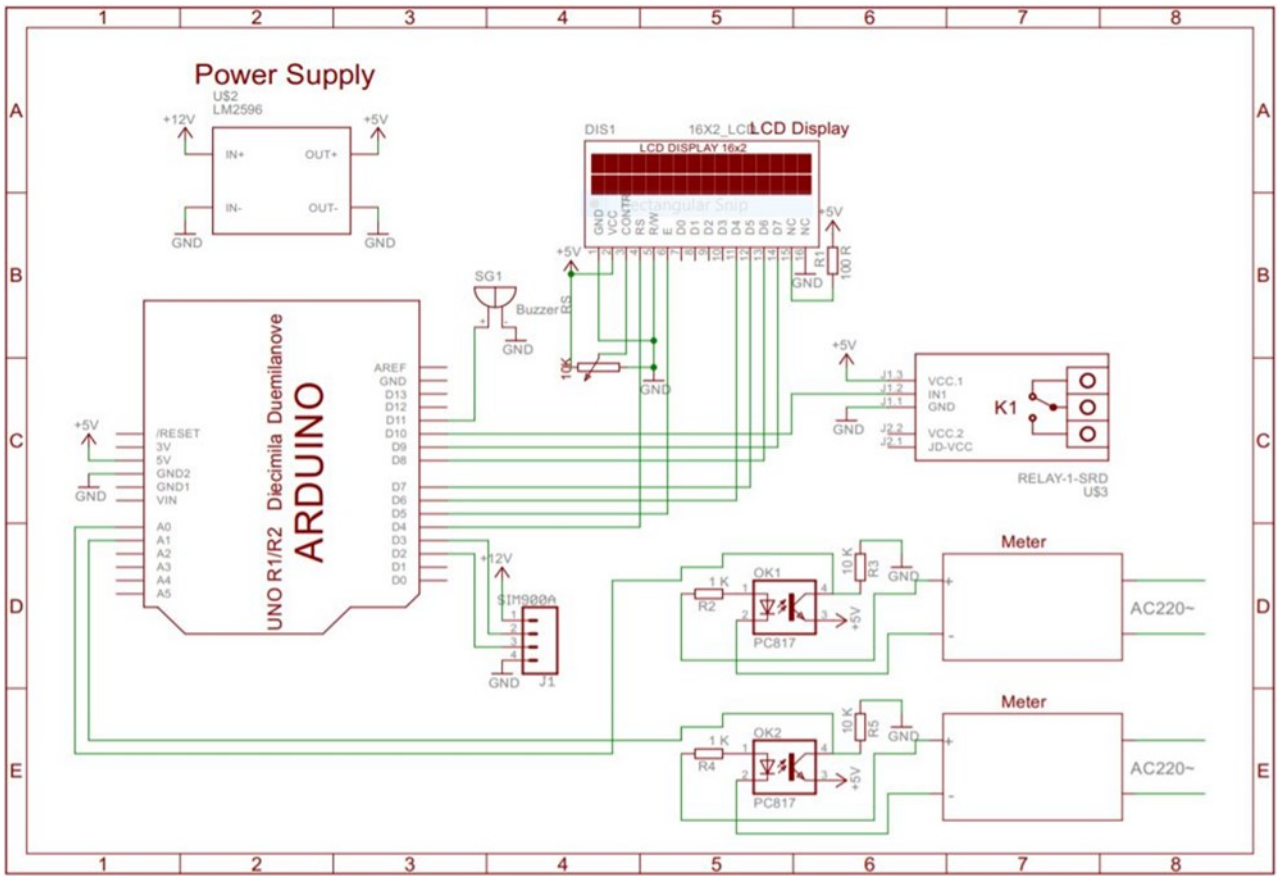


Figure 3: Circuit diagram

Hardware Design

The 230V AC power supply is given to energy meter and then connected to the load through the relay. The input of 5V for Arduino is given by a voltage regulator LM2596 from the 12V supply. The output pin of optocoupler PC817 is given to Arduino through A0 and A1. The interfacing of GSM and Arduino is made. The input 12V GSM is given through a 220 to 12V AC to DC adaptor. The corresponding message is sent through GSM.

Working

The power supply from the distributor is given to the system. This supply consists of 230 volts. The transmission lines are used to provide power supply to the consumer using electric poles at a particular distance. The system has an Arduino Atmega328, energy meter, and GSM module attached to it. Here the power supply is step down from the 230-volt power supply into 5 volts by the power supply circuit because our system uses a 5-volt power supply. The 230-volt power supply is stepped down 12 volts via a 12-volt step-down transformer so now the power supply will be 12 volts. It uses an LM2596 voltage regulator to get a 5-volt power supply which is given to our system. This 5-volt power supply given to an Atmega328 and GSM module. The system has Arduino, Energy meter, and GSM module. The master energy meter tracks the quantity of the energy units supply through it then it sends the unit consumed to Atmega328. Now Atmega328 stores this data in memory. As the supply reaches to slave energy meter it will count the number of energy units consumed. The data at the slave energy meter will be transfer to Atmega328. The Atmega328 will analyze the two data present. If the number of energy units reached varies. The Atmega328 will trigger the GSM module. The GSM module sends the SMS to the consumer about the amount of difference occurred in the transmission cable. The distributor will track the amount of loss happen and puts a penalty to the consumer. The distributer is informed about the energy loss happen to them by the system's action.

Programming

The code for the program is done in Arduino IDE and then it is burned into the Arduino. For the simulation part, Keil software is used to convert the program into .hex file. The programming logic followed is depicted in the flow chart.

978-1-7281-5371-1/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:55:26 UTC from IEEE Xplore. Restrictions apply.

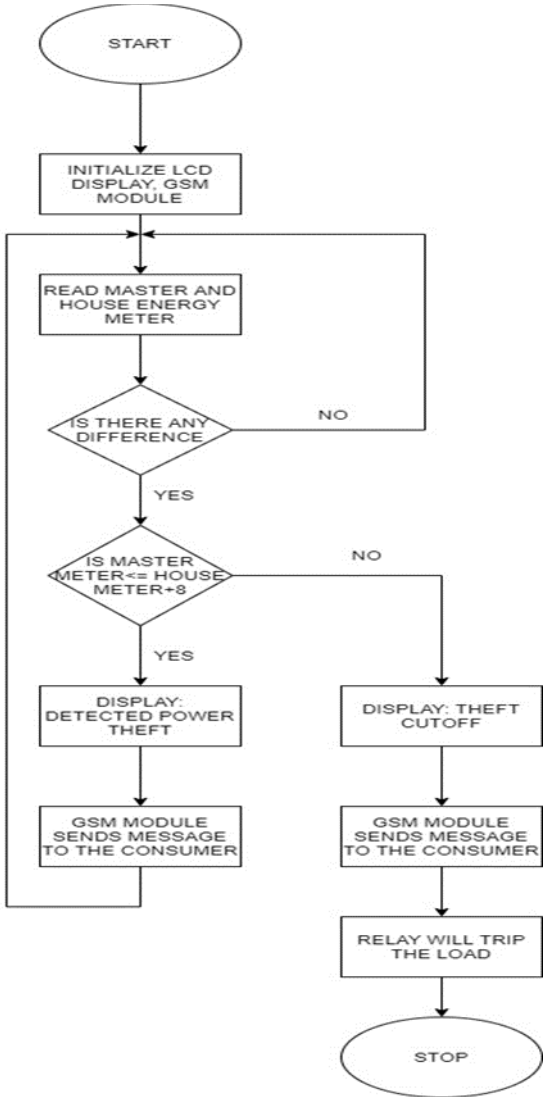


Figure 4 -programming flow chart

Simulation Results

Normal condition-Both the meter shows the same readings

NO ALERT as no theft observed.

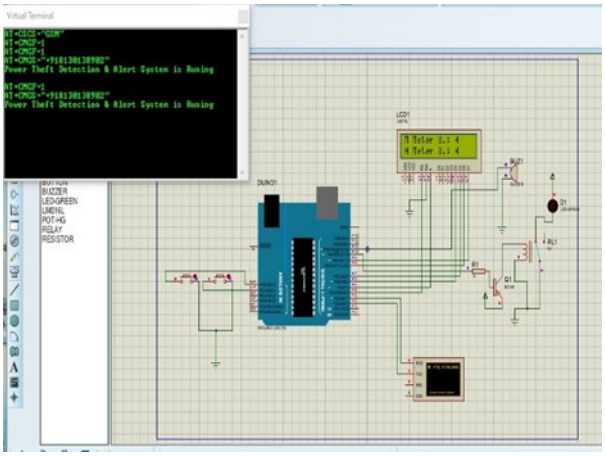


Figure 5-Normal condition

Alert stage: - When theft is observed. master meter reading and house meter readings have a difference of 3 units. alert message is sent in the message screen that power theft has been observed.

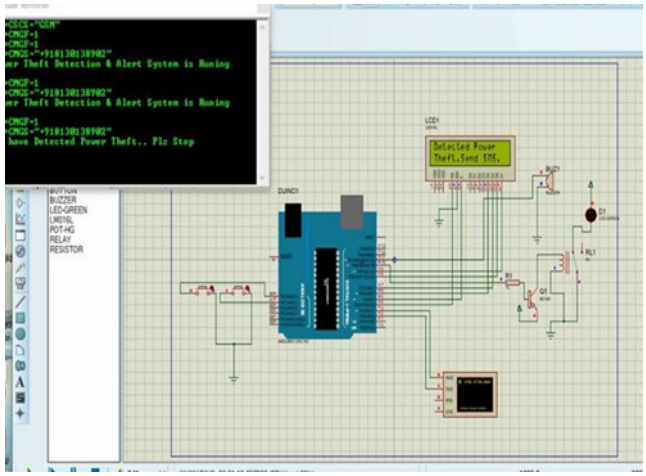


Figure 6-Alert condition

Cut off stage- when the difference between master meter reading and house meter reading is more than 8. Relay is tripped and the corresponding message of cut off is sent in the message screen.

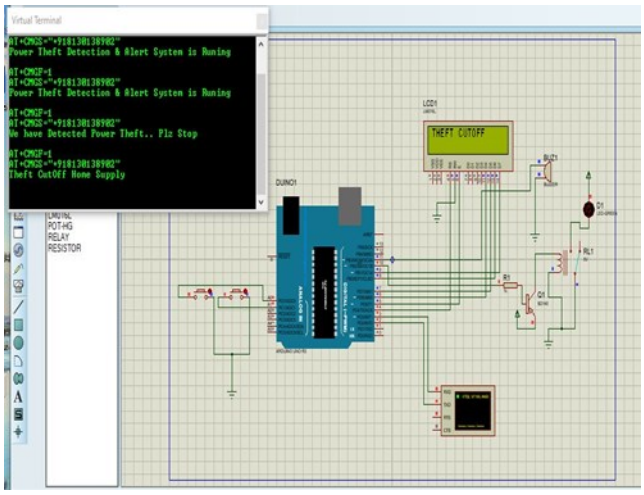


Figure 7-Cut-off condition

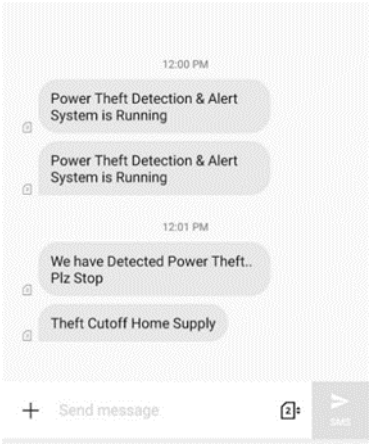


Figure 8 - Message Screenshot

Hardware Results

Under normal condition i.e. house meter reading is equal to master meter shown in figure-9. Alert stage: - when theft is observed. Master meter reading and house meter readings have a difference of 5 units, an alert message is sent inside the screen that power theft has been observed as shown in figure-10. Cut off stage: - when the difference between

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:55:26 UTC from IEEE Xplore. Restrictions apply.

master meter reading and house meter reading is more than 10. Relay is tripped and the corresponding message of cut off is send as shown in figure-11



Figure 9-Normal condition



Figure 10-Alert condition



Figure 11-Cut-off condition

Conclusion

The proposed method of power theft minimizes the heavy power and revenue loss that results from the theft of power by customers. With the completion of the project, it is concluded that power theft can be effectively reduced by detecting when does the power theft occurs and simultaneously informing the authorities to disconnect the power to the house or customer trying to steal the power supply by an integrated automatic breaker that is attached to the unit. The system is used to send or communicate data to the consumer domain as well through a GSM wireless network. The proposed system will reduce energy waste and save more for future use.

References

[1] T. B. Smith, "Electricity theft: a comparative analysis," Elsevier Journal Energy Policy, vol. 32, no. 18, pp. 2067-2076, Dec. 2004.

[2] K. Dineshkumar, Prabhu Ramanathan, Sudha Ramasamy. "Development of ARM processor based electricity theft control system using GSM network", 2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015], 2015

[3] N Mohammad,A Barua and M Arafat,"A smart prepaid energy metering system to control electricity theft", *in proceedings of International Conference on Power, Energy and Control (ICPEC)*, vol.562, no.565, pp.6-8, Feb.2013.

[4] Nitin K Mucheli, Umakanta Nanda, D Nayak, P K Rout, S K Swain, S K Das, S M Biswal. "Smart Power Theft Detection system", 2019 Devices for Integrated Circuit (DevIC), 2019

[5] M. Saad, M. F. Tariq, A. Nawaz and M. Y. Jamal, "Theft detection based GSM prepaid electricity system", *IEEE International Conference on Control Science and Systems Engineering (ICCSSE)*, pp. 435-438, 2017.

[6] S.S.S.R. Depuru, L. Wang, and V. Devabhaktuni, "A conceptual designusing harmonics to reduce pilfering of electricity," in proceedings of IEEE Power and Energy Society General Meeting, pp. 1-7, Jul. 2010.

[7] J. Nagi, K.S. Yap, S.K. Tiong, S.K. Ahmed, and A.M. Mohammad,"Detection of abnormalities and electricity theft using genetic Support Vector Machines," in proceedings of IEEE Region 10 Conference, pp. 1-6, Nov. 2008.

[8] A.Pasdar and S.Mirzakuchaki, "A solution to remote detecting of illegalelectricity usage based on smart metering," proceedings of 2nd International Workshop on Soft Computing Applications, pp. 163-167, Aug. 2007.

[9] ATMEL website.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:55:26 UTC from IEEE Xplore. Restrictions apply.

OPSEARCH
https://doi.org/10.1007/s12597-020-00464-7

THEORETICAL ARTICLE

Best criteria selection based PROMETHEE II method

Anjali Singh[1,2] · **Anjana Gupta**[2] · **Aparna Mehra**[3]

Accepted: 8 July 2020
© Operational Research Society of India 2020

Abstract

In recent years, PROMETHEE II has emerged as one of the most efficient and successfully applied outranking method in multi-criteria decision making (MCDM) problems. Conventional PROMETHEE II method presumes that the weights of the criteria are known beforehand, and this is rather a stringent assumption. Moreover, the number of preference indices to be evaluated in the method grows enormously with an increase in the number of alternatives to be ranked. In this work, we look at these concerns and propose a methodology inspired by PROMETHEE II that identifies the best criteria first and then, the preference indices are computed with respect to these criteria only. The algorithm is customized in such a way that it curtails the computational time complexity of PROMETHEE II to a great extent. Here, a large number of alternatives-criteria evaluation matrices are simulated with random numeric data to illustrate the proposed method and evaluate its performance. Statistical analysis on the stability of the rankings and the conformity measure proves that the proposed approach may be preferred over PROMETHEE II and TOPSIS methods. Rank correlation coefficients are evaluated, and they indicate the efficacy of the proposed method.

Keywords Multi-criteria decision making · PROMETHEE II · TOPSIS · Stability degree · Conformity metric · Kendall's tau-b rank correlation coefficient

Mathematics Subject Classification 91B06 · 90B50

✉ Anjana Gupta
anjanagupta@dce.ac.in

[1] Department of Mathematics and Humanities, Mahatma Gandhi Institute of Technology, JNTU, Hyderabad, Telangana 500075, India

[2] Department of Applied Mathematics, Delhi Technological University, Delhi 110042, India

[3] Department of Mathematics, Indian Institute of Technology, Delhi 110016, India

Published online: 22 July 2020

1 Introduction

MCDM problems involve analysing the subjective assessments of a finite number of alternatives under a certain predetermined, finite number of criteria by an expert/decision-maker (DM) or a group of experts/decision-makers (DMs). In the latter case, it is called a multi-criteria group decision making (MCGDM) problem, see[1]. Over the last few decades, numerous MCDM methods have been proposed under different set-ups for selecting the best alternative among the available ones. For ready reference, see[2–5]. The advancement in MCDM methods is stimulated not only by different applications to real-life problems but also by researchers' inclination towards introducing extended decision-making methodologies as discussed in[6].

1.1 Research issues and need for the present work

Conventional PROMETHEE (Preference Ranking Organization Method for Enrichment Evaluations) method is one of the most successfully practiced MCDM method. The technique was designed by Brans[7] and extended by Vincke and Brans[8]. The PROMETHEE II version is a protracted method that requires a priori knowledge of the criteria weights. It presumes that the decision-maker is a well-versed expert in the domain of the problem to decide upon these weights, reflecting the importance of criteria in ranking the alternatives. This imparts a subjective character to criteria weights varying which alter the ranking of the alternatives. Another concern is that the computational efforts in PROMETHEE II method increase with the number of alternatives and criteria. For a given decision-making problem with 'm' alternatives and 'n' criteria, one needs to evaluate '$mn(m-1)$' preference indices. This paper combines these two interrelated issues. One it noticed that these many computations are necessarily not required and another that the prior knowledge of the criteria weights can be dropped from the PROMETHEE II method.

1.2 Key differences between closely related studies and the work presented in this paper

The significant differences between the proposed work and closely related studies available in the literature are as follows:

1. In literature[9–11], the conventional PROMETHEE II method and its extensions are framed to evaluate '$mn(m-1)$' preference indices for a given decision making problem with 'm' alternatives and 'n' criteria. In this study, we have customised the existing procedure in such a manner that the reckoning process of preference indices lowers to a great extent.
2. The classical PROMETHEE II method presumes the weights of the criteria beforehand. In the available literature, various techniques like AHP, Entropy[12, 13], etc. are also used to evaluate the weight vector of the criteria. Although,

many researchers have designed various tools like weight stability intervals, walking weights, PROMETHEE VI[14, 15], etc. to study the repercussion of these values, but it brings subjectivity in the decision making process that can change the ranking of the alternatives. So, in this paper, we have first recognized the best criterion(a)/feature(s) for each alternative and then assign the uniform weight to the best criteria and no importance i.e., zero weight to the vestigial criteria. Such procedure of weight assignment to criteria is anomalous as it involves both the alternatives and criteria and vary with each alternative in an MCDM problem. Even after the withdrawal of prerequisite of the prior knowledge of weight vector, the statistical discussion given in Sect. 5 reveals that the suggested scheme performs efficiently. The ranking obtained by the proposed methodology is highly correlated as well as more stable than the classical PROMETHEE II ranking. Also, this practice dissolves the element of subjectivity from the criteria importance in an MCDM problem.

3. The computational complexity of an algorithm is also an indicator of its efficiency. The traditional PROMETHEE II has computational time complexity of $O(m^2n)$. Recently, Calders and Assche[16] suggested the method that curtails the time complexity of PROMETHEE with linear preference function to $O(mn\log(m))$. In our study, the computational time complexity of the proposed approach is of $O(mn)$ that makes the performance of the algorithm remarkable.
4. The statistical analysis is an essential and significant element of any methodology to prove its efficiency. In view of statistical studies discussed in Sect. 5, the ranking of alternatives obtained by the proposed algorithm is not merely highly correlated with the classical PROMETHEE II and TOPSIS methods' ranking but also more stable for 10 distinct pairs of numbers of alternatives and criteria. The conformity values also assert the analogous efficacy of proposed algorithm and PROMETHEE II. The conformity of designed method may be agreeing the traditional PROMETHEE II but the reduced complexity of the suggested scheme is the highlight of the work.

1.3 Focus of the present study

In this paper, we propose a methodology inspired by PROMETHEE II, which deliberates on first identifying the best criteria for each alternative and then evaluates the preference indices for these criteria only. The procedure is customised in such a way that the number of preference indices to be estimated reduces to a greater extent. Moreover, the method proposes to assign uniform weights to the best criteria and zero weight to the remaining criteria with each alternative. In this way, no single weight vector is attached to the criteria; instead, the weight vectors depend both on the alternatives and the criteria and vary with alternatives in an MCDM problem. The algorithm is contrived in such a manner that it also curtails the computational time complexity of the classical PROMETHEE II to a noteworthy extent.

The suggested method is validated by randomly generating a large number of alternatives-criteria evaluation matrices which are also used to compute stability degree and conformity measure of the method. To demonstrate the efficiency of the

designed algorithm, the study includes PROMETHEE II and TOPSIS methods in the statistical analysis which exhibits that the proposed method is comparable with the existing methods and can be opted by the practitioners for real-life decision-making problems.

1.4 Organization of the paper

The organization of remainder of this paper unfolds as follows. Section 2 briefs the literature of the two MCDM methods - PROMETHEE II and TOPSIS. Section 3 describes the proposed method. Section 4 presents an example by simulating alternatives-criteria evaluation matrix to illustrate the proposed method. Section 5 discusses the analysis and efficiency of the method via stability degree, conformity measure and rank correlation coefficient. Section 6 concludes the paper.

2 Literature survey

The outranking methods aim to build a preference relation, called an outranking relation, through a series of pairwise comparisons of alternatives to be evaluated on criteria. The study of the outranking method PROMETHEE commenced with the pioneer works of Brans and his collaborative in the early eighties. The PROMETHEE-I method for partial ranking and PROMETHEE II method for the complete ranking of the alternatives were proposed in the work of Brans[7]. The success of these two methods led to the development of more versions of PROMETHEE method, namely, PROMETHEE III, IV, and V. Human brain representation was incorporated in PROMETHEE VI method, proposed by Brans and Mareschal[15] followed by PROMETHEE GDSS for group decision-making, introduced by Macharis et al. [17] and the visual interactive module GAIA (Geometrical Analysis for Interactive Aid) for graphical representation, given by Mareschal and Brans[18, 19]. Moreover, Figueira et al.[20] introduced two extended approaches: PROMETHEE-TRI for handling sorting problem and PROMETHEE CLUSTER for nominal classification. All these versions and extensions of PROMETHEE have been applied in many fields, including environment management[9, 11], business and financial management[10, 21], chemistry[22, 23], energy management[24, 25], social management[26, 27], and many more. For a comprehensive survey of PROMETHEE methods see[28]. Among all the methods of this family, PROMETHEE II is the most acclaimed and frequently used one.

Among numerous other MCDM methods, TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) is a compensatory method that determines the solution which is at the shortest distance from the positive ideal solution and the farthest distance from the negative ideal solution of the MCDM problem. It has been often used across different application areas since its inception by Hwang and Yoon[29]. TOPSIS has been successfully applied in supply chain management and logistics[30, 31], health, safety and environment management[32, 33], human resources management[34, 35], energy management[36, 37], and many more. One can refer to[29, 38] for a detailed

description and working of the method and to the work of Behzadian[39] for a comprehensive survey on the use of TOPSIS method in MCDM. We have briefly mentioned this method here for it has been used in the latter part of this paper for performance analysis and comparison. Both methods require a priori knowledge of criteria weights for aggregating the information in data for penultimate ranking of alternatives. This assumption makes the methods sensitive to the choice of criteria weights.

3 Proposed methodology

An MCDM problem involves m alternatives A_i, $i = 1, \ldots, m$, to be evaluated on n criteria C_j, $j = 1, \ldots, n$. Let $I = \{1, \ldots, m\}$ and $J = \{1, \ldots, n\}$ be the indices sets of alternatives and criteria respectively. Let an alternative-criteria evaluation matrix be $R = [r_{ij}]_{m\times n}$, where r_{ij} represents the evaluation of ith alternative on jth criteria. The proposed method involves the following steps:

Step 1: The classical PROMETHEE II method characterizes six types of preference functions to compare the alternatives under various cost and beneficial criteria. The provision of distinct preference functions makes the normalization process a choice in classical PROMETHEE II. In this paper, the pairwise difference is adopted as the preference function that requires the uniformity of all criteria. Hence, the normalization process becomes a fundamental requirement for the method.

Normalize the matrix R using linear normalization given as $R_{norm} = [\widetilde{r}_{ij}]_{m\times n}$, where

$$\widetilde{r}_{ij} = \begin{cases} \dfrac{r_{ij} - \min\limits_{i\in I}\{r_{ij}\}}{\max\limits_{i\in I}\{r_{ij}\} - \min\limits_{i\in I}\{r_{ij}\}} & \text{if } j\text{th criterion is a beneficial criterion;} \\ \dfrac{\max\limits_{i\in I}\{r_{ij}\} - r_{ij}}{\max\limits_{i\in I}\{r_{ij}\} - \min\limits_{i\in I}\{r_{ij}\}} & \text{if } j\text{th criterion is a cost criterion.} \end{cases}$$

The max-min normalization is preferred as the above process converts all criteria into the beneficial criteria that makes nature of all criteria uniform and it is also an essential requirement for the upcoming step.

Step 2: For a given alternative A_i, find the maximum value over all criteria

$$M_i = \max_{j\in J}\{\widetilde{r}_{ij}\}.$$

Let M_i be achieved on k_i number of criteria ($k_i \leq n$), say, $C_{\gamma_i(1)}, \ldots, C_{\gamma_i(k_i)} \in \{C_1, \ldots, C_n\}$; set $J_i = \{\gamma_i(1), \ldots, \gamma_i(k_i)\}$.

Step 3: For $j \in J_i$, evaluate the preference index $P_j(A_i, A_{i'})$, $i, i' \in I$, $i \neq i'$, as follows:

$$P_j(A_i, A_{i'}) = \begin{cases} 0 & \text{if } \widetilde{r}_{ij} \leq \widetilde{r}_{i'j} \\ \widetilde{r}_{ij} - \widetilde{r}_{i'j} & \text{otherwise.} \end{cases}$$

Here $P_j(A_i, A_{i'})$ is the preference index (value) of the alternative A_i over alternative $A_{i'}$ under criterion C_j. So, for the ith alternative, the preference indices $P_j(A_i, A_i')$, $j \in J_i$, are computed only under those criteria which correspond to the maximum evaluation of the alternative A_i.

The conventional PROMETHEE II method requires $mn(m-1)$ preference indices to participate in the procedure to rank m alternatives on n criteria whereas the designed method reduces it to evaluate only $((m-1)\sum_{i=1}^{m} k_i)$ number of preference indices.

Step 4: Compute the cumulative preference index $\pi(A_i, A_i')$, for $i, i' \in I$, $i \neq i'$, as follows:

$$\pi(A_i, A_i') = \frac{1}{k_i} \sum_{j \in J_i} P_j(A_i, A_{i'}).$$

The conventional PROMETHEE II method assumes the criteria weight vector beforehand. In the proposed method, since the criterion(a) having maximum value represents the best feature(s) of an alternative, hence, for each $i \in I$, we propose to attach the uniform weight $1/k_i$ with criteria in J_i, and weight zero with all remaining criteria in $J \setminus J_i$.

The concern of the prior knowledge of weight vector may be dropped as the statistical results given in Sect. 5 validate the efficiency and competency of the proposed algorithm. The ranking obtained by proposed algorithm is highly correlated as well as more stable than the classical PROMETHEE II ranking. It also makes the approach independent of the decision-makers' perspective on the criteria importance/weights.

Step 5: Determine the outgoing and the incoming flows described respectively as follows:

$$\phi^+(A_i) = \frac{1}{m-1} \sum_{\substack{i'=1 \\ i \neq i'}}^{m} \pi(A_i, A_{i'}), \quad i \in I,$$

$$\phi^-(A_i) = \frac{1}{m-1} \sum_{\substack{i'=1 \\ i \neq i'}}^{m} \pi(A_{i'}, A_i), \quad i \in I,$$

and, compute the net flow

Table 1 Computational time complexity of PROMETHEE II and designed methodology

Principle activity	PROMETHEE II	Proposed method
1. Normalization	$O(mn)$	$O(mn)$
2. Preference indices and cumulative preference indices	$O(mn(m-1))$	$O(m(m-1)+mn)$
3. Net flows	$O(m(m-1))$	$O(m(m-1))$
4. Sorting the net flows values to rank	$O(m\log m)$	$O(m\log m)$
Computational time complexity	$O(m^2n)$	$O(mn)$

$$\phi(A_i) = \phi^+(A_i) - \phi^-(A_i), \qquad i \in I.$$

Step 6: The alternatives are ranked with the largest value of $\phi(A_i)$ considered as the best alternative.

Remark 3.1 For a given $m \times n$ alternatives-criteria evaluation matrix, the computational time complexity of suggested scheme is $O(mn)$ which is indeed less than the complexity $O(m^2n)$ of the conventional PROMETHEE-II method. Table 1 depicts the same.

The time complexity of proposed method has also decreased to a noteworthy extent. Hence, the designed methodology abridges the complexity of PROMETHEE II method.

The flowchart given in Fig. 1 summarizes the proposed method.

4 Illustrative example

Consider an MCDM problem with 7 alternatives and 9 criteria, out of which, criteria C_1, C_2, C_3, C_6, C_8, C_9 are the beneficial criteria and criteria C_4, C_5, C_7 are the cost criteria. The 7×9 alternative-criteria evaluation matrix is populated with a random data from $U(0, 1)$ distribution to illustrate the proposed method. Let this matrix be R.

$$\begin{bmatrix} 0.2956 & 0.9715 & 0.5639 & 0.2711 & 0.1345 & 0.3183 & 0.3315 & 0.5127 & 0.6675 \\ 0.5696 & 0.6743 & 0.3813 & 0.104 & 0.0997 & 0.2724 & 0.4562 & 0.2391 & 0.4185 \\ 0.5231 & 0.1166 & 0.0101 & 0.5562 & 0.4602 & 0.0472 & 0.7755 & 0.876 & 0.9642 \\ 0.6077 & 0.9339 & 0.9153 & 0.3164 & 0.717 & 0.2123 & 0.8049 & 0.8063 & 0.238 \\ 0.4758 & 0.5839 & 0.8552 & 0.9422 & 0.1279 & 0.3202 & 0.346 & 0.2821 & 0.6372 \\ 0.2278 & 0.4051 & 0.1107 & 0.924 & 0.7876 & 0.2157 & 0.0529 & 0.2402 & 0.1467 \\ 0.4636 & 0.2086 & 0.5459 & 0.6851 & 0.4435 & 0.6745 & 0.4678 & 0.6737 & 0.898 \end{bmatrix}.$$

The normalized matrix is evaluated using the normalization process defined for cost and beneficial criteria in step 1 of Sect. 3 and is given as R_{norm}.

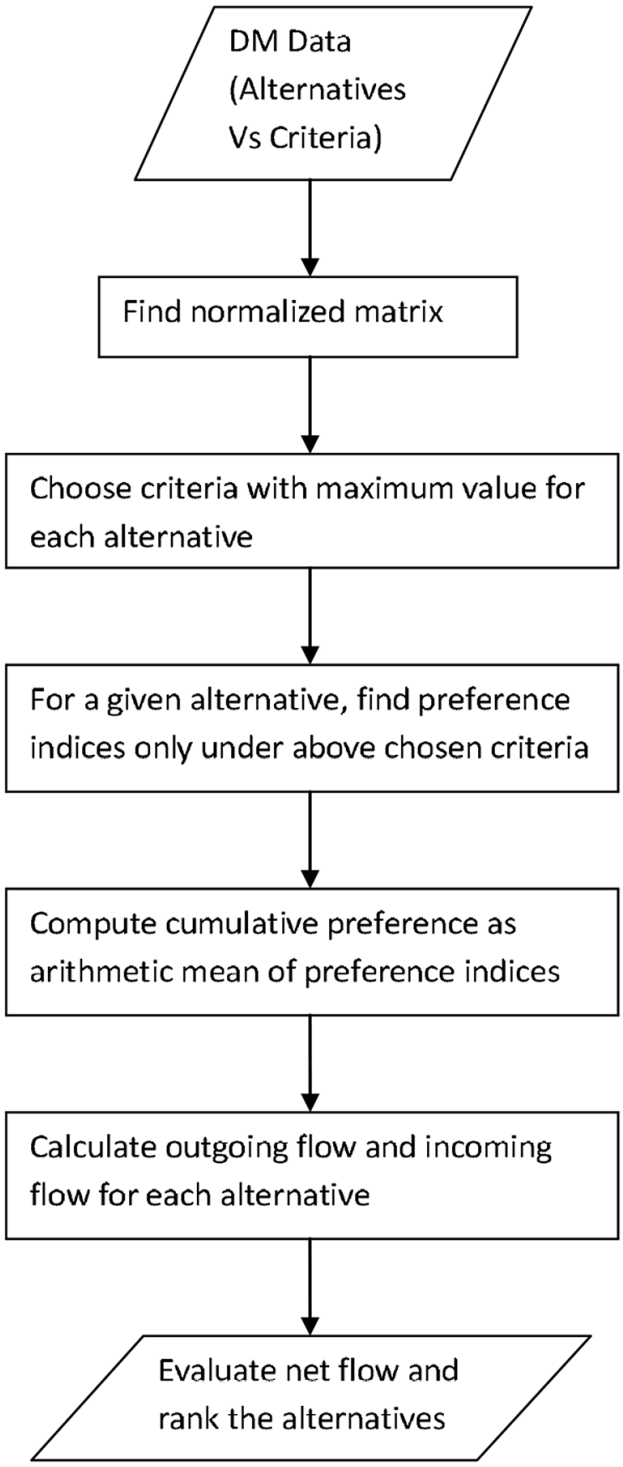


Fig. 1 Flow chart for the proposed methodology

$$R_{norm} = \begin{bmatrix} 0.1785 & 1.0000 & 0.6118 & 0.8006 & 0.9494 & 0.4322 & 0.6295 & 0.4296 & 0.6371 \\ 0.8997 & 0.6524 & 0.4101 & 1.0000 & 1.0000 & 0.3590 & 0.4637 & 0.0000 & 0.3325 \\ 0.7773 & 0.0000 & 0.0000 & 0.4605 & 0.4759 & 0.0000 & 0.0391 & 1.0000 & 1.0000 \\ 1.0000 & 0.9560 & 1.0000 & 0.7466 & 0.1026 & 0.2632 & 0.0000 & 0.8906 & 0.1117 \\ 0.6528 & 0.5466 & 0.9336 & 0.0000 & 0.9590 & 0.4352 & 0.6102 & 0.0675 & 0.6000 \\ 0.0000 & 0.3375 & 0.1111 & 0.0217 & 0.0000 & 0.2686 & 1.0000 & 0.0017 & 0.0000 \\ 0.6207 & 0.1076 & 0.5919 & 0.3067 & 0.5002 & 1.0000 & 0.4483 & 0.6824 & 0.9190 \end{bmatrix}$$

Table 2 Maximum value and the index set of criteria at which the maximum value is achieved

Alternative	Maximum value (M_i)	Index set of criteria J_i
A_1	1	{2}
A_2	1	{4, 5}
A_3	1	{8, 9}
A_4	1	{1, 3}
A_5	0.9590	{5}
A_6	1	{7}
A_7	1	{6}

The maximum value M_i and the index set J_i of each alternative are given in Table 2.

The preference indices are evaluated under the criteria that attain the maximum value for given alternative as mentioned in Table 3.

Note the number of preference indices to be evaluated is 60, whereas the conventional PROMETHEE II method would have taken 378 evaluations.

The cumulative preference indices are given in Table 4.

For each alternative, the outgoing flow, the incoming flow, and the net flow are depicted in Table 5.

By net flow values, A_1 is the best alternative, followed by A_7 and A_2.

To validate the proposed method, the ranking of alternatives are also obtained by conventional PROMETHEE II and the TOPSIS method. The weights of the criteria in the former approach are taken to be uniform for the nine criteria. The two variants of TOPSIS method described in the literature, please see[40], differ only in the normalization process. Here, we use TOPSIS with linear normalization (same as in Step 1 of the designed methodology) to make a conducive comparison with the ranking of the suggested method. In this work, the solution procedures for PROMETHEE II and TOPSIS methods are not provided, but one can see[9, 38] for the reference. The rankings of seven alternatives by the three methods are tabulated in Table 6.

The ranking achieved by the designed method is same as ranking from the two conventional methods, making the proposed method promising.

5 Empirical analysis and discussion

Although in certain situations, the proposed methodology, PROMETHEE II method, and TOPSIS (with linear normalisation) can result in similar rankings (as in above example), but eventually they can lead to different rankings. Therefore, compatibility and stability of rankings obtained by various decision-making methods must be analysed for the validity and efficiency of the new method.

Table 3 Preference indices $P_j(A_i, A_{i'}),\ j \in J_i,\ i, i' \in I,\ i \neq i'$

	C1	C2	C3	C4	C5	C6	C7	C8	C9
$P(A_1, A_2)$		0.3476							
$P(A_1, A_3)$		1							
$P(A_1, A_4)$		0.0440							
$P(A_1, A_5)$		0.4534							
$P(A_1, A_6)$		0.6625							
$P(A_1, A_7)$		0.8924							
$P(A_2, A_1)$				0.199356	0.050589				
$P(A_2, A_3)$				0.5395	0.5241				
$P(A_2, A_4)$				0.2534	0.8974				
$P(A_2, A_5)$				1	0.041				
$P(A_2, A_6)$				0.9783	1				
$P(A_2, A_7)$				0.6933	0.4998				
$P(A_3, A_1)$								0.5704	0.3629
$P(A_3, A_2)$								1	0.6675
$P(A_3, A_4)$								0.1094	0.8883
$P(A_3, A_5)$								0.9325	0.4
$P(A_3, A_6)$								0.9983	1
$P(A_3, A_7)$								0.3176	0.0810
$P(A_4, A_1)$	0.8215		0.3882						
$P(A_4, A_2)$	0.1003		0.5899						
$P(A_4, A_3)$	0.2227		1						
$P(A_4, A_5)$	0.3472		0.0664						
$P(A_4, A_6)$	1		0.8889						
$P(A_4, A_7)$	0.3793		0.4081						
$P(A_5, A_1)$					0.0096				
$P(A_5, A_2)$					0				
$P(A_5, A_3)$					0.4831				
$P(A_5, A_4)$					0.8564				
$P(A_5, A_6)$					0.9590				
$P(A_5, A_7)$					0.4588				
$P(A_6, A_1)$							0.3705		
$P(A_6, A_2)$							0.5363		
$P(A_6, A_3)$							0.9609		
$P(A_6, A_4)$							1		
$P(A_6, A_5)$							0.3898		
$P(A_6, A_7)$							0.5517		
$P(A_7, A_1)$						0.5678			
$P(A_7, A_2)$						0.6410			
$P(A_7, A_3)$						1			
$P(A_7, A_4)$						0.7368			
$P(A_7, A_5)$						0.5648			
$P(A_7, A_6)$						0.7314			

Table 4 Cumulative preference indices

Cumulative preference indices			
$\pi(A_1, A_2) = 0.3476$	$\pi(A_2, A_1) = 0.1250$	$\pi(A_3, A_1) = 0.4667$	$\pi(A_4, A_1) = 0.6049$
$\pi(A_1, A_3) = 1.0000$	$\pi(A_2, A_3) = 0.5318$	$\pi(A_3, A_2) = 0.8338$	$\pi(A_4, A_2) = 0.3451$
$\pi(A_1, A_4) = 0.0440$	$\pi(A_2, A_4) = 0.5754$	$\pi(A_3, A_4) = 0.4989$	$\pi(A_4, A_3) = 0.6113$
$\pi(A_1, A_5) = 0.4534$	$\pi(A_2, A_5) = 0.5205$	$\pi(A_3, A_5) = 0.6662$	$\pi(A_4, A_5) = 0.2068$
$\pi(A_1, A_6) = 0.6625$	$\pi(A_2, A_6) = 0.9891$	$\pi(A_3, A_6) = 0.9991$	$\pi(A_4, A_6) = 0.9444$
$\pi(A_1, A_7) = 0.8924$	$\pi(A_2, A_7) = 0.5965$	$\pi(A_3, A_7) = 0.1993$	$\pi(A_4, A_7) = 0.3937$
$\pi(A_5, A_1) = 0.0096$	$\pi(A_6, A_1) = 0.3705$	$\pi(A_7, A_1) = 0.5678$	
$\pi(A_5, A_2) = 0.0000$	$\pi(A_6, A_2) = 0.5363$	$\pi(A_7, A_2) = 0.6410$	
$\pi(A_5, A_3) = 0.4831$	$\pi(A_6, A_3) = 0.9609$	$\pi(A_7, A_3) = 1.0000$	
$\pi(A_5, A_4) = 0.8564$	$\pi(A_6, A_4) = 1.0000$	$\pi(A_7, A_4) = 0.7368$	
$\pi(A_5, A_6) = 0.9590$	$\pi(A_6, A_5) = 0.3898$	$\pi(A_7, A_5) = 0.5648$	
$\pi(A_5, A_7) = 0.4588$	$\pi(A_6, A_7) = 0.5517$	$\pi(A_7, A_6) = 0.7314$	

Table 5 Different flows for each alternative

Alternative	Outgoing flow	Incoming flow	Net flow
A_1	0.5667	0.3574	0.2093
A_2	0.5564	0.4506	0.1057
A_3	0.6107	0.7645	− 0.1539
A_4	0.5177	0.6186	− 0.1009
A_5	0.4611	0.4669	− 0.0058
A_6	0.6349	0.8809	− 0.2461
A_7	0.707	0.5154	0.1916

Table 6 Ranking of alternatives by three methods

	Method	Rank of alternatives
1.	Proposed method	$A_1 > A_7 > A_2 > A_5 > A_4 > A_3 > A_6$
2.	PROMETHEE II	$A_1 > A_7 > A_2 > A_4 > A_5 > A_3 > A_6$
3.	TOPSIS with linear normalization	$A_1 > A_7 > A_2 > A_4 > A_5 > A_3 > A_6$

Wu et al.[41] discussed the quantitative method for measuring the stability of rankings in group decision-making. The stability degree of an MCDM method is a quantitative measure, with values in [0, 1], and the highest degree gives the most stable ranking.

Conformity measure evaluates the performance of an MCDM method vis a vis other MCDM methods, and it is to be minimised to improve the performance of an MCDM method, see[42].

The rank correlation coefficient is used to analyse the similarity between the rankings given by each method. Commonly used rank correlation coefficients are

Spearman's rank-order correlation coefficient, Kendall's tau-b and Pearson's correlation coefficient. The Kendall's tau-b is a non-parametric rank correlation coefficient which computes the degree of similarity between two sets of ranks for the same set of objects. The Kendall's tau-b coefficient also takes into account ties in rankings, a feature desirable in the context of our present study.

For m (number of alternatives), n (number of criteria), and q (number of randomly generated MCDM matrices), the procedure of computing the aforementioned measures is discussed below.

Step 1: For $t = 1$ to q,
generate a random alternative-criteria evaluation matrix R_t of size $m \times n$.
Step 2: Pass each R_t, $t = 1, \ldots, q$, into three methods {Proposed Method, PROMETHEE II, TOPSIS} to obtain rankings for the m alternatives. The total number of such rankings is q for each of the three methods.
Step 3: Computing the stability degree:

3(a)(i) Determine the score matrices $(SM)_{Algo} = [N_{is}]_{m \times m}$, where N_{is} is the number of times an alternative A_i appears at the sth position in q ranks obtained in Step 2, and $Algo \in \{$ Proposed Method, PROMETHEE II, TOPSIS $\}$. For example, in matrix $(SM)_{ProposedMethod}$, the N_{21} depicts the number of times the alternative A_2 appears at 1st position while ranking the alternatives using the proposed methodology.

3(a)(ii) For each method, evaluate the weighted score vector as $S_{Algo} = (SM)_{Algo} W$, where the weight vector is taken as $W = (m-1, m-2, \ldots, 1, 0)^T$ from the voting rule of Borda count method, as discussed in[41], and $Algo \in \{$ Proposed Method, PROMETHEE II, TOPSIS $\}$. The alternatives are ranked with the first one having the highest score in S_{Algo} vector[1].
The combined (overall) ranking of alternatives and the score vector S_{Algo}, where $Algo \in \{$ Proposed Method,PROMETHEE II, TOPSIS $\}$, are used to calculate the stability degree of the rankings.

3(b)(i) Let $I^{(1)}_{Algo}$ and $I^{(2)}_{Algo}$ be the sets of indices in I corresponding to the alternatives having rank respectively, 1 and 2 in the combined ranking obtained by the $Algo$. Note that the sets $I^{(\ell)}_{Algo}$, $\ell = 1, 2$, can have more than one single index. Calculate the vector $U_{Algo} = (u^1_{Algo}, \ldots, u^m_{Algo})$, as follows:

[1] The vector S_{Algo} is formed by combining q alternative-criteria matrices to rank the alternatives; this scheme can easily be extended to the Group Decision MCDM problems.

$$u^k_{Algo} = \begin{cases} S^k_{Algo} & \text{if } \; k \notin I^{(1)}_{Algo} \cup I^{(2)}_{Algo} \\ \frac{S^k_{Algo}+S^b_{Algo}}{2} & k \in I^{(1)}_{Algo}, \\ & \text{and for any index } \; b \in I^{(2)}_{Algo} \\ \frac{S^k_{Algo}+S^b_{Algo}}{2} & k \in I^{(2)}_{Algo}, \\ & \text{and for any index } \; b \in I^{(1)}_{Algo}. \end{cases}$$

Here, S^k_{Algo}, S^b_{Algo} are the k-th and b-th components of the vector S_{Algo} respectively.

3(b)(ii) Find the stability degree ρ_{Algo} of $Algo \in$ { Proposed Method, PROMET-HEE II, TOPSIS }, by computing sine of the angle between the vectors S_{Algo} and U_{Algo} as follows:

$$\rho_{Algo} = \frac{||S_{Algo} \times U_{Algo}||}{||S_{Algo}||\,||U_{Algo}||}, \quad \times \text{ stands for the cross product .}$$

Step 4: Computing the conformity values:

4(i) For $t = 1$ to q, compute the conformity values

$$C^t_{Algo} = \sum_{i=1}^{m} \left(\text{ Net Flow}^t_{Algo,(i)} - \text{ Mean Flow}^t_{(i)} \right)^2,$$

$Algo \in$ { Proposed Method, PROMETHEE II, TOPSIS }. For TOPSIS, the net flow value of each alternative is replaced by its relative closeness value. And the mean flow for each alternative is evaluated by taking the arithmetic mean of the net flow values (or relative closeness values for TOPSIS) of m alternatives, for each of the three methods.

4(ii) Determine the mean and the variance of these conformity values.

$$\overline{C}_{Algo} = \frac{1}{q} \sum_{t=1}^{q} C^t_{Algo}$$

$$\sigma^2_{C_{Algo}} = \frac{1}{q} \sum_{t=1}^{q} (C^t_{Algo} - \overline{C}_{Algo})^2.$$

4(iii) Apply the hypothesis testing to test the significance of the difference between the three vectors $C_{Algo} = \{C^1_{Algo}, C^2_{Algo}, \ldots, C^q_{Algo}\}$, $Algo \in$ { Proposed Method, PROMETHEE II, TOPSIS }.

Step 5: Computing the rank correlation coefficient:

5(i) Go to step 3 and obtain the combined rank for each of the three methods using 3(a)(i) and 3(a)(ii).

5(ii) Evaluate Kendall's tau-b correlation coefficient to analyse the similarity between the rankings given by each method.

Table 7 Stability degrees from three methods

S. No.	No. of alternatives	No. of criteria	Proposed method	PROMETHEE	TOPSIS
1	3	5	0.00097	0.00026	0.00032
2	5	6	0.00106	9.3287×10^{-5}	0.00019
3	7	9	0.00033	5.3452×10^{-6}	1.4254×10^{-5}
4	10	10	0.00056	0.00015	4.9690×10^{-5}
5	10	15	0.00023	3.2299×10^{-5}	0.00013
6	12	15	0.00014	0.00011	3.9711×10^{-5}
7	15	20	0.00014	2.9994×10^{-5}	3.8341×10^{-5}
8	20	25	8.7711×10^{-5}	6.1415×10^{-5}	5.4092×10^{-5}
9	25	40	0.00016	4.0659×10^{-5}	6.6232×10^{-5}
10	30	50	0.00011	0.0001	0.0001

In the empirical analysis, we took different pairs (m, n) for alternatives-criteria and for each such pair we randomly simulated 1,00,000 independent and identically distributed matrices of size $m \times n$. The global stability degrees of the combined rankings of these 1,00,000 matrices with different numbers of alternatives and criteria are reported in Table 7.

It is evident from Table 7 that, for each pair of alternatives and criteria, the rankings delivered by the designed method are more stable than those given by the other two methods. This approves the advantage of the proposed method in solving MCDM problems.

Next, we compute the conformity vectors C_{TOPSIS}, $C_{\text{PROMETHEE}}$ and $C_{\text{Proposed Method}}$ for 1,00,000 matrices of size $m \times n$, as generated in stability analysis. To check if there is a significant difference between these three vectors, we set the two null hypotheses $H_0^{(1)}$ and $H_0^{(2)}$ and their corresponding alternate hypotheses $H_1^{(1)}$ and $H_1^{(2)}$ as follows:

$$
\begin{aligned}
H_0^{(1)} &: \overline{C}_{\text{TOPSIS}} - \overline{C}_{\text{Proposed Method}} = 0,\\
H_1^{(1)} &: \overline{C}_{\text{TOPSIS}} > \overline{C}_{\text{Proposed Method}},\\
H_0^{(2)} &: \overline{C}_{\text{Proposed Method}} - \overline{C}_{\text{PROMETHEE}} = 0, \quad \text{and}\\
H_1^{(2)} &: \overline{C}_{\text{Proposed Method}} > \overline{C}_{\text{PROMETHEE}}.
\end{aligned}
$$

The one-tailed Z-test is conducted at 99% level of significance. The results are reported in Table 8.

From Table 8, the proposed method performs better than the TOPSIS method. And the Z-statistics confirms the rejection of the null hypothesis $H_0^{(1)}$ at 99% significance level which concludes that the mean conformity of the proposed algorithm is less than that of the TOPSIS method to a noteworthy extent. It is also worth noticing that although the mean conformity of the PROMETHEE II

Table 8 First two moments of the conformity vectors and their Z-statistics

S. No.	No. of alternatives	No. of criteria	TOPSIS		Proposed method		PROMETHEE II		Z-statistics	
			Mean	S.d.	Mean	S.d.	Mean	S.d.	TOPSIS—proposed method	Proposed method—PROMETHEE
1	3	5	0.3661	0.0488	0.1303	0.0390	0.1119	0.0268	3.7742	0.3899
2	5	6	0.5882	0.0677	0.1972	0.0364	0.1543	0.0191	5.0877	1.0422
3	7	9	0.8038	0.0734	0.2547	0.0356	0.2091	0.0193	6.7296	1.1262
4	10	10	1.1420	0.0954	0.3561	0.0387	0.2951	0.0239	7.6376	1.3424
5	10	15	1.1319	0.0770	0.3354	0.0326	0.2919	0.0198	9.5262	1.1424
6	12	15	1.3578	0.0900	0.4032	0.0357	0.3496	0.0227	9.8578	1.2674
7	15	20	1.6901	0.0948	0.4878	0.0352	0.4335	0.0239	11.8830	1.2738
8	20	25	2.2478	0.1084	0.6382	0.0375	0.5751	0.0272	14.0383	1.3634
9	25	40	2.7988	0.1035	0.7667	0.0337	0.7120	0.0260	18.6740	1.2827
10	30	50	3.3546	0.1081	0.9082	0.0341	0.8517	0.0272	21.5876	1.2969

is lower than the designed method yet the null hypothesis $H_0^{(2)}$ is accepted given the Z-statistics at 99% confidence. It signifies that there is essentially no significant difference between $\overline{C}_{\text{PROMETHEE II}}$ and $\overline{C}_{\text{Proposed Method}}$. Hence, the efficiency of the proposed method and the PROMETHEE II method are analogous.

To analyse the performance of the suggested scheme, for a small sample size data, the above procedure is followed for sample size 10 and 20. The conformity vectors C_{TOPSIS}, $C_{\text{PROMETHEE}}$ and $C_{\text{Proposed Method}}$ are evaluated and hypothesis testing is carried out to check if there is a significant difference between these three vectors. The two null hypotheses $H_0^{(1)}$ and $H_0^{(2)}$ are same as before.

The one-tailed t-test is conducted at 99% level of significance. The results are reported in Table 9 for sample size 10 and in Table 10 for sample size 20. The p-value is mentioned in the bracket.

From Tables 9 and 10, we can conclude that the proposed algorithm performs better than the TOPSIS method with the same reason as for the large sample size. And also, the efficiency of the suggested method and PROMETHEE II are comparable. The analysis indicates the efficiency and adeptness of the designed method.

For the next analysis, various pairs (m, n) for alternatives-criteria are taken and for each such pair, 10 independent and identically distributed matrices of size $m \times n$ are randomly simulated. The combined rankings for each of the three methods are obtained, and Kendall's tau-b correlation coefficients are evaluated using IBM SPSS Statistics 21 software as given in Table 11.

It is evident from Table 11, that the ranking obtained by the proposed method is highly correlated with the rankings acquired by PROMETHEE II and TOPSIS with statistically significant p-values.

In summary, the proposed algorithm is an efficient and conducive approach. The method obtains the most stable ranking of the alternatives that is highly correlated with the rankings acquired by the existing methods. Also, the conformity values assert the better performance of the suggested scheme in comparison to TOPSIS while the efficacy of the suggested approach and PROMETHEE II are analogous.

Note: The experts and researchers working in this field may arise a concern that the weights of the criteria reflect the preferences of the DM and hence, the omission of pre-defined criteria weight vector is not justifiable. In reference to this issue, the authors agree with the fact that the individual's preference of the criteria is of significant importance, but it is not always possible for DM to quantify his preference for all available criteria. It may also be claimed that various well-known techniques to evaluate the exact criteria weights exist in the literature. The adoption of such techniques is the DM's choice, but these methodologies add to the complexity of the decision situation and emanate the decision-making as a time-consuming process. In such a scenario, the suggested method calculates the criteria weights itself to overcome the subjectivity of human thinking in the decision-making. The statistical results given in Sect. 5 confirm the legitimacy of the proposed method and hence, validate the efficacy and competency of the scheme.

Table 9 First two moments of the conformity vectors and their t-statistics for sample size 10

S. No.	No. of alternatives	No. of criteria	TOPSIS		Proposed method		PROMETHEE II		t-statistics	
			Mean	S.d.	Mean	S.d.	Mean	S.d.	TOPSIS—proposed method	Proposed method—PROMETHEE
1	3	5	0.3663	0.0546	0.1368	0.0307	0.1053	0.0183	3.6637	0.8846 (0.0126)
2	5	6	0.6024	0.0706	0.2039	0.0323	0.1573	0.0184	5.1342	1.2525 (0.0034)
3	7	9	0.7803	0.0782	0.2463	0.0380	0.2037	0.0169	6.1421	1.0239 (0.0020)
4	10	10	1.1383	0.0870	0.3301	0.0268	0.2925	0.0231	8.8836	1.0635 (0)
5	10	15	1.1328	0.0617	0.3242	0.0255	0.2896	0.0156	12.1055	1.1549 (0)
6	12	15	1.3608	0.0688	0.3768	0.0225	0.3477	0.0174	13.5967	1.0230 (0)
7	15	20	1.7167	0.0563	0.4746	0.0112	0.4384	0.0149	21.6561	1.9405 (0)
8	20	25	2.2268	0.1048	0.6038	0.0400	0.5644	0.0275	14.4756	0.8113 (0)
9	25	40	2.8097	0.0780	0.7372	0.0248	0.7081	0.0195	25.3284	0.9210 (0)
10	30	50	3.3425	0.0445	0.8688	0.0124	0.8415	0.0115	53.5804	1.6126 (0)

Table 10 First two moments of the conformity vectors and their t-statistics for sample size 20

S. No.	No. of alternatives	No. of criteria	TOPSIS		Proposed method		PROMETHEE II		t-statistics	
			Mean	S.d.	Mean	S.d.	Mean	S.d.	TOPSIS—proposed method	Proposed method—PROMETHEE
1	3	5	0.3629	0.0540	0.1271	0.0381	0.1129	0.0337	3.5698	0.2797 (0.0251)
2	5	6	0.5625	0.0562	0.2066	0.0335	0.1505	0.0169	5.4409	1.4944 (0)
3	7	9	0.7800	0.0947	0.2485	0.0344	0.2047	0.0243	5.2762	1.0387 (0)
4	10	10	1.1293	0.0996	0.3563	0.0374	0.2921	0.0249	7.2648	1.4282 (0)
5	10	15	1.1339	0.0742	0.3239	0.0201	0.2911	0.0178	10.5360	1.2197 (0)
6	12	15	1.3707	0.1020	0.4073	0.0394	0.3540	0.0253	8.8107	1.1364 (0)
7	15	20	1.6514	0.0710	0.4780	0.0358	0.4240	0.0210	14.7582	1.3006 (0)
8	20	25	2.2724	0.0855	0.6363	0.0280	0.5797	0.0206	18.1891	1.6268 (0)
9	25	40	2.8155	0.1350	0.7722	0.0414	0.7167	0.0344	14.4742	1.0295 (0)
10	30	50	3.3469	0.1274	0.9075	0.0384	0.8498	0.0310	18.3385	1.1691 (0)

Table 11 Kendall's tau-b correlation coefficients of proposed method ranking with PROMETHEE and TOPSIS rankings, the 2-tailed significance values are given in brackets

S. No.	No. of alternatives	No. of criteria	PROMETHEE	TOPSIS
1	3	5	1.000** (0)	1.000** (0)
2	5	6	0.894* (0.037)	0.894* (0.037)
3	7	9	0.683* (0.033)	0.586 (0.068)
4	10	10	0.395 (0.121)	0.524* (0.043)
5	15	20	0.546** (0.005)	0.476* (0.015)

*indicates significance of correlation at 1% and **indicates significance of correlation at 5%

6 Conclusions

PROMETHEE II is a highly researched and most applied outranking method. In this paper, inspired by the PROMETHEE II method, we proposed a scheme which overcomes two concerns with the PROMETHEE II method. One is the prior knowledge of the criteria weight vector, and second is the number of computations going in the performance indices, especially when the number of alternatives is large. The computational time complexity of the classical PROMETHEE II has also reduced from $O(m^2n)$ to $O(mn)$ that makes the algorithm efficient. The stability analysis confirms the suitability and usage of the proposed method by experts/decision-makers. The conformity analysis proves that the designed method is equally as good as the classical PROMETHEE II method. The correlation coefficient approves the ranking given by the proposed method makes the method more desirable.

We envision that the proposed method can be extended to the scenarios when information is expressed through linguistic variables, interval-valued linguistic variables, or fuzzy numbers.

Acknowledgements The authors wish to express their gratitude to the esteemed referees for their valuable observations and comments in improving the quality of this research work.

Authors' Contributions Anjali Singh performed the computations and drafted the manuscript. Anjana Gupta and Aparna Mehra conceived of the presented theory and verified the analytical methods. All authors provided critical feedback and helped shape the research, analysis and manuscript. All authors read and approved the final manuscript.

Funding Not applicable.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Availability of data and material Not applicable.

References

1. Lootsma, F.A.: Multi Criteria Desicion Analysis via Ratio and Difference Judgement. Kluwer Academic Publishers, Dordrecht (1999)
2. Beccali, M., Cellura, M., Mistretta, M.: Decision-making in energy planning: application of the ELECTRE method at regional level for the diffusion of renewable energy technology. Renew. Energy **28**, 2063–2087 (2003)
3. Hokkanen, J., Salminen, P.: ELECTRE III and IV decision aids in an environmental problem. J. MultiCriteria Decis. Anal. **6**, 215–226 (1997)
4. Keyser, W.D., Peeters, P.: Theory and methodology: a note on the use of PROMETHEE multicriteria methods. Eur. J. Oper. Res. **89**, 457–461 (1996)
5. Shiha, H.S., Shyurb, H.J., Lee, E.S.: An extension of TOPSIS for group decision-making. Math. Comput. Modell. **45**, 801–813 (2007)
6. Wiecek, M.M., Ehrgott, M., Fadel, G., Figueira, J.R.: Multiple criteria decision-making for engineering. Omega **36**, 337–339 (2008)
7. Brans, J.P.: Lingenierie de la decision. Elaboration dinstruments daide a la decision. In: Methode PROMETHEE, Nadeau, R., Landry, M., (eds.) Laide a la Decision: Nature, Instruments et Perspectives Davenir. Presses de Universite Laval, Quebec, Canada, pp. 183–214 (1982)
8. Vincke, Ph, Brans, J.P.: A preference ranking organization method: the PROMETHEE method for multiple criteria decision-making. Manag. Sci. **31**, 647–656 (1985)
9. Pohekar, S.D., Ramachandran, M.: Multi-criteria evaluation of cooking energy alternatives for promoting parabolic solar cooker in India. Renew. Energy **29**, 1449–1460 (2004)
10. Albadvi, A., Chaharsooghi, S.K., Esfahanipour, A.: Decision-making in stock trading: an application of PROMETHEE. Eur. J. Oper. Res. **177**, 673–683 (2007)
11. Queiruga, D., Walther, G., Gonza'lez-Benito, J., Spengler, T.: Evaluation of sites for the location of WEEE recycling plants in Spain. Waste Manag. **28**, 181–190 (2008)
12. Singh, A., Gupta, A., Mehra, A.: An AHP-PROMETHEE II method for 2-tuple linguistic multicriteria group decision-making. In: 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), IEEE Xplore, pp. 1–6 (2015). https://doi.org/10.1109/ICRITO.2015.7359374
13. Singh, A., Gupta, A., Mehra, A.: Energy planning problems with interval valued 2-tuple linguistic information. Oper. Res. Int. J. (2016). https://doi.org/10.1007/s12351-016-0245-x
14. Mareschal, B.: Weight stability intervals in multicriteria decision aid. Eur. J. Oper. Res. **33**, 54–64 (1988)
15. Brans, J.P., Mareschal, B.: The PROMETHEE VI procedure: how to differentiate hard from soft multicriteria problems. J. Decis. Syst. **4**, 213–223 (1995)
16. Calders, T., Assche, D.V.: PROMETHEE is not quadratic: an $O(qn \log(n))$ algorithm. Omega. (2017). https://doi.org/10.1016/j.omega.2017.04.003
17. Macharis, C., Brans, J.P., Mareschal, B.: The GDSS PROMETHEE procedure: a PROMETHEE-GAIA based procedure for group decision support. J. Decis. Syst. **7**, 283–307 (1998)
18. Mareschal, B., Brans, J.P.: Geometrical representations for MCDA: the GAIA module. Eur. J. Oper. Res. **34**, 69–77 (1988)
19. Brans, J.P., Mareschal, B.: The PROMETHEE GAIA decision support system for multicriteria investigations. Investig. Oper. **4**, 107–117 (1994)
20. Figueira, J., de Smet, Y., Brans, J.P.: MCDA methods for sorting and clustering problems: Promethee TRI and Promethee CLUSTER, Universite Libre de Bruxelles, Service deMathematiques de la Gestion, Working Paper 2004/02 (2004)
21. Halouani, N., Chabchoub, H., Martel, J.M.: PROMETHEE-MD-2T method for project selection. Eur. J. Oper. Res. **195**, 841–895 (2009)
22. Hendriks, M.M.W.B., De Boer, J.H., Smilde, A.K., Doornbos, D.A.: Multicriteria decision-making. Chemometr. Intell. Lab. Syst. **16**, 175–191 (1992)
23. Herngren, L., Goonetilleke, A., Ayoko, G.A.: Analysis of heavy metals in roaddeposited sediments. Anal. Chim. Acta **571**, 270–278 (2006)
24. Pohekar, S.D., Ramachandran, M.: Application of multi-criteria decision-making to sustainable energy planning—a review. Renew. Sustain. Energy Rev. **8**, 365–381 (2004)
25. Zhou, P., Ang, B.W., Poh, K.L.: Decision analysis in energy and environmental modeling: an update. Energy **31**, 2604–2622 (2006)

26. Johnson, M.P.: Spatial decision support for assisted housing mobility counseling. Decis. Support Syst. **41**, 296–312 (2005)
27. Brans, J.P., Macharis, C., Kunsch, P.L., Chevalier, A., Schwaninger, M.: Combining multicriteria decision aid and system dynamics for the control of socio-economic processes, an iterative real-time procedure. Eur. J. Oper. Res. **109**, 428–441 (1998)
28. Behzadian, M., Kazemzadeh, R.B., Albadvi, A., Aghdasi, M.: PROMETHEE: a comprehensive literature review on methodologies and applications. Eur. J. Oper. Res. **200**(1), 198–215 (2010)
29. Hwang, C.L., Yoon, K.P.: Multiple Attribute Decision-Making: Methods and Applications. Springer, New York (1981)
30. Chen, C.T., Lin, C.T., Huang, S.F.: A fuzzy approach for supplier evaluation and selection in supply chain management. Int. J. Prod. Econ. **102**, 289–301 (2006)
31. Zhang, G., Shang, J., Li, W.: An information -granulation entropy-based model for third-party logistics providers evaluation. Int. J. Prod. Res. **50**, 177–190 (2012)
32. Grassi, A., Gamberini, R., Mora, C., Rimini, B.: A fuzzy multi-attribute model for risk evaluation in workplaces. Saf. Sci. **47**, 707–716 (2009)
33. Aiello, G., Enea, M., Galante, G., La Scalia, G.: Clean agent selection approached by fuzzy TOPSIS decision-making method. Fire Technol. **45**, 405–418 (2009)
34. Boran, F.E., Genc, S., Kurt, M., Akay, D.: Personnel selection based on intuitionistic fuzzy sets. Hum. Factors Ergon. Manuf. Serv. Ind. **21**(5), 493–503 (2011)
35. Saremi, M., Mousavi, S.F., Sanayei, A.: TQM consultant selection in SMEs with TOPSIS under fuzzy environment. Expert Syst. Appl. **36**, 2742–2749 (2009)
36. Doukas, H., Karakosta, C., Psarras, J.: Computing with words to assess the sustainability of renewable energy options. Expert Syst. Appl. **37**, 5491–5497 (2010)
37. Chamodrakas, I., Martakos, D.: A utility-based fuzzy TOPSIS method for energy efficient network selection in heterogeneous wireless networks. Appl. Soft Comput. **11**, 3734–3743 (2011)
38. Wei, G.W.: Extension of TOPSIS method for 2-tuple linguistic multiple attribute group decision-making with incomplete weight information. Knowl. Inf. Syst. **25**, 623–634 (2010)
39. Behzadian, M., Otaghsara, S.K., Yazdani, M., Ignatius, J.: A state-of the-art survey of TOPSIS applications. Expert Syst. Appl. **39**(17), 13051–13069 (2012)
40. Opricovic, S., Tzeng, G.H.: Compromise solution by MCDM methods: a comparative analysis of VIKOR and TOPSIS. Eur. J. Oper. Res. **156**, 445–455 (2004)
41. Wu, F., Zhao, Y., Chen, Y.: Stability analysis of group decision-making under weighted scoring rules. Int. J. Syst. Sci. (2015). https://doi.org/10.1080/00207721.2015.1128577
42. Razaei, J.: Best-worst multi-criteria decision-making method. Omega **53**, 49–57 (2015)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal of Public Health | pp. 1–9 | doi:10.1093/pubmed/fdaa099

Breast cancer burden in Africa: evidence from GLOBOCAN 2018

Rajesh Sharma

University School of Management and Entrepreneurship, Delhi Technological University, Delhi 110095, India
Address correspondence to Rajesh Sharma, E-mail: rajesh.sharma@dtu.ac.in.

ABSTRACT

Background Breast cancer is the leading malignancy in African females. This study aims to examine the breast cancer burden in Africa using recently released GLOBOCAN 2018 estimates.

Methods The incidence and mortality estimates of age- and country-wise burden of breast cancer in 54 African countries were obtained from GLOBOCAN 2018.

Results In Africa, breast cancer caused 74 072 deaths, and 168 690 cases were estimated to have occurred in 2018. The age-standardized incidence rate stood at 37.9/100 000 in Africa, varying from 6.9/100 000 in the Gambia to 69.6/100 000 in Mauritius. The age-standardized mortality rate stood at 17.2/100 000 in 2018, ranging from 4/100 000 in the Gambia to 29.1/100 000 in Somalia in 2018. Nigeria was the leading country in terms of absolute burden with 26 310 cases and 11 564 deaths, followed by Egypt with 23 081 new cases and 9254 deaths. The mortality-to-incidence ratio for Africa stood at 0.44, varying from 0.24 in Libya to 0.68 in the Central African Republic.

Conclusion To tackle breast cancer burden in Africa, the main challenges are late-stage disease presentation, lack of screening and therapeutic infrastructure, lack of awareness and limited resources.

Keywords Africa, breast cancer, GLOBOCAN, incidence, mortality

Introduction

Breast cancer is the leading cancer group in African females and its burden has been rising steadily in the last three decades owing to factors such as longevity, westernization of lifestyle characterized by delayed marriages, a shorter period of breast-feeding and lower age at menarche.[1–5] The mortality rates are higher and survival rates are lower in Africa than developed countries due to poor oncologic care, ranging from prevention to palliation.[6] The different disease outcomes and survivorship in Africa were ascribed to genetic variation[7] and tumor stage at diagnosis.[8] Much of breast cancer (80–90%) arises sporadically due to a multitude of risk factors, whereas a small percentage is associated with inherited gene mutations (e.g. BRCA 1 or BRCA 2).[9, 10]

This study examines the burden of breast cancer in Africa using recently released GLOBOCAN 2018 estimates.[11] We examined the estimates of age- and country-wise, and age-standardized rates of incidence and mortality of breast cancer in 54 African countries in 2018 in the light of country-level human development index (HDI).[12] The HDI is a summary indicator of three indicators of development: health (life expectancy), education (mean years of schooling and expected years of schooling) and income (gross national income (GNI) per capita). It is a better indicator of human development than its components as it not only reflects healthcare resources (through income per capita) in a country but also encompasses general conditions of population health (through life expectancy) and general awareness levels (through years of schooling). Country-level survival rates of breast cancer were approximated using mortality-to-incidence ratio (MIR), which is a useful proxy of survival rates of different neoplasia.[2, 13, 14] The examination of leading malignancy in African females is expected to help realign public health priorities toward emerging public health concerns such as breast cancer and shape the public health policy toward cancer control and management to achieve

Rajesh Sharma, Assistant Professor

© The Author(s) 2020. Published by Oxford University Press on behalf of Faculty of Public Health. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Downloaded from https://academic.oup.com/jpubhealth/article-abstract/doi/10.1093/pubmed/fdaa099/5869901 by Delhi Technological University user on 20 July 2020

better disease outcomes and improve the quality of life of breast cancer patients.

Methods

Breast cancer estimates for 54 African countries were obtained from GLOBOCAN 2018.[11, 15, 16] The estimates were produced only for invasive carcinoma classified as C50 by the International Classification of Diseases, tenth version (ICD-10). GLOBOCAN 2018 estimates were produced for 18 nonoverlapping age-groups: under 5, 5 to 9, 10 to 14, 15 to 19, 20 to 24, 25 to 29, 30 to 34, 35 to 39, 40 to 44, 45 to 49, 50 to 54, 55 to 59, 60 to 64, 65 to 69, 70 to 74, 75 to 79, 80 to 84 and 85 plus. All-age and age-specific incidence and death counts were reported along with 95% uncertainty intervals (UIs), which reflects the uncertainties associated with cancer registry coverage and death registration, data quality and timeliness of data reporting.[11, 15]

All the GLOBOCAN 2018 estimates are available at global cancer observatory (gco.iarc.fr), and a brief account of Globocan methodology is presented in Appendix 1. The 5-year survival rate of breast cancer was approximated using the country-level MIR, which was calculated by dividing all-age death count with all-age incidence count.

Human development index

We also examined the association of breast cancer burden and country-level HDI.[12, 17] HDI is an indicator of country-level development, which is a composite index of three indices: income index, education index and health index. Each of these indices had values varying between 0 (lowest) and 1 (highest), and the final HDI was calculated by taking the geometric mean of three indicators.[12]

Odd os breast cancer incidence

We also examined the cumulative risk of developing and dying from breast cancer in African countries during their lifetime (below the age of 75 years) using the following formula.

$$\text{Cumulative Risk} = 1 - \exp^{-\text{CR}}$$

In the above equation, CR denotes cumulative rate and was calculated by $\text{CR} = \sum_i \text{Rate}_i * T_i$ and Rate_i is the rate of incidence or mortality in ith age-group and T_i is the number of years in the ith age-group. The odds of developing or dying from breast cancer were calculated as per the following.

$$\text{ODDS} = \frac{1}{\text{Cumulative Risk}}$$

Results

Geographical distribution of breast cancer

In Africa, breast cancer caused an estimated 74 072 deaths [95% UI, 67 345–81 472] with 168 690 [152 332–186 804] cases were estimated to have occurred in 2018. The age-standardized incidence rate (ASIR) stood at 37.9/100 000 in Africa, whereas the age-standardized mortality rate (ASMR) stood at 17.2/100 000 in 2018. Nigeria was the leading country in terms of absolute burden with 26 310 [23 610–29319] cases and 11 564 [10 302–12 981] deaths followed by Egypt with 23 081 [21 734–24 512] new cases and (9254 [8707–9836]) deaths in 2018 (Fig. 1a and b; Supplementary Table S1). Geographically, high all-age incidence counts were also reflected in high all-age death numbers, the age-standardized rates, however, were quite asymmetric between incidence and mortality in Africa (Fig. 1a–d).

The ASIR was highest in Mauritius (69.6/100 000) and lowest in the Gambia (6.9/100 000) in 2018 (Fig. 1c; Supplementary Table S1). Besides Mauritius in eastern Africa, there were pockets of high incidence rates in north Africa (e.g. Algeria: 55.6/100 000; Egypt: 52.4/100 000), central Africa (e.g. Republic of Congo: 37.8/100 000) and southern Africa (e.g. South Africa: 49/100 0000; Zimbabwe: 38.4/100 000) (Fig. 1c). The ASMR was the highest in Somalia (29.1/100 000), with the Gambia recording the lowest ASMR of 4/100 000 in 2018. The high ASMRs were mostly concentrated near central and northeast Africa toward the great horn of Africa (e.g. Central African Republic: 23.9/100 000; Somalia: 29.1/100 000; Eritrea: 24.8/100 000; Sierra Leone 25.4/100 000). Low ASMRs were largely concentrated toward the southeastern region of Africa (e.g. Botswana: 7.1/100 000; Madagascar: 7.5/100 000).

The MIR for Africa stood at 0.44 in 2018 with widespread variation across the region ranging from 0.24 in Libya to 0.68 in Central African Republic (Fig. 1e; Supplementary Table S1). High MIRs were mostly concentrated in central sub-Saharan Africa (SSA) (e.g. Central African Republic: 0.68; Equatorial Guinea: 0.64), eastern SSA (e.g. Somalia: 0.675; Eritrea: 0.58) and western SSA (e.g. Niger: 0.64; Chad:0.59). The MIRs were the lowest in north Africa (e.g. Algeria: 0.28; Libya: 0.24).

Age-wise, breast cancer incidence and deaths increased up to the 45–49 age-group and declined after that (Fig. 2). This pattern was preserved in nearly all African countries, with the incidence being the highest in middle age-groups (30–49) (Supplementary Figure S1a and b).

Association between breast cancer burden and HDI

Figure 3a–c displays the relationship between various metrics of breast cancer burden and country-level HDI. Both

Downloaded from https://academic.oup.com/jpubhealth/article-abstract/doi/10.1093/pubmed/fdaa099/5869901 by Delhi Technological University user on 20 July 2020

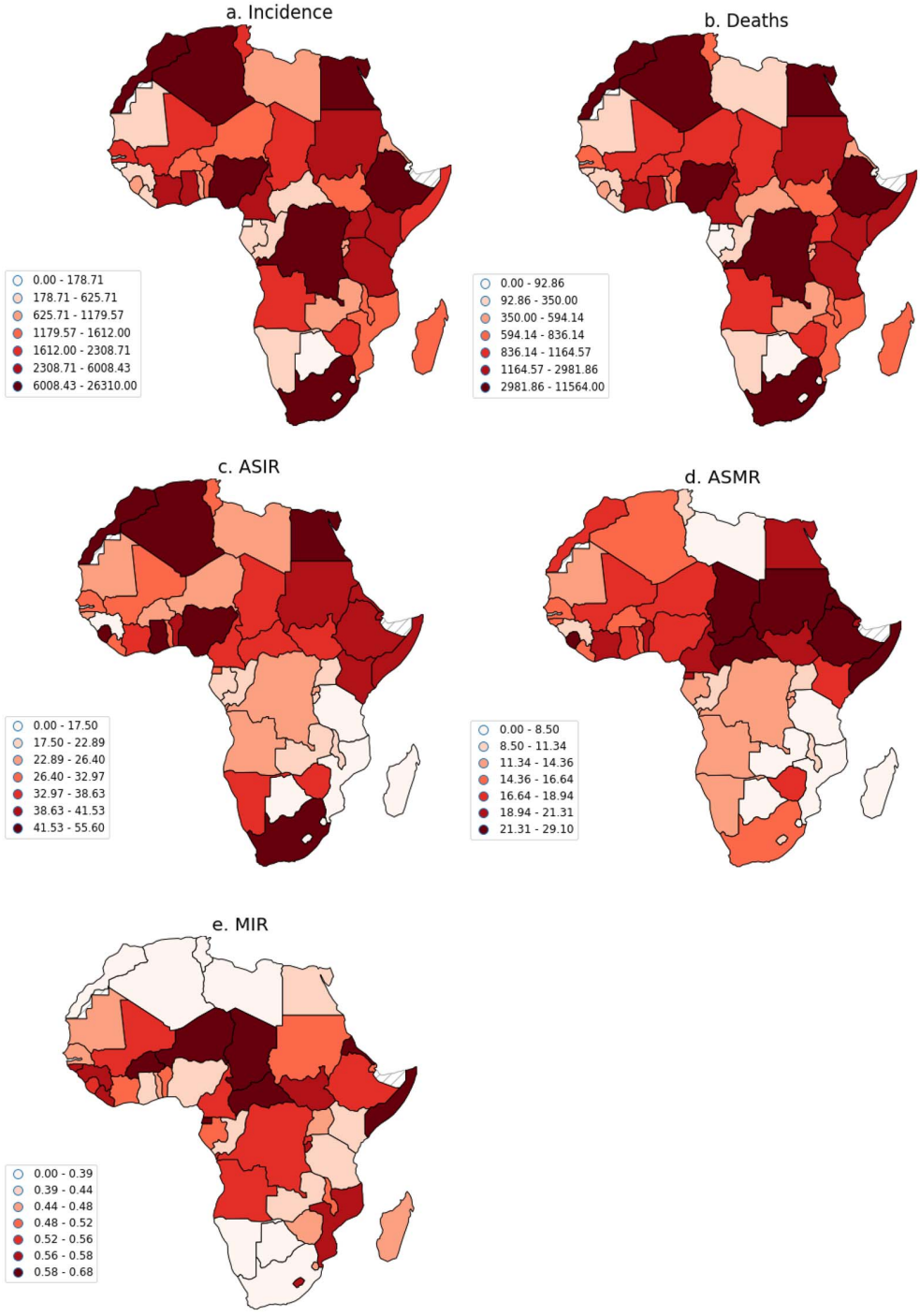


Fig. 1 Geographical distribution of breast cancer burden in Africa (a) Incidence (b) Deaths (c) ASIR (d) ASMR (e) MIR. Incidence: All-age new cases; Deaths: All-age death counts; ASIR: Age-standardized incidence rate; ASMR: Age-standardized mortality rate; MIR: Mortality-to-incidence ratio. Data Source: GLOBOCAN 2018 (IARC). Figure couldn't show the heatmap for the countries Cape Verde, Comoros, Mauritius and Sao Tome and Principe whose country-specific values can be obtained from Supplementary Table S1.

all-age incidence count and ASIRs were positively linked with HDI, albeit with few exceptions, with comparatively developed countries having high counts as well as incidence rates (Fig. 3a). The relationship between all-age incidence and HDI, although statistically insignificant (at 5% level of significance), was positive; the relationship between ASIR and HDI, however, was positive and statistically significant (Supplementary Table S2). Mortality, however,

Downloaded from https://academic.oup.com/jpubhealth/article-abstract/doi/10.1093/pubmed/fdaa099/5869901 by Delhi Technological University user on 20 July 2020

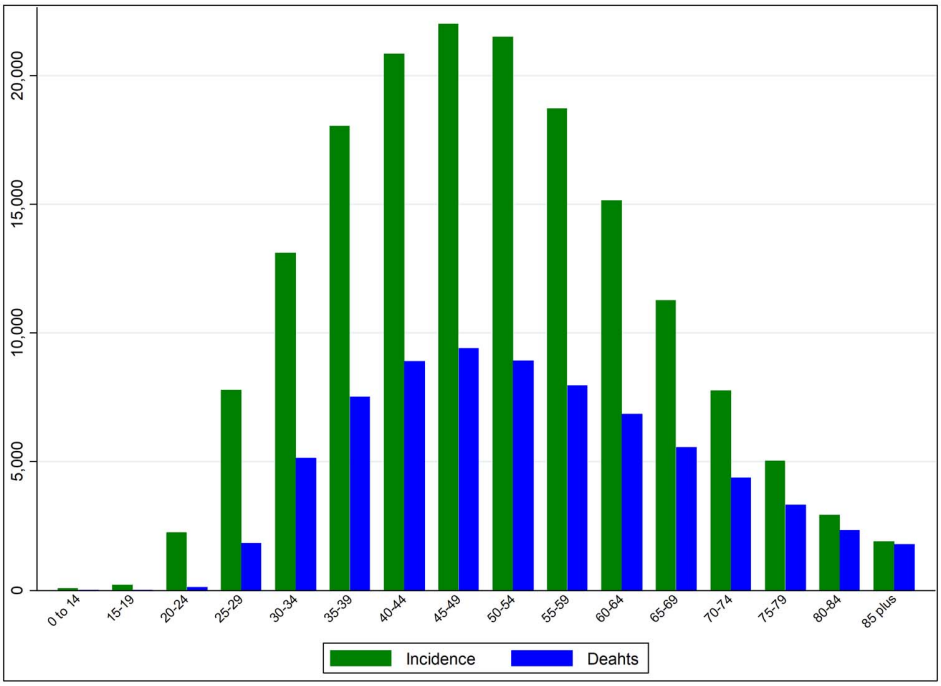


Fig. 2 Age-group wise distribution of breast cancer incidence and death counts in 2018. Incidence: Age-wise new cases; Deaths: Age-wise death counts. Data Source: GLOBOCAN 2018 (IARC).

showed a slightly different relationship with HDI; all-age death counts were positively linked (although statistically insignificant) with HDI, the ASMRs, however, were negatively associated (although statistically insignificant) with HDI (Fig. 3b; Supplementary Table S2). MIR, a proxy for 5-year survival rate had strong negative and statistically significant association (at 1% level of significance) with HDI (R-squared: 0.6384), with high HDI countries of Africa (e.g. Algeria, HDI:0.75 and MIR:0.28; Mauritius, HDI: 0.79 and MIR: 0.33) registering comparatively higher survival rates (i.e. low MIR; Fig. 3c and Supplementary Table S2).

Odds of breast cancer

In Africa, the odds of developing breast cancer were 1 in 25 females (probability: 3.94%) during their lifetime (before the age 75 years), whereas 1 in 57 females (probability: 1.77%) were expected to die of breast cancer during their lifetime (Fig. 4; Supplementary Table S3). The risk of developing breast cancer varied 10-fold in Africa; it was the highest in Mauritius (7.27%, odds: 1 in 14) and the lowest in the Gambia (0.71%, odds: 1 in 141). The probability of dying from breast cancer varied 7-fold in Africa, from 0.42% (odds: 1 in 240) in the Gambia to Somalia (2.97%, odds: 1 in 34). We noted that the odds of developing cancer were higher in comparatively developed countries of Africa, whereas the risk of breast cancer death displayed a negative association with HDI (Supplementary Figure S2a and b).

Discussion

Main finding of this study

Across Africa, the breast cancer burden rose from 92 600 cases and 50 000 deaths in 2008 (as per GLOBOCAN 2008[18]) to 168 690 cases and 74 072 deaths in 2018 (as per GLOBOCAN 2018). The age-standardized rates also displayed increases, with ASIR rising from 28/100 000 in 2008[18] to 37.9/100 000 in 2018, whereas the ASMR rose marginally from 16/100 000 in 2008[18] to 17.2/100 000 in 2018. While the incidence rates increased in developed countries, too, the rise in mortality rates is specific to Africa in line with previous research.[2] As per GLOBOCAN 2018, the ASIR was 84.8/100 000 in North America and 86.7/100 000 in Oceania, whereas the same stood at 37.2/100 000 in Africa in 2018. ASMR, in contrast, was lower in North America (12.6/100 000) and Oceania (14.9/100 000) in comparison with Africa (17.2/100 000). As per GLOBOCAN 2018, in terms of ASIR, no African country featured in top 10, whereas three African countries featured in top 10 in terms of ASMR (In fact, no African country featured in top 25 in terms of ASIR, but 12/25 countries belonged to Africa in terms of ASMR.). High incidence rates, along with low mortality rates in developed regions of the world, reflect diagnosis of breast cancer at an early stage and the fact that better treatment is available in these countries.[19]

Downloaded from https://academic.oup.com/jpubhealth/article-abstract/doi/10.1093/pubmed/fdaa099/5869901 by Delhi Technological University user on 20 July 2020

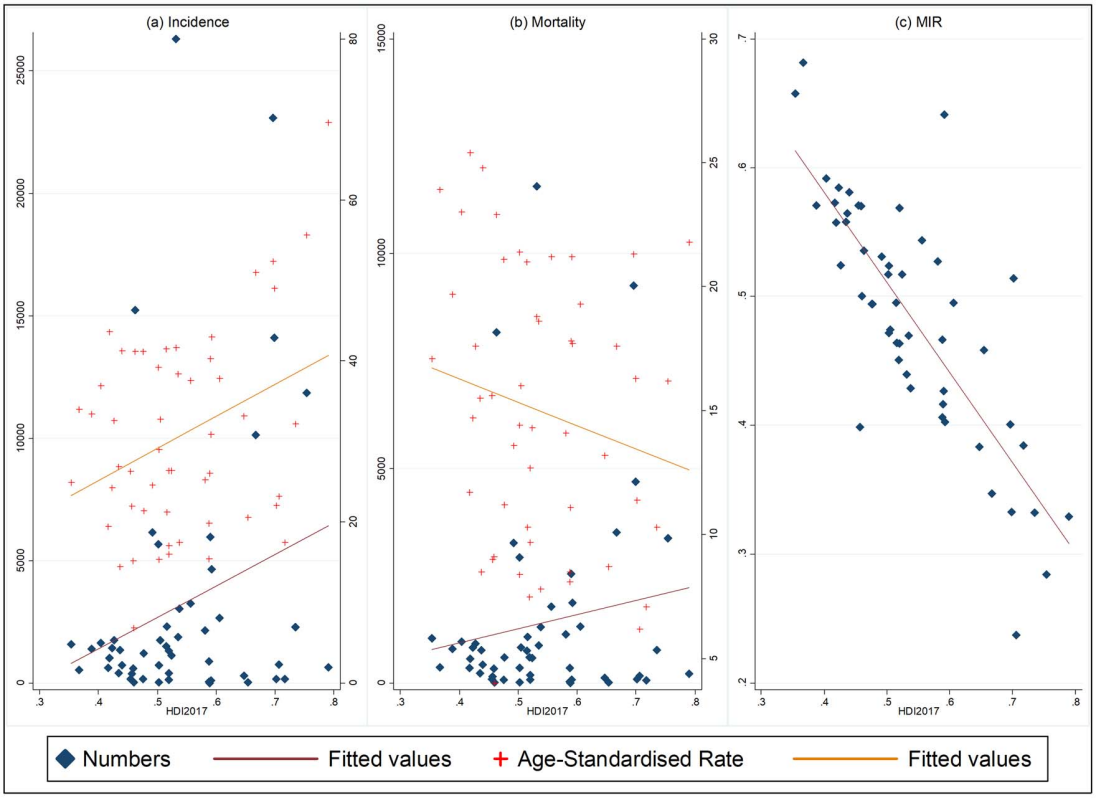


Fig. 3 Bivariate relationship between country-level breast cancer burden and HDI (a) Incidence versus HDI (b) Deaths versus HDI (c) MIR versus HDI. Incidence: All-age new cases in 2018; Deaths: All-age death counts; ASIR: Age-standardized incidence rate; ASMR: Age-standardized mortality rate; MIR: Mortality-to-incidence ratio; HDI: Human Development Index. The data of HDI for 2018 was procured from united nations development program (UNDP). Data Source: GLOBOCAN 2018 (IARC).

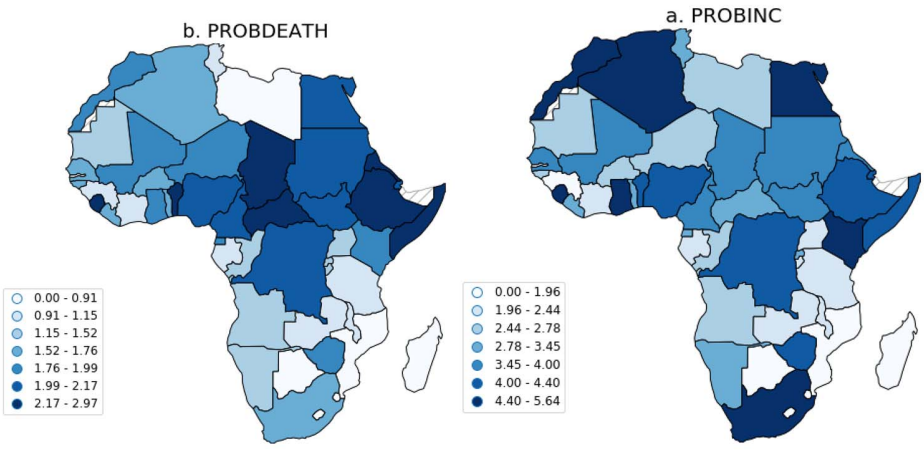


Fig. 4 Geographical distribution of risk of developing and dying from breast cancer in Africa (a) PROBINC (b) PROBDEATH PROBINC: Probability of developing breast cancer during lifetime (before the age 75); PROBDEATH: Probability of dying from breast cancer during lifetime (before the age 75). Figure couldn't show the heatmap for the countries Cape Verde, Comoros, Mauritius and Sao Tome and Principe whose country-specific values can be obtained from Supplementary Table S1.

Downloaded from https://academic.oup.com/jpubhealth/article-abstract/doi/10.1093/pubmed/fdaa099/5869901 by Delhi Technological University user on 20 July 2020

What is already known on this topic

International Agency for Research on Cancer (IARC) previously published GLOBOCAN estimates for the year 2008 and 2012.[20, 21] In the latest iteration, the cancer estimates were provided for the most recent year 2018. While few studies examined the global burden of breast cancer in multicountry context,[2, 5] no study has examined the most recent estimates of breast cancer burden in Africa.

What this study adds

This study examined the burden of breast cancer in 54 African countries measured in terms of age-wise and age-standardized rates of incidence and mortality. The MIR ratio was employed to approximate the 5-year survival rates and the country-wise metrics of breast cancer burden were comprehensively examined in light of the development status of a country measured by its HDI. With few exceptions, greater socioeconomic status measured by HDI was linked with high ASIRs and low ASMRs (Fig. 3a and b). High ASIRs in developing countries of Africa relates to the similar risk factors as in developed countries as well as better screening of breast cancer in a few of these countries. MIR, a proxy for survival rate, had a strong negative and statistically significant association with HDI. The poor survival rates in SSA countries have been documented before[6, 22] and are due to low awareness levels, lack of screening availability and cancer care infrastructure, among other factors.[23]

We found that 50% of patients were below the age 50 (74% below the age of 60) with higher HDI countries had more breast cancer patients in older age-groups as high HDI characterizes greater development and longer life spans (Supplementary Fig S3). Previous studies have highlighted the issue of younger age-profiles of breast cancer patients in Africa.[24] Although lower life expectancy in African countries compared to developed countries is considered to be the biggest reason, younger age of breast cancer patients among African−American women[25] also suggests there must be few factors other than life expectancy such as genetic factors or environmental factors that predispose females of African origin to breast cancer at earlier age.[24] These patterns are particularly concerning given that breast cancer among young women tends to be clinically and pathologically more aggressive with higher mortality rate than that in older women.[26]

Late-disease presentation is a hallmark of breast cancer burden in Africa.[27] A meta-analysis on breast cancer presentation in Africa by Jedy-Agba *et al*.[27] found late-stage presentation varying from 30% in South Africa to 98% in Nigeria with disease presentation also differing among black and nonblack women. As per Abdulrahman and Rahman,[28] 50–75% of African women present with advanced disease, whereas only 11% of breast cancer patients presented with advanced disease (stage III or IV) in the USA.[29] In a country-specific study, Islami *et al*.[30] reported 74% of breast cancer patients in Cote d'Ivoire and 81% in the Democratic Republic of Congo presented with advanced disease.

The advanced disease presentation in Africa results in poor disease outcomes and hence, lower survival rates.[6] The survival rates were found to be 12 and 46% in Gambia and Uganda, respectively.[22] In Africa, lack of awareness, fear of mastectomy and societal fears were a few of the patient-specific reasons responsible for the late-disease presentation and higher mortality rates.[2–4, 23] Moreover, many African women also show up to a traditional healer on first noticing symptoms, who promises treatment without any painful surgery or medicine, which further delays proper diagnosis and poor prognosis ensues.[31–33] In addition to the above patient-specific factors, late-disease presentation also reflects poor oncologic infrastructure, week referral systems and lack of screening services in Africa[23].

While many developed countries can reduce case fatality rates for breast cancer by promoting early detection using mammography screening, many African countries lack either access to screening equipment and services or the means to promote them due to misconception and lack of population awareness. As far as screening is concerned, the developed countries' approach of screening through mammograms cannot be emulated in African countries due to a lack of resources and awareness.[34] As per breast health global initiative, there must be phased approaches toward breast cancer detection as per country's resource availability; in limited-resource settings, the focus must be on raising awareness and capacity building and in highest-resource settings, the focus must be on screening of asymptomatic women.[35]

As mammographic screening is not cost effective in low-resource settings in Africa, clinical breast examination (CBE) may prevent the delay in diagnosis and can facilitate earlier disease presentation. Preliminary results from randomized control trials (RCTs) on the efficacy of CBE in low- and medium-income countries have revealed success in downstaging and early detection of breast cancer.[36–38] Although CBE has not been recommended as primary screening device due to nonconclusive evidence of its mortality-reduction effects, its cost-effectiveness and relative success in downstaging of cancer in RCTs makes it suitable to be recommended for those women who do not receive regular mammograms either because it is not recommended (women aged < 40 and > 69) or because breast mammograms are costly. It can also present an opportunity for healthcare providers to educate women on breast cancer, risk factors, role of early detection in progression and lastly, self-breast examination.[39, 40]

Downloaded from https://academic.oup.com/jpubhealth/article-abstract/doi/10.1093/pubmed/fdaa099/5869901 by Delhi Technological University user on 20 July 2020

Breast cancer risk factors pertain to the duration of breast-feeding, age at menarche and menopause, hormone status, life expectancy, dietary risk factors, exposure to ionizing radiation and genetic predisposition.[41, 42] Few of these risk factors were not prevalent in Africa earlier; in recent times, however, life expectancy is increasing, infant mortality is falling along with declining fertility rates, marital behaviors, and reproductive patterns are changing in Africa.[43, 44] Cumulatively, these behavioral changes have been termed westernization of lifestyle, which has resulted in higher breast cancer risk in African women.[3]

Cancer care infrastructure ranging from preventive medicine to palliative care is often scarce in low- and medium-income countries, including Africa.[24, 45] For instance, even though surgery is the primary treatment option of breast cancer and sometimes the only treatment option, a small percent of breast cancer patients in low- and medium-income countries such as in Africa have access to safe, affordable and timely surgery.[46] Only 25 of the 43 responding African countries had CBE available at public primary healthcare level, whereas 15 of 43 African countries had mammograms available at the public primary health level (Supplementary Table S4) (Only 16 of 43 African countries had national level cancer control plan or strategy by 2014.).

Limitations of the Study

The quality and reliability of estimates depend crucially upon the degree of representativeness and the quality of source information[11, 15]. As per Bray *et al*.[47], across SSA, there were only four population-based registries (Kampala, Uganda; Harare, Zimbabwe; Blantyre, Malawi and PROMEC, South Africa). Moreover, as population-based cancer registries are not available in many of the African countries and estimates were obtained using data from neighboring countries, which may be underreported in few of these countries. Therefore, the breast cancer estimates reported and examined in this paper are expected to be biased. Moreover, the modeling effort may result in too wide UIs, which may render policy deductions lesser useful in few countries.

Conclusion

With ~ 170 000 cases and ~ 75 000 deaths, breast cancer has become one of the most significant public health threats in Africa with incidence rates fast approaching that of developed countries and mortality rates higher than that of developed countries. Breast cancer is a cause of concern in Africa owing to increasing incidence, high mortality, low survival rates, late-stage disease presentation and lack of cancer infrastructure. Increasing life expectancy and westernizing lifestyles will result in greater breast cancer burden in Africa in the near future, and the public health systems in Africa have to gear up to tackle this enormous public health threat. In addition to boosting oncologic infrastructure, there must be effective educational campaigns to educate women about the importance of early detection of malignancy and access to care[48]; for this purpose, low-cost interventions such as CBE can be employed to educate women on all issues relating to breast cancer. Lastly, there must be an increased focus on establishing cancer registries and expanding the coverage of already existing ones in Africa.

Supplementary data

Supplementary data mentioned in the text are available to subscribers in *PUBMED* online.

Ethical approval and Consent to Participate

The research was conducted using data available in the public domain and does not include any human participants and/or animals. Therefore, no ethical approvals were required.

Consent to Publish

The authors of the study have the consent and responsibility of submission to this journal.

Conflict of Interest

The authors declare that they have no conflict of interest.

Funding

The authors received no specific funding for this research.

References

1 Fitzmaurice C, Akinyemiju TF, Al Lami FH *et al.* Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2016: a systematic analysis for the global burden of disease study. *JAMA oncol* 2018;**4**:1553–68.

2 Sharma R. Breast cancer incidence, mortality and mortality-to-incidence ratio (MIR) are associated with human development, 1990–2016: evidence from global burden of disease study 2016. *Breast Cancer* 2019a;**26**:428–45.

Downloaded from https://academic.oup.com/jpubhealth/article-abstract/doi/10.1093/pubmed/fdaa099/5869901 by Delhi Technological University user on 20 July 2020

3 Porter P. "Westernizing" women's risks? Breast cancer in lower income countries. *N Engl J Med* 2008;**358**:213–6.

4 Arnold M, Pandeya N, Byrnes G *et al.* Global burden of cancer attributable to high body-mass index in 2012: a population-based study. *Lancet Oncol* 2015;**16**:36–46.

5 Li N, Deng Y, Zhou L *et al.* Global burden of breast cancer and attributable risk factors in 195 countries and territories, from 1990 to 2017: results from the global burden of disease study 2017. *J Hematol Oncol* 2019;**12**:140.

6 Allemani C, Matsuda T, Di Carlo V *et al.* Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet* 2018;**391**:1023–75.

7 Fregene A, Newman LA. Breast cancer in sub-Saharan Africa: how does it relate to breast cancer in African-American women? *Cancer* 2005;**103**:1540–50.

8 Awadelkarim KD, Arizzi C, Elamin EOM *et al.* Pathological, clinical and prognostic characteristics of breast cancer in Central Sudan versus northern Italy: implications for breast cancer in Africa. *Histopathology* 2008;**52**:445–56.

9 Bradbury AR, Olopade OI. Genetic susceptibility to breast cancer. *Rev Endocr Metab Disord* 2007;**8**:255–67.

10 Kim H, Choi DH. Distribution of BRCA1 and BRCA2 mutations in Asian patients with breast cancer. *J Breast Cancer* 2013;**16**:357–65.

11 Bray F, Ferlay J, Soerjomataram I *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;**68**:394–424.

12 UNDP. *Human Development Indices and Indicators 2018 Statistical Update*. http://hdr.undp.org/sites/default/files/2018_human_development_statistical_update.pdf (3 May 2019, date last accessed)

13 Vostakolaei F, Karim-Kos HE, Janssen-Heijnen ML *et al.* The validity of the mortality to incidence ratio as a proxy for site-specific cancer survival. *Eur J Public Health* 2010;**21**:573–7.

14 Sharma R. The burden of prostate cancer is associated with human development index: evidence from 87 countries, 1990–2016. *EPMA J* 2019b;**10**:137–52.

15 Ferlay J, Colombet M, Soerjomataram I *et al.* Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer* 2019;**144**:1941–53.

16 Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F. *Global Cancer Observatory*. https://gco.iarc.fr/ (1 May 2019, date last accessed)

17 Human Development Database. http://hdr.undp.org/en/data# (3 May 2019, date last accessed)

18 Jemal A, Bray F, Forman D *et al.* Cancer burden in Africa and opportunities for prevention. *Cancer* 2012;**118**:4372–84.

19 Youlden DR, Cramb SM, Dunn NA *et al.* The descriptive epidemiology of female breast cancer: an international comparison of screening, incidence, survival and mortality. *Cancer Epidemiol* 2012;**36**:237–48.

20 Ferlay J, Shin HR, Bray F *et al.* Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010;**127**: 2893–917.

21 Ferlay J, Soerjomataram I, Dikshit R *et al.* Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* 2015;**136**:E359–86.

22 Sankaranarayanan R, Swaminathan R, Brenner H *et al.* Cancer survival in Africa, Asia, and central America: a population-based study. *Lancet Oncol* 2010;**11**:165–73.

23 Brinton LA, Figueroa JD, Awuah B *et al.* Breast cancer in sub-Saharan Africa: opportunities for prevention. *Breast Cancer Res Treat* 2014;**144**:467–78.

24 Forouzanfar MH, Foreman KJ, Delossantos AM *et al.* Breast and cervical cancer in 187 countries between 1980 and 2010: a systematic analysis. *Lancet* 2011;**378**:1461–84.

25 Brinton LA, Sherman ME, Carreon JD, Anderson WF. Recent trends in breast cancer among younger women in the United States. *J Natl Cancer Inst* 2008;**100**:1643–8.

26 Anders CK, Johnson R, Litton J *et al.* Breast cancer before age 40 years. *Semin Oncol* 2009;**36**:237–49.

27 Jedy-Agba E, McCormack V, Adebamowo C, dos-Santos-Silva I. Stage at diagnosis of breast cancer in sub-Saharan Africa: a systematic review and meta-analysis. *Lancet Glob Health* 2016;**4**:e923–35.

28 Abdulrahman GO, Rahman GA. Epidemiology of breast cancer in Europe and Africa. *J Cancer Epidemiol* 2012;**2012**:915610.

29 Sariego J. Patterns of breast cancer presentation in the United States: does geography matter? *Am Surg* 2009;**75**:545–50.

30 Islami F, Lortet-Tieulent J, Okello C *et al.* Tumor size and stage of breast cancer in cote d'Ivoire and republic of Congo—results from population-based cancer registries. *Breast* 2015;**24**:713–7.

31 Clegg-Lamptey JN, Dakubo JC, Attobra YN. Psychosocial aspects of breast cancer treatment in Accra. *Ghana. East Afr Med J.* 2009;**86**(7):348–53.

32 Ezeome ER. Delays in presentation and treatment of breast cancer in Enugu. *Nigeria. Niger J Clin Pract* 2010;**13**(3):311–6.

33 De Ver DT, Bogale S, Hobden C *et al.* A mixed-method assessment of beliefs and practice around breast cancer in Ethiopia: implications for public health programming and cancer control. *Glob Public Health* 2011;**6**:719–31.

34 Harford JB. Breast-cancer early detection in low-income and middle-income countries: do what you can versus one size fits all. *Lancet Oncol* 2011;**12**:306–12.

35 Yip CH, Smith RA, Anderson BO *et al.* Breast health global initiative early detection panel. Guideline implementation for breast healthcare in low-and middle-income countries: early detection resource allocation. *Cancer* 2008;**113**:2244–56.

36 Pisani P, Parkin DM, Ngelangel C *et al.* Outcome of screening by clinical examination of the breast in a trial in the Philippines. *Int J Cancer* 2006;**118**:149–54.

37 Sankaranarayanan R, Ramadas K, Thara S *et al.* Clinical breast examination: preliminary results from a cluster randomized controlled trial in India. *J Natl Cancer Inst* 2011;**103**:1476–80.

38 Mittra I, Mishra GA, Singh S *et al.* A cluster randomized, controlled trial of breast and cervix cancer screening in Mumbai, India: methodology and interim results after three rounds of screening. *Int J Cancer* 2010;**126**:976–84.

Downloaded from https://academic.oup.com/jpubhealth/article-abstract/doi/10.1093/pubmed/fdaa099/5869901 by Delhi Technological University user on 20 July 2020

39 Smith RA, Saslow D, Sawyer KA *et al.* American Cancer Society guidelines for breast cancer screening: update. *CA Cancer J Clin* 2003;**53**:141–69.

40 Saslow D, Hannan J, Osuch J *et al.* Clinical breast examination: practical recommendations for optimizing performance and reporting. *CA: a cancer journal for clinicians* 2004;**54**:327–44.

41 Beral V, Bull D, Pirie K *et al.* Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. *Lancet Oncol* 2012;**13**:1141–51.

42 Colditz GA, Bohlke K. Priorities for the primary prevention of breast cancer. *CA Cancer J Clin* 2014;**64**:186–94.

43 Dicker D, Nguyen G, Abate D *et al.* Global, regional, and national age-sex-specific mortality and life expectancy, 1950–2017: a systematic analysis for the global burden of disease study 2017. *Lancet* 2018;**392**:1684–735.

44 Murray CJ, Callender CS, Kulikoff XR *et al.* Population and fertility by age and sex for 195 countries and territories, 1950–2017: a systematic analysis for the global burden of disease study 2017. *Lancet* 2018;**392**:1995–2051.

45 Kingham TP, Alatise OI, Vanderpuye V *et al.* Treatment of cancer in sub-Saharan Africa. *Lancet Oncol* 2013;**14**:e158–67.

46 Sullivan R, Alatise OI, Anderson BO *et al.* Global cancer surgery: delivering safe, affordable, and timely cancer surgery. *Lancet Oncol* 2015;**16**:1193–224.

47 Bray F, Ferlay J, Laversanne M *et al.* Cancer incidence in five continents: inclusion criteria, highlights from volume X and the global status of cancer registration. *Int J Cancer* 2015;**137**: 2060–71.

48 Zelle SG, Nyarko KM, Bosu WK *et al.* Costs, effects and cost-effectiveness of breast cancer control in Ghana. *Trop Med Int Health* 2012;**17**:1031–43.

Downloaded from https://academic.oup.com/jpubhealth/article-abstract/doi/10.1093/pubmed/fdaa099/5869901 by Delhi Technological University user on 20 July 2020

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-020-09199-5

Capsmf: a novel product recommender system using deep learning based text analysis model

Rahul Katarya[1] · Yamini Arora[2]

Received: 29 November 2019 / Revised: 19 April 2020 / Accepted: 8 June 2020
Published online: 29 June 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Researchers and data scientists have developed different Recommender System Algorithms such as Content-Based and Collaborative-Based in order to filter a large amount of information available on the internet and hence, recommend only the relevant and essential content based on the personalized interests of users. Information acquired explicitly by collecting users' ratings for an item lead to the problem of data sparsity. Many researchers have been currently working towards the improvement of rating prediction accuracy by integrating the auxiliary information along with the ratings provided by the users. This paper proposes a novel product recommender system called as "CapsMF", it applies the advanced neural network architecture Capsule Networks (Caps) for document representation, and MF represents Matrix factorization. In the proposed approach, we have enhanced a deep neural network text analysis model by adding a newly discovered neural network architecture; Capsule Networks stacked on bi-directional Recurrent Neural Network (Bi-RNN) for the robust representation of textual descriptions of items and users. The Deep Neural Network text analysis model is integrated with the Probabilistic Matrix Factorization to generate improved recommendations. The experiment has been performed on two real amazon datasets resulting in the enhancement of rating prediction accuracy, the recall, and the precision of top-n recommendations, in comparison to the basic and hybrid Recommendation System Algorithms. Also, text analysis model involving Capsule Networks stacked with Recurrent Neural Networks (RNNs) have outperformed the baseline models that have single Convolutional Neural Networks (CNN) or CNN combined with Bi-RNN in text analysis.

Keywords Recommender system · Matrix factorization · Capsule networks · Collaborative filtering

✉ Rahul Katarya
rahulkatarya@dtu.ac.in

Yamini Arora
arorayamini08@gmail.com

Extended author information available on the last page of the article

1 Introduction

In this digital era, people use e-commerce websites for shopping and social media sites for news and related information. The problem arises when there are varieties of options available while searching for a thing. It is tough for the end users to find the critical content they are looking for according to their interest. The solution for this new digital era issue is Recommender Systems (RS). These are being deployed on e-commerce websites and on the social media platforms that extract the personal information related to the users and recommend them the relevant content and information that they want depending on their interest, activity, and behavior. For example, movies, books, clothes, tweets are being recommended to the people visiting different web sites. Recommender System [22] is an application of Web Mining. This process discovers interests and behavior from large amounts of data without human interference and hence, ranks the most relevant content or items using the rating prediction of the users visiting a website. The researchers and data scientists that filter the information according to the personalized interests of the users have developed different types of Recommendation System Algorithms and techniques. A recommender system can extract information related to a user in two different ways: explicitly or implicitly.

One can get the information related to the users explicitly by collecting the ratings for an item given by a user, and it can be acquired implicitly by monitoring the behavior of users on the different websites related to the product [2]. Their behavior includes songs heard, applications downloaded, websites visited, movies watched, books read. Social information can be compiled by scraping the users' accounts on different social websites, for example, what kind of posts they like, their followers, followees and different activities on social media can tell a lot about the preference of the users, and hence, personalized recommendations can be more accurate and precise [15, 18].

Based on the information about different items and users, two categories of recommendation algorithms developed are Content-Based, Collaborative-based. Out of these, Matrix Factorization technique based on Collaborative Filtering (CF) gives the best recommendations [21]. The limitation observed in the current system is that information acquired explicitly by collecting users' ratings for an item lead to the problem of data sparsity because not every user rate every other movie. The accuracy of rating prediction for an item reduces, thereby not satisfying the needs of the customer. Many researchers inclined towards the improvement of rating prediction accuracy by integrating the auxiliary information along with the ratings provided by the users. It has been observed in related works that integrating the textual data along with rating data has brought an improvement in the rating prediction accuracy. However, document modeling approaches are different in different research papers.

Various topic modeling approaches and deep neural network models [36] are being used to represent the textual description of items and users for improving the rating prediction accuracy and recall and precision of top-n recommendations. The recent models integrate matrix factorization with textual data generate a textual representation of documents using deep neural network architectures; they are more effective as they extract semantic information from the text.

Capsule Networks bring an improvement in the old neural network architecture, and it works better than the Convolutional Neural Networks (CNN). There are certain disadvantages of using the convolutional neural networks, few areas where CNN lacks, Capsule Networks have overcome those limitations, and now it is a better neural network architecture for

developing models to solve the problems in different domains. Capsule Networks are primarily used for Classification of Images in different categories [8].

CN can also be applied in the areas of NLP and RS to utilize textual information more efficiently. Neural Networks can be trained to learn numeric representations of various words and phrases using models such as Word2Vec and Glove and can be applied to classify the data in different categories.

This research paper has successfully executed the following points:

We have addressed the limitations of CNN based document modeling approach as well as CNN and RNN based text representation approach.

We have proposed a novel and powerful Deep Neural Network text analysis model combining Capsule Networks and Bi-directional RNN to accomplish the textual representation of items' and users' reviews to improve recommendation performance.

Experiments results prove that exploiting ratings, items' and users' reviews have improved rating prediction accuracy, precision, and recall of top-n recommendations.

The rating Prediction accuracy on proposed model "CapsMF" has shown improvement over PMF, DRMF-Item, DRMF-User, and DRMF by 26%, 4.7%, 3.5%, and 1.7% in terms of MAE.

Different hybrid Recommendation Models using items' and users' reviews as auxiliary information have been described in the Related Work. The theory related to Capsule Networks and Bi-RNN and Recommendation algorithm, Probabilistic Matrix Factorization is explained in the Methodologies section. The proposed deep neural network text analysis model and the integration of PMF with a novel text analysis model are explained in the Proposed Approach section of the Paper. The last section of the paper deals with Experiment and Results along with Conclusion and Future Work associated with this approach.

2 Related work

Here, we have discussed different models that combine textual data with Recommendation System algorithms providing recommendations to the users. In [11], a brief about text classification using Capsule Networks(CN) has been explained. The different hybrid recommendation systems that have been proposed by the researchers that combine textual data and recommendation algorithms are explained further.

CTR, Collaborative Topic Regression [30], incorporates the advantages of collaborative-based, MF algorithm and probabilistic topic modeling approach for recommending research based articles to the scientists online. Content analysis is done using the Topic modeling approach (LDA) combined with latent factor models using MF for recommending the unseen articles [28]. Their approach worked well as it recommended completely unrated articles to the users that were useful and hence, predictions were made in the right manner. Various efficient algorithms have been proposed in further years that have worked upon these models and have improved them. In HFT [19] model, the authors have developed a statistical model that combines hidden dimensions in the rating matrix with the topics extracted from the text of the review given by users. This methodology helped in predicting users' rating on unseen movies and further recommends them. The models developed by them helped in the discovery of genre and identified the most informative reviews. Their model "HFT", Hidden Factor Topics has addressed the most occurred problem in recommendation tasks, Cold-Start. Their model performed better than the latent factor-based recommender system improving accuracy

by 5–10% [32]. The model RMR [17], proposed by the authors, have combined the content-based and collaborative based filtering techniques to develop a novel integrated recommender system. Their model "RMR", Rating meets reviews combined the rating model with a topic model to generate recommendations and thereby solving the "cold-start" problem, that is recommending the new products to the users that have not been rating using the description of the item. Their model also works on the side of the item. For modeling the ratings, they have used a mixture of Gaussian, unlike above researchers that have used MF method. They have compared their model with the above described, HFT and CTR and has shown that their model has performed better in terms of rating prediction accuracy. CDL [32], Collaborative Deep Learning brought a change in the topic modeling approach by learning the numeric representation of textual data by applying deep learning techniques on the item description document. In this paper, they have proposed the hierarchal Bayesian Model, which applies deep learning on the textual data and Collaborative filtering approach on rating data in a joint manner. They have used SDAE, Stacked Denoising Autoencoders, neural network approach combined with the Collaborative Filtering method in order to integrate the rating and content information, thereby advancing the quality of recommender systems. Convolutional Matrix Factorization (ConvMF) [14], this model addresses the rating data sparsity by learning the text document using single CNN. The CNN considers the surroundings of the word order that has improved the representation of latent features in the description of the items' documents. Their model integrates PMF and CNN to upgrade the rating prediction task and hence, refine the quality of recommendation of items or movies to the end-users. LZhang et al. [35] proposed combining CF algorithm with an artificial neural network (ANN) to enhance the scalability and remove the sparsity of recommendation systems. Quadric Polynomial Regression model extracts the latent features, and these features then becomes the input to the deep neural network for predicting the rating scores. The model used to extract the hidden features improves the traditional matrix factorization method [35]. Dual-Regularized Matrix Factorization (DRMF) [34], this is the recent research work in the field of integrating textual data description with PMF. They have exploited the textual description of items and users both and then predicted the rating score. They have also adopted a unique multi-layered neural network model that combines CNN and bi-directional GRU that has given a better latent representation of the text. The content representation regularized the latent models for the items and users in MF.

Different expert recommendation systems have also been designed to recommend the experts in different domains for solving problems using the information retrieval process. It helps in the detection of knowledgeable people expert in his or her area. The combination of search engines and NLP can be used to retrieve the experts in this kind of recommendation systems [12]. Different tags and their syntactic patterns associated with Web 2.0 such as audio, video, movies can be exploited for recommendation to solve the cold start problem [1, 23].

This research paper is being inspired by the above-related work explained. Our proposed approach brings an improvement in the text analysis model so that the textual representation gets better, thereby improving the quality of recommendations.

3 Why capsule networks over CNN?

Capsule Networks were newly discovered Neural Network Architecture by Geoffrey Hinton in 2017. They have been applied in various domains such as Image Classification, Natural Language Processing (NLP), text classification, etc. [31]. This is a breakthrough in the domain

of Deep Learning. It has overcome the various drawbacks in the CNN and shown effective results in the classification problems. The max pooling function used in CNN to extract the highest features is not efficient and results in reduced computational complexity. It has been replaced by Dynamic Routing Algorithm that is very efficient in recognizing the spatial relationships and hence improving the classification accuracy. "Routing-by-agreement" proposed by Geoffrey Hinton, helps in retaining the relevant information and features of the content and sends the features from lower-level to higher level only if the content is being matched [6].

CNN does not consider the spatial information, whereas CN does, and that increases its power. Replacing the CNN by CN in the deep neural network text analysis model has increased the rating prediction accuracy and hence reducing the root mean squared error (RMSE). The experiment and results are presented in the further sections.

4 Why deep neural networks for text analysis?

In recent times, the research in deep learning has enhanced. It is being used in different domains such as computer vision, image classification, NLP, Recommender Systems, etc. They have improved the accuracy in different classification and regression problems. They are more complex to understand and take more time in training, but the results generated are more effective and precise. They perform better in extracting semantics from documents and hence, generating an effective text representation of documents [9]. As we saw Capsule Networks perform better than CNN, as they have a dynamic routing algorithm that simultaneously takes feedback from simple entities.

Similarly, Bi-directional GRU is good at temporal modelling, that is, it stores information related to past and present. These neural networks are good in feature extraction and dimensionality reduction. Hence, we have proposed a DNN model for text analysis and generating the representation of items' reviews and users' reviews.

5 Methodologies

This section discusses the different methodologies adopted in order to implement a product recommender system.

6 Probabilistic matrix factorization

PMF [25] is collaborative filtering, model-based technique. The task of the algorithm is to infer the preferences of the users based on the ratings given by users. And what other similar users have liked? In this algorithm, the rating matrix of items and users are given, but it is a sparse matrix because not every user rate every other movie. The main aim is to fill the missing entries to predict the rating that a user would give to an unseen movie and then recommend those to the user based on predicted rating score.

PMF factorizes the rating matrix into a product of two smaller (low-rank) matrices. Given a rating matrix of M products and N users, and integer rating 1 to 5, the matrix is factorized into two latent feature matrices of users and items.

The two users and items latent feature-based matrices is learned from the observed rating matrix. PMF models ratings as drawn from a normal distribution. User and item features are determined using mean values. It assumes that ratings are normally distributed.

$$\widehat{r}_{ui} \sim N\left(p_u q_i^T, \sigma^2\right) \tag{1}$$

Where r_{ui} represents predicted rating score, N represents normal distribution, p_u represents user-feature vector and q_i represents item-feature vector.

σ^2 represents variance used to model noise of rating matrix.

User Feature Vector can be represented as:

$$p_{uf} \sim N\left(0, \sigma_p^2\right) \tag{2}$$

Here, p_{uf} represents user feature vector, N represents the normal distribution, σ_p^2 represents variance used to model noise of user latent feature vector matrix.

Item Feature Vector can be represented as:

$$q_{if} \sim N\left(0, \sigma_p^2\right) \tag{3}$$

Here, q_{if} represents item feature vector, N represents the normal distribution, σ_p^2 represents variance used to model noise of item latent feature vector matrix.

In this research, we have compared our model with the basic Probabilistic Matrix Factorization approach. In this proposed integrated model, parameters can be learned using the MAP approach.

7 Deep neural network for text representation

The Deep Neural Network (DNN) Text Analysis Model is being used for generating the text representation of both items and users. A unique and powerful DNN model has been implemented to learn the hidden features of textual reviews. The DNN Text Analysis Model first goes through three Capsule Network (CN) Layers and learns the text representation with these CNs. After learning the semantics and orientation of the text by concatenating three CN Layers, the text is encoded into numbers using Bi-directional Gated Recurrent Unit (Bi-GRU), a variation of RNN. The Framework of DNN Text Analysis Model consists of the Embedding Layer first, and then three CN Layers stacked on a Bi-directional GRU Layer, and finally the Output Layer that projects the hidden features of the reviews.

8 Embedding layer

The input data needs to be encoded in integer; the algorithm does not take strings as input and work on it. Here, the Embedding Layer in Keras is essential. It helps in providing a dense representation of words and learns the relationship between the words. In this research, we have used a pre-trained embedding (Glove) [20] to get the dense representation of the reviews

of items and users and learn the relationship between the words in the reviews. It helps in extracting the semantic relationship between the words. It encodes the meaning of words into vectors.

9 Capsule networks

Capsule Networks [24] are newly developed Deep Neural Network architecture introduced by Geoffrey Hinton in 2017. These networks brought an improvement in the Convolutional Neural Networks (CNN). Capsule Networks have overcome the limitations observed in CNN's. The main advantages of CNs that makes it different from CNN and more efficient are dynamic routing and equi-variance. Dynamic routing Algorithm [7] is the central concept behind capsule networks. Analogous to backpropagation in Artificial Neural Network, Dynamic Routing between capsules help the capsules in communicating with each other and creating the text representations [27].

10 Routing by agreement

A capsule network is built up by combining several layers. The Capsules present in the lower level of capsule network architecture correspond to simple entities that further combine to form an image. For example, in the image of a house (complex entity), the two smaller entities are rectangle and triangle. According to the concept of Routing by Agreement as specified by Geoffrey Hinton, the capsules at lower level bet on the presence of a complex entity at a higher level. Moreover, these bets are combined to get the output of capsules at a higher level [30]. For example, the presence of two simple entities, rectangle, and triangle, works with each other to bet on the presence of a complex entity at a higher level, i.e. house. A coupling effect is observed in this concept. Here, capsules at lower level try to predict the output of the capsule at the higher level and if the predictions made by capsules at lower level conform to the actual output of capsule at a higher level, the coupling coefficient between the two capsules increases.

11 Mathematics behind capsule networks

Consider two layers t and t + 1 having m and n capsules respectively. The main aim is to compute the activations of the capsules at layer t + 1 given the activations at lower layer t. Let u denotes the activations of capsules at layer t. The task is to calculate v, activations of capsules at layer t + 1 [7].

For a capsule j at layer t + 1,

12 Calculation of prediction vectors

The first step involves the calculation of the prediction vectors of the upper layer using the capsules at lower layer. The prediction vector is calculated by a capsule i (of layer t) for the capsule j (of layer t + 1):

$$\overline{U_{j\backslash i}} = W_{ij}U_i \tag{4}$$

Where Wij is the weight matrix

13 Calculation of the output vector

The output vector s_j is calculated using the weighted sum of all the prediction vectors given by the capsules of layer t for the capsule j [19].

$$s_j = \sum_{i=1}^{m} c_{ij}\overline{u_{j\backslash i}} \tag{5}$$

The scalar cij is the coupling coefficient between capsule i (of layer t) and j (of layer t + 1). The iterative dynamic routing algorithm determines the coupling coefficient.

14 Apply the squashing function

After calculating the output vector, the squashing function is applied to the output vector v_j of the capsule j:

$$\mathrm{v_j} = \mathrm{squash}\left(\mathrm{s_j}\right) \tag{6}$$

Squash Function performs the non-linear transformation and works as an activation function for capsule networks like ReLU function used in convolution networks.

15 Bi-directional gated recurrent unit

We have used Bi-Directional Gated Recurrent Unit stacked on Capsule Networks to enhance the text representation. Gated Recurrent Unit (GRU) solves the different problems observed in RNNs [29] such as long-term dependency and vanishing/ exploding gradient problem. These are the most effective neural networks when the context of the input is required. GRU uses two gates, update gate and reset gate in order to maintain the information passed long ago.

Mathematically, Update gate is represented using z_t for the time step t using the following equation,

$$z_t = \sigma\left(W^{(z)}x_t + U^{(z)}h_{t-1}\right) \tag{7}$$

Where x_t represents input unit, W (z) represents the weight, h_t-1 represents the information from previous input units and U (z) is the weight associated with the previous input units. They are both added, and the output is obtained between zero and one due to the sigmoid function applied.

The update gate reduces the vanishing gradient problem as it helps in determining the amount of past information required to be forwarded in the time ahead [5].

The Reset gate also plays an important role to forget the not so required past information. The equation for the reset gate is:

$$r_t = \sigma\left(W^{(r)}x_t + U^{(r)}h_{t-1}\right) \tag{8}$$

Where x_t represents input unit, W(r) represents weights associated with the input unit, h_t-1 represents information from previous input units and U(r) represents weights associated with the previous input units. The sigmoid function is applied after summing the results of the input unit and previous input units.

Now, the following formula shows the final output that would be generated using the reset gate that restores the past information and removes the irrelevant information from the past [25]. The new memory content generated using the Reset gate is given as:

$$h_t^{'} = \tanh(W_{x_t} + r_t \odot U h_{t-1}) \tag{9}$$

The h_t' represents the ongoing memory content. The present memory content is generated by multiplying the corresponding weights to the input x_t and multiplying the corresponding weight of previous input units. After that, the element-wise product is calculated between reset gates and weighted previous input units. Then, the non-linear activation function tanh is applied to generate the results. The closing Memory at the present time step t is calculated using the following equation:

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot h_t^{'} \tag{10}$$

In this equation, h_t represents the closing memory content vector generated, z_t represents the update gate and h'_t represents present memory content.

16 Output layer

After going through the embedding layer, Capsule Network Layer and bi-GRU Layers, the features extracted by the last layer are outlined on a k-latent dimensional space by using hyperbolic tanh activation function.

$$\alpha = \tanh\left(G^{*}h + b\right) \tag{11}$$

Here, G represents the projection matrix obtained in the output Layer, h represents the output obtained from the bi-directional GRU layer, and b is the bias. Dropout [26] is also used in the model to prevent the over-fitting in the projection matrix. The main aim behind Dropout is dropping hidden or visible neurons in a neural network model to prevent the over-fitting. Dropout is a regularization technique that helps in reducing the interdependence between the nodes in the neural network.

The above-mentioned text analysis neural network layers described converting a raw document into a densely distributed representation, as shown in Fig. 1. Given X, the document of items' reviews and Y, the documents of users' reviews, the DNN text analysis model transforms X to beta and Y to alpha. Alpha and beta are the densely distributed representation of the raw documents of users and items, respectively.

Keras functional API [4] is being used to define the neural network model for text analysis and getting the projection matrix for the further process. The model parameters are learnt by applying the back-propagation algorithm. The input sequence, output, and the corresponding loss function are defined for the model. In the Proposed Work section, the algorithm is being explained used for implementing the DNN text analysis model based on Keras.

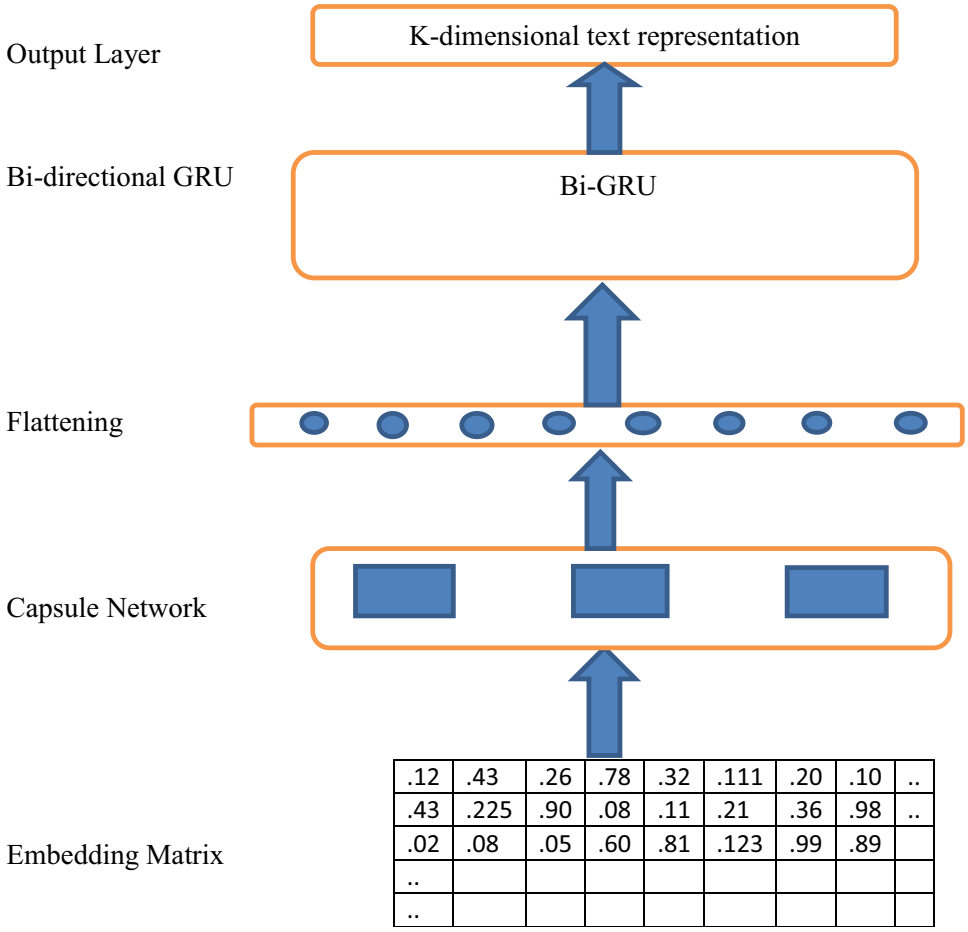


Fig. 1 Deep Neural Network text analysis model generating the representation of users' and items' reviews

17 Proposed work

In this segment, we have described the algorithm implemented for text analysis of reviews of items and users of different products. The matrix obtained from the text analysis is then integrated with the collaborative filtering based PMF method to generate more accurate and precise recommendations.

18 The model

The Algorithm gives the details about the implementation of Text Analysis Neural Network Model that involves different layers starting from Embedding Layer, Capsule Network Layers and Bi-directional GRU layers that extract features from the description documents and project them into k-dimensional space. It outputs a latent document k-dimension matrix that helps in improving the product recommendations.

The earlier similar models have used convolutional neural networks (CNN) [37] for latent document matrix that had certain limitations. These limitations have been overcome using capsule network layers.

The Algorithm is documented below, showing all the different layers that have been applied and implemented to get the k-dimensional matrix of the user and item reviews document. The document is first converted into a numerical matrix using the embedding layer representing the

text as a sequence of word embedding. Then, the document representation is learned through capsule networks, and the semantics of the reviews are encoded by going through the bi-directional GRU layers.

Algorithm: Deep neural network (dnn) text analysis model

```
Input: vocabulary size, e, d, s, gru_dim, k, init_W, n_capsule, n_routings, caps_dim;
Output: DNN text analysis Model
{Embedding Layers}
input_sequence = Input (shape= (d * s,))
seq_emb = Embedding (vocabulary size, e, weights= [init_W / 20], trainable=False) (input_sequence)
{Capsule Layers}
caps_1 = Capsule (num_capsule=n_capsule, dim_capsule = caps_dim,
routings = n_routings, share_weights=True) (seq_emb)
x = Flatten () (caps_1)
caps_2 = Capsule (num_capsule=n_capsule, dim_capsule=caps_dim,
    routings = n_routings, share_weights=True) (x_3)
x = Flatten () (caps_2)
caps_3 = Capsule (num_capsule=n_capsule, dim_capsule=caps_dim,
    routings=n_routings, share_weights=True) (caps_3)
x = Flatten () (caps_3)
x = concatenate ([caps_1, caps_2, caps_3])
{Bi-directional GRU Layers}
x_gru = Bidirectional (GRU (gru_dim, return_sequences=True)) (x)
x = Flatten () (x_gru)
{Output Layer}
seq_dropout = Dropout (dropout_rate) (x)
out_seq = Dense (output_dimesion, activation='tanh') (seq_dropout)
dnn_text_analysis = Model (input_sequence, out_seq)
dnn_text_analysis (optimizer='rmsprop', loss='mse')
return dnn_text_analysis;
```

19 CapsMF: integration of deep neural network text analysis model and PMF

Here, we describe the integration of PMF with the text analysis model and the equations involved in generating user latent feature vector and item latent feature vector. The main aim of the MF technique is to generate user and item hidden feature vectors [16] that particularly describe the choices of the users, and they are obtained from both the ratings and the reviews of items and users.

20 User latent vector

User i as a latent vector represented as:

$$p_i = €_i + \alpha_i \tag{12}$$

Where p_i represents user latent vector, $€ \sim N(0, \sigma_p^2 I)$, α_i represents user document latent vector.

Integrating the above we obtain the user latent model as:

$p_i \sim N(\alpha_i, \sigma_p^2)$, here I represent the identity matrix.

21 Item latent vector

Item j as a latent vector represented as:

$$q_j = €_j + \beta_j \tag{13}$$

Where q_j represents item latent vector, $€ \sim N(0, \sigma_q^2 I)$, β_j represents item document latent vector.

Integrating the above, we obtain item latent model as:

$q_j \sim N(\beta_j, \sigma_p^2 I)$, here I is the identity matrix.

22 Prediction

Suppose, M represents the number of users, and N represents the number of items, R represents the given rating matrix. The missing rating in the rating matrix is predicted using the formula "P^TQ", where P represents the user latent matrix of dimension k X M and Q represents the item latent matrix of dimension k X N.

$$\overline{R} = N\left(P_i^T Q_j, \sigma^2\right) \tag{14}$$

Next is the Flow Chart representing the Learning of different parameters in this model.

Figure 2 above describes the learning of different parameters used in the model "CapsMF" using Maximum a Posteriori (MAP) approach. First, rating matrix, item document, and user document are given as input to the model. Then, Item Content and User Content are pre-processed by removing the stop words and applying lemmatization and convert into embedding input using a pre-trained embedding (Glove) [20] to get the dense representation of the

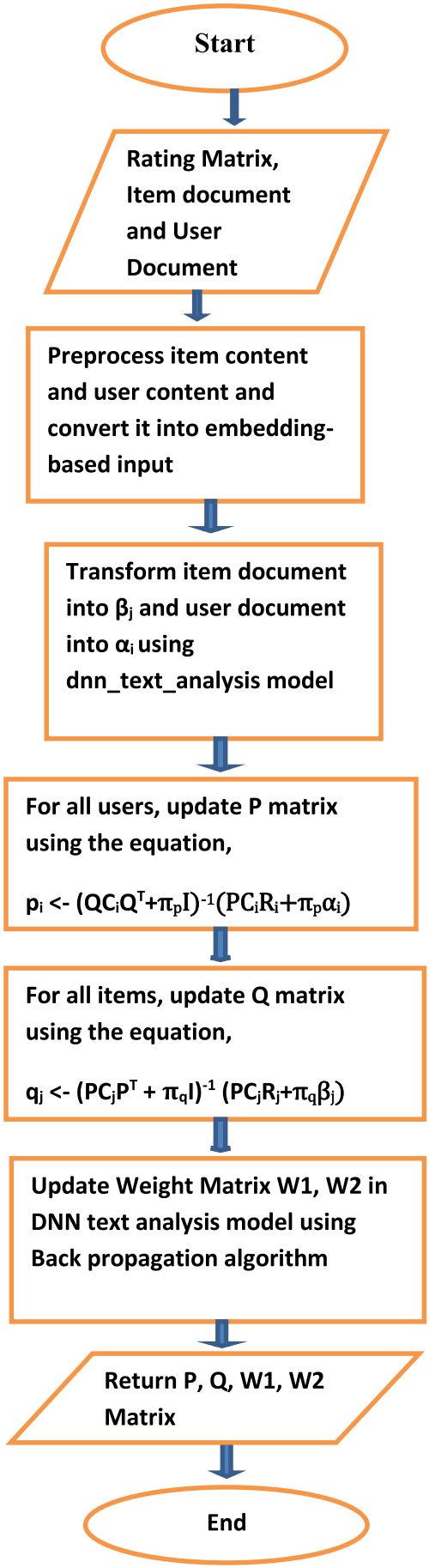


Fig. 2 Learning parameters in the proposed model CapsMF

Table 1 Description of the datasets

Datasets	Number of Users	Number of Items	Total Reviews/Ratings	Rating Scale
AIV (Amazon Instant Video)	5130	1685	37,126	1–5
AA (Apps for Android)	87,721	13,209	752,937	1–5

reviews of items and users. The document matrices generated for users and items and the text representation is learned through DNN text analysis model. Item document is transformed into β_j, and user document is transformed into α_i using the algorithm described in Section 6.1. After transforming the item reviews document and user reviews document, the user latent matrix and item latent matrix is updated using the equations given in the flow chart. The Weight Matrices W1 and W2 applied in DNN text analysis model are updated using Back Propagation Algorithm. When the number of iterations is over, return the User Latent Matrix, Item Latent Matrix, and Weight Matrices.

23 Experiment and results

We have conducted the experiments on the two amazon datasets, one is "Amazon Instant Video (AIV)", and the other is "Apps for Android (AA)". Both the datasets are taken from Amazon Product Reviews site (http://jmcauley.ucsd.edu/data/amazon/). This site contains reviews of different kind of products as well as metadata regarding the product. We have considered only the reviews and ratings of the products. We have taken 5-core datasets, in these 5-core datasets, the users and items have five reviews each. Table 1 describes the datasets that have been used for implementation. The Rating is given from one to five to an item by a user.

We have conducted the experiments on Google Colaboratory Notebook using GPU since the dataset is too large and deep learning models take time in training. We have used different evaluation metrics to compare the proposed model with the recommendation models already published. Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) have been extensively used for comparing the performance of rating predictions.

Table 2 Parameter settings of π_p and π_q for rating prediction

	Amazon Instant Video		Apps for Android	
Models	Π_p	Π_q	Π_p	Π_q
PMF	0.1	0.01	0.1	0.01
DRMF-Item	1	10	1	100
DRMF-User	10	0.1	100	0.1
DRMF	100	10	10	100
CapsMF	100	10	10	100

Where $\Pi p = \sigma^2 / \sigma_p^2$ and $\Pi q = \sigma^2 / \sigma_q^2$.

σ^2 represents variance used to model noise of rating matrix, σ_P^2 represents variance used to model noise of user latent feature vector matrix and, σ_P^2 represents variance used to model noise of user latent feature vector matrix.

Also, parameters related to Capsule Layer are set accordingly for "CapsMF":

$n_c = 50$; $n_r = 5$; $n_d = 100$

Here, n_c = number of capsules, n_r = number of routings and n_d = Capsule Dimension.

Table 3 Parameter settings of Π_p and Π_q for top-N recommendation

	Amazon Instant Video		Apps for Android	
Models	Π_p	Π_q	Π_p	Π_q
PMF	1	5	5	10
DRMF-Item	0.5	5	10	1
DRMF-User	5	0.5	10	1
DRMF	5	0.5	10	1
CapsMF	5	0.5	10	1

Where $\Pi p = \sigma^2 / \sigma_p^2$ and $\Pi q = \sigma^2 / \sigma_q^2$.

σ^2 represents variance used to model noise of rating matrix, σ_P^2 represents variance used to model noise of user latent feature vector matrix and, σ_P^2 represents variance used to model noise of user latent feature vector matrix.

Also, parameters related to Capsule Layer are set accordingly for "CapsMF":

$n_c = 50$; $n_r = 5$; $n_d = 100$

Here, n_c = number of capsules, n_r = number of routings and n_d = Capsule Dimension.

$N = 300$ for top-N recommendations.

We have also calculated the precision and recall for assessing the accuracy of top- n recommendations.

24 Different evaluation metrics

The formulas used for different evaluation metrics is described here:

$$\mathrm{MAE} = \sum_{i,j} \left| r_{ij} - \overline{r_{ij}} \right| / N \tag{15}$$

Where r_{ij} represents the actual rating, r'_{ij} represents the predicted rating and N represents the count of samples.

$$\mathrm{RMSE} = \text{square root} \left(\sum_{i,j} \left(r_{ij} - r'_{ij} \right)^2 \right) / N \Big) \tag{16}$$

Precision and Recall are calculated to assess the accuracy of top-N recommendations, we have calculated these metrics for top-300 recommendations given to the users.

$$\text{Precision@N} = (\#\text{of recommended items@N that are relevant}\} / (\#\text{of recommended items@N}) \tag{17}$$

Table 4 Rating prediction performance of different models on AIV dataset

Models	MAE	RMSE
PMF	0.9627	1.2088
DRMF-Item	0.7419	1.0101
DRMF-User	0.7325	0.9866
DRMF	0.7190	0.9661
CapsMF	0.7064	0.9593

Table 5 Rating prediction performance of different models on AA dataset

Models	MAE	RMSE
PMF	1.1829	1.4773
DRMF-Item	1.0664	1.3763
DRMF-User	0.9638	1.2617
DRMF	0.9071	1.1930
CapsMF	**0.8878**	**1.157**

The best values are shown in bold.

According to the accuracy comparisons shown in the tables above, it can be concluded that CapsMF, the model described has shown improvement in terms of MAE and RMSE on the two datasets: AIV and AA.

$$\text{Recall@N} = (\#\text{of recommended items@N that are relevant})/(\text{total}\#\text{of relevant items}) \quad (18)$$

25 Parameter settings

Different values that have been set for π_p and π_q are given in Tables 2 and 3.

26 Results

The different values of evaluation metrics obtained on implementing the different models are shown in this section. The baseline models are PMF, DRMF-Item, DRMF-User, and DRMF, as explained earlier in the related work. For the right results, we have averaged the results of five-folds in order to estimate the prediction and recommendation performance accurately.

Below Table 4 shows the results of rating prediction performance of different models on AIV dataset in term of MAE and RMSE.

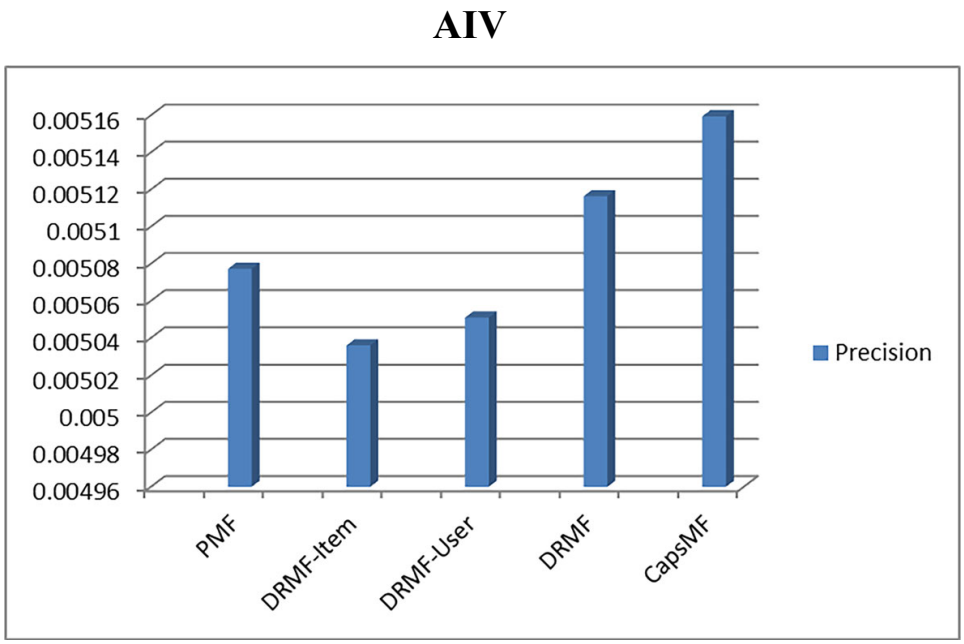


Fig. 3 Top-300 ranking performance in terms of precision

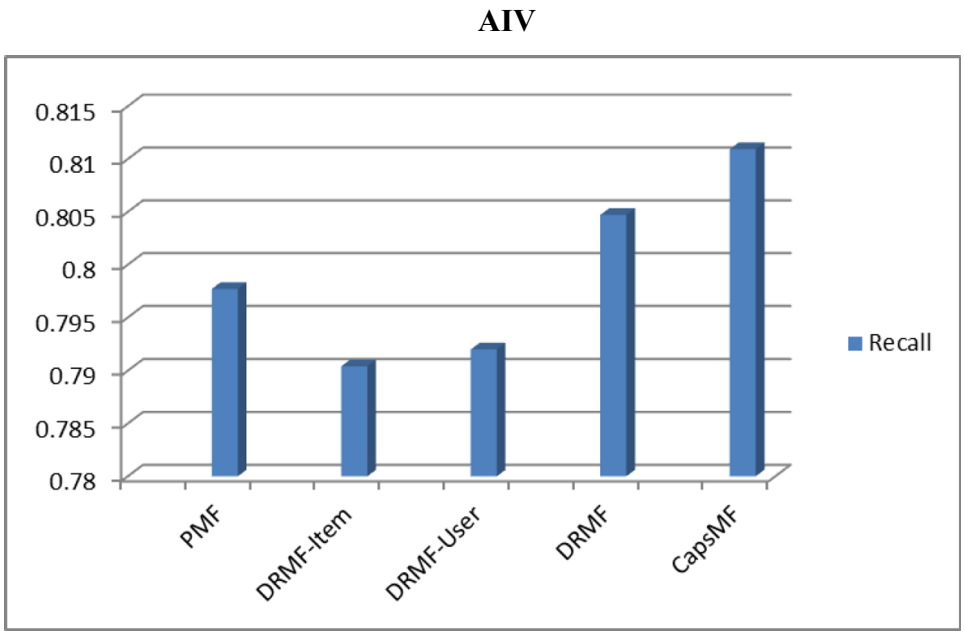


Fig. 4 Top-300 ranking performance in terms of recall

Table 5 shows the results obtained when rating prediction is implemented using different Models on AA dataset. Rating Prediction performance of different models is shown based on MAE and RMSE, respectively.

27 Graphical analysis of results

Here, we have shown the ranking performance of different Models on the two datasets AIV and AA graphically in term of Precision and Recall.

Below Figs. 3 and 4 graphically represent the top-300 recommendation performance of different Models on AIV dataset in term of precision and recall respectively.

Figures 5 and 6 show the results obtained while implementing the task of Top-N recommendations on AIV dataset in term of precision and recall, respectively. Here, the value of N varies from 50 and ends at 300. It is noteworthy that there is not much gain in top-n recommendation performance.

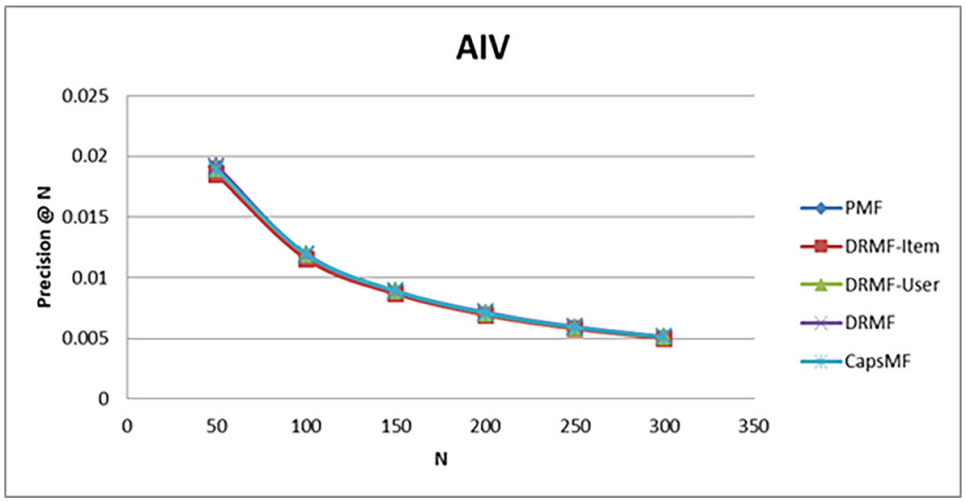


Fig. 5 Top – N recommendation in term of precision

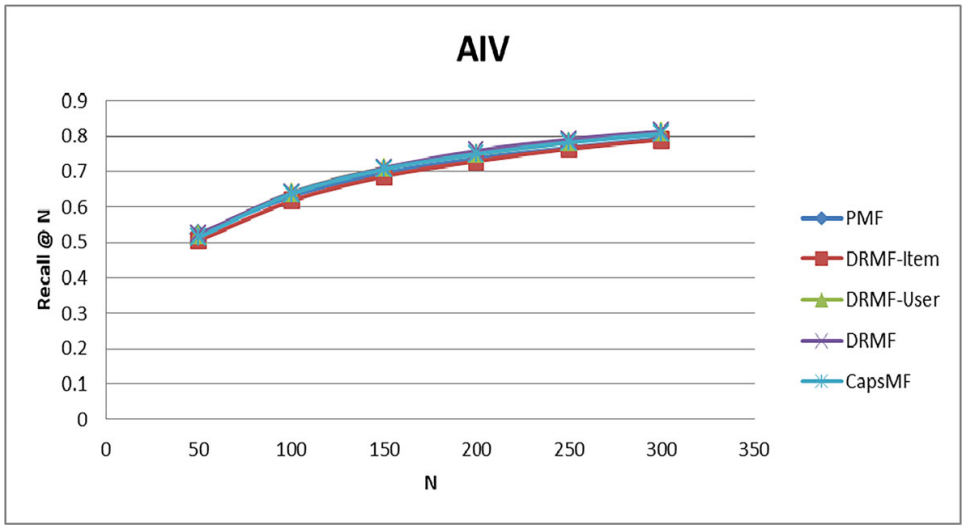


Fig. 6 Top-N ranking performance in term of recall

We can observe in the charts above that the proposed approach "CapsMF" has shown better recommendation performance on AIV dataset. Both RMSE and MAE have smaller values than the baseline Models implemented. All the models have been implemented on Google Colaboratory Notebook in Python using GPU for getting the results faster. Higher values of Precision and Recall prove that recommendation performance has improved over the baseline Models described in the related work.

Following are the graphical representations of the results obtained on "Apps for Android (AA)" dataset. Figure 7 shows the ranking performance of top-300 recommendations in term of Precision graphically on the AA dataset.

Figures 8 and 9 graphically represents the ranking performance of items in term of Precision and Recall on AA dataset. The value of N in Top-N recommendation performance varies from 50 to 300.

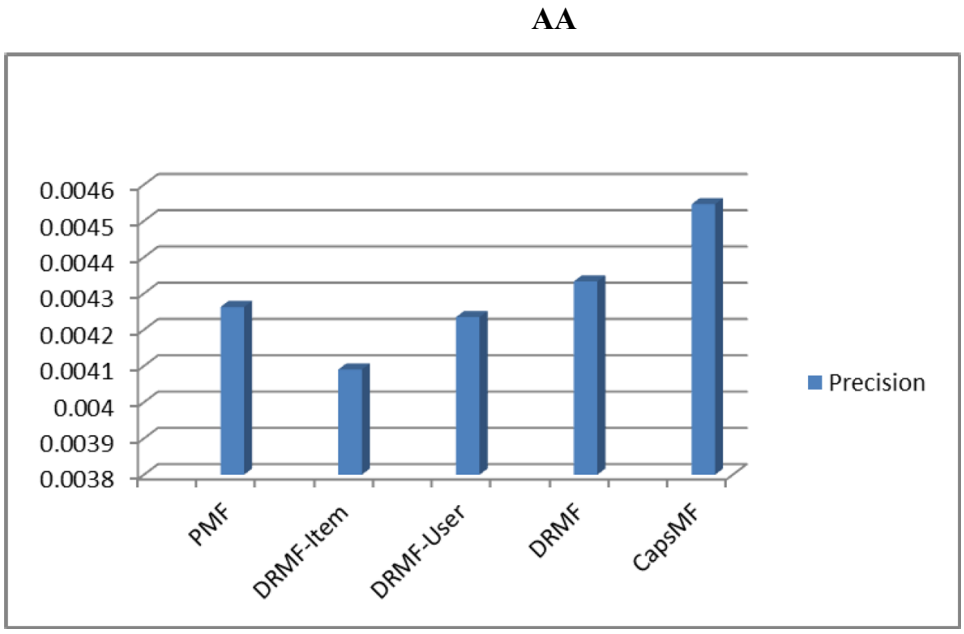


Fig. 7 Top-300 recommendation performance in terms of precision

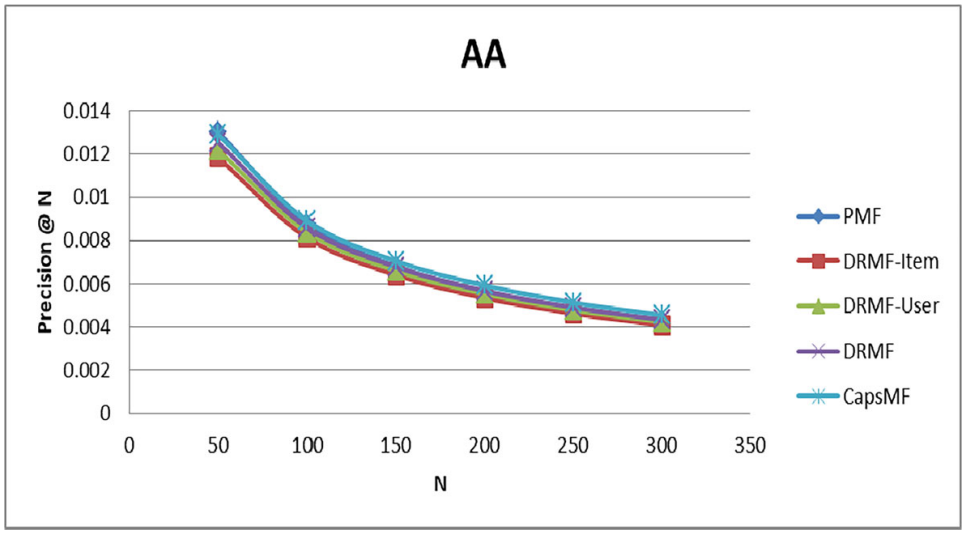


Fig. 8 Top-N recommendation performance in term of precision

28 Performance comparisons

Above we have shown the results in the tables and analysis of the different models on rating prediction and top-n recommendations using column charts. As we can observe that CapsMF, the proposed deep neural network model for text analysis has shown better performance when integrated with Matrix factorization (MF) Technique. For the task of rating prediction on AIV dataset, CapsMF has shown improvement over PMF, DRMF-Item, DRMF-Use, and DRMF by 26%, 4.7%, 3.5%, and 1.7% respectively in term of MAE. In term of RMSE, CapsMF has improved over PMF, DRMF-Item, DRMF-User, and DRMF by 20.6%, 5.02%, 2.76%, and 0.7% respectively.

For the task of top-n recommendation on AIV dataset, CapsMF has shown significant improvement when compared with the other models of similar type in terms of both precision and recall. We have chosen $n = 300$. It has improved by 1.6% when compared with PMF and 2.3%, 2%, and 0.8% improvement w.r.t DRMF-Item, DRMF-User and DRMF in term of Precision. Similar significant improvements have been observed in terms of Recall @ 300 on AIV Dataset.

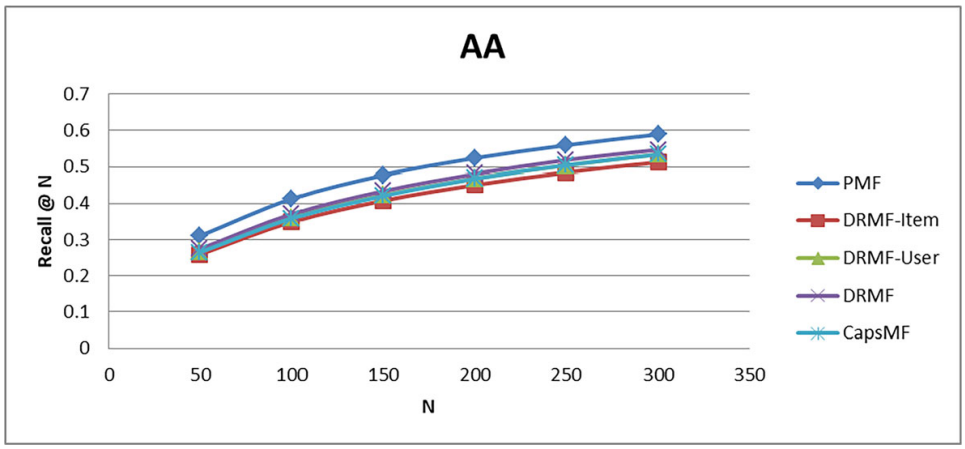


Fig. 9 Top-N ranking in term of recall similarly, when the proposed approach "CapsMF" is implemented on Apps for Android (AA) dataset has shown improvement over the baseline models on two different tasks, one of rating prediction and another, the ranking performance of items that are being recommended to the users

When the task of rating prediction and top-n recommendation are implemented on "Apps for Android (AA)" dataset, significant improvements have been recorded. CapsMF has shown better rating prediction accuracy over the baseline models. It has outperformed the baseline models PMF, DRMF-Item, DRMF-User, and DRMF by 25%, 16.7%, 7.8%, and 2.12% respectively in term of MAE. Similarly, RMSE values have also been reduced when calculated on the models by 21.6% when compared with PMF, 15.9%, 8.29%, 3.01% when compared with DRMF-Item, DRMF-Use, and DRMF respectively.

We can observe that the proposed model CapsMF has shown better results on a larger dataset, "Apps for Android (AA)".

29 Conclusion and future work

In this research paper, we have addressed the limitations of CNN based document modeling approach as well as CNN and RNN based text representation approach. We have proposed a novel and powerful DNN text analysis model stacking Capsule Networks and Bi-directional RNN to generate the textual representation of users' and items' reviews integrated with PMF to improve recommendation performance. Experiments results have proved that exploiting ratings, items' and users' reviews have improved rating prediction accuracy and precision and recall of top-n recommendations. It can also be concluded that the proposed model, "CapsMF", performs better on more massive datasets as compared to smaller datasets. Both the users and items reviews have been considered to build the separate deep neural network for text analysis. Hence, it has resulted in better recommendations to the users and also solves the cold-start problem in recommender systems.

There are still certain limitations that come with Capsule Networks; one that has been observed while performing the experiments is that capsule networks take times in training. The advantage is that it takes fewer data points in training as compared to CNN.

In future works, we would refine the deep neural network for text analysis with new deep learning technologies getting discovered for analysis of text for getting a better representation of the text. We would apply the proposed model in different recommendation datasets and scenarios such as social recommendations [3], joint-recommendation, group recommendation [33], etc. Other contextual information related to users and items such as images, tweets [10, 13], metadata about products can be integrated with collaborative based filtering technique to enhance the personalized recommendations.

Compliance with ethical standards

Conflict of interest There is "NO any" direct or indirectly related conflict of interest for this manuscript and with authors.

References

1. Belém FM, Heringer AG, Almeida JM, Gonçalves MA (2019) Exploiting syntactic and neighborhood attributes to address cold start in tag recommendation. Inf Process Manag 56(3):771–790
2. Bobadilla J, Ortega F, Hernando A, Gutiérrez A (2013) Knowledge-based systems recommender systems survey. Knowledge-Based Syst 46:109–132

3. Chen K, Chen T, Zheng G, Jin O, Yao E, Yu Y (2012) Collaborative personalized tweet recommendation. Proc. 35th Int. ACM SIGIR Conf Res Dev Inf Retr - SIGIR '12, no 800, p 661
4. Chollet F et al. Keras: deep learning library for theano and tensorflow. https://keras.io/
5. Chung J (2014) Gated recurrent neural networks on sequence modeling. arXiv : 1412. 3555v1 [cs . NE], pp 1–9
6. Hinton GE, Krizhevsky A, Wang SD (2011) Transforming auto-encoders. In: Honkela T, Duch W, Girolami M, Kaski S (eds) Artificial neural networks and machine learning – ICANN 2011. Lecture notes in computer science, vol 6791. Springer, Berlin, Heidelberg
7. Hinton G, Sabour S, Frosst N (2018) Matrix capsules with EM routing. International Conference on Learning Representations (ICLR), pp 1–15
8. Jaiswal A, AbdAlmageed W, Natarajan P (2018) CapsuleGAN: generative adversarial capsule network. Cite arxiv:1802.06167 pp 1–10
9. Johnson R, Zhang T Effective use of word order for text categorization with convolutional neural networks. In Proceedings of the 2015 Conference of the North {A}merican Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 103–112. https://doi.org/10.3115/v1/N15-1011
10. Katarya R, Arora Y (2018) A survey of recommendation systems in twitter. 2018 4th international conference on Computational Intelligence & Communication Technology (CICT), Ghaziabad, pp 1–5. https://doi.org/10.1109/CIACT.2018.8480264
11. Katarya R, Arora Y (2019) Study on text classification using capsule networks. Accepted in International Conference on Advanced Computing & Communication Systems (ICACCS), IEEE
12. Khasmakhi NN, Balafar MA, Derakhshi MRF (2019) Engineering Applications of Artificial Intelligence The state-of-the-art in expert recommendation systems ☆. *Eng. Appl. Artif. Intell.* 82(August 2018):126–147
13. Khater S, Elmongui HG, Gra D (2014) Tweets you like: personalized tweets recommendation based on dynamic users interests. IEEE, pp 1–10
14. Kim D, Park C, Oh J, Lee S, Yu H (2016) Convolutional matrix factorization for document context-aware recommendation. Proceedings of the 10th ACM Conference on Recommender Systems, page 233–240. New York, NY, USA, ACM
15. Kong X, Member S (2018) A survey of collaborative filtering-based recommender systems : From traditional models to hybrid models based on social networks. IEEE Access 6:64301–64320
16. Koren Y, Bell R, Volinsky C (2009) Matrix Factorization Techniques for Recommender Systems. Computer 42(8):30–37. https://doi.org/10.1109/MC.2009.263
17. Ling G, Lyu MR, King I (2014) Ratings meet reviews, a combined approach to recommend. In Proceedings of the 8th ACM Conference on Recommender systems (RecSys '14). ACM, New York, NY, USA, 105–112. https://doi.org/10.1145/2645710.2645728
18. Liphoto M, Du C, Ngwira S (2016) A survey on recommender systems. 2016 International Conference on Advances in Computing and Communication Engineering (ICACCE), Durban, pp 276–280. https://doi.org/10.1109/ICACCE.2016.807376
19. Mcauley J (2013) Hidden factors and hidden topics : understanding rating dimensions with review text. Proceedings of the 7th ACM Conference on Recommender Systems, RecSys '13, pages 165–172, ACM
20. Pennington JJ, Socher R, Manning CD (2014) GloVe : global vectors for word representation. Empirical Models in Natural Language Processing (EMNLP), pp 1532–1543
21. Ramlatchan A, Yang M, Liu Q, Li M, Wang J, Li Y (2018) A survey of matrix completion models for recommendation systems. In Big Data Mining and Analytics, vol 1, no 4, pp 308–323(IEEE), December. https://doi.org/10.26599/BDMA.2018.9020008
22. Razghandi M, Golpaygani SAH (2017) A context-aware and user behavior-based recommender system with regarding social network analysis. 2017 IEEE 14th Int Conf E-bus Eng pp 208–213
23. Rojo D, Valdiviezo-díaz P (2018) Hybrid collaborative filtering based on users rating behavior. IEEE Access vol 6
24. Sabour S, Nov CV, Hinton GE (2017) Dynamic routing between capsules. arXiv:1710.09829, Nips
25. Salakhutdinov R, Mnih A (2007) Probabilistic matrix factorization. In Proceedings of the 20th International Conference on Neural Information Processing Systems (NIPS'07), Curran Associates Inc., USA, 1257–1264
26. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: A simple way to prevent neural networks from overfitting. J Mach Learn Res 15(1):1929–1958
27. Srivastava S, Khurana P, Tewari V (2018) Identifying aggression and toxicity in comments using capsule network. Proceedings of the First Workshop on Trolling, Aggression, and Cyberbullying (TRAC-2018), pp 98–105

28. Steyvers M, Smyth P, Rosen-Zvi M, Griffiths T (2004) Probabilistic author-topic models for information discovery. Proc. 2004 ACM SIGKDD Int Conf Knowl Discov Data Min. - KDD '04, no. 1990, p 306
29. Van Merri B, Fellow CS (2014) Learning phrase representations using RNN encoder – decoder for statistical machine translation. In Proceedings of the 2014 Conference on empirical models in natural language processing (EMNLP), Association for Computational Linguistics, pp 1724–1734, October
30. Wang C, Blei DM (2011) Collaborative topic modeling for recommending scientific articles. Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, page 448–456, New York, NY, USA, ACM
31. Wang Y, Sun A, Han J (2018) Sentiment analysis by capsules. WWW '18 Proceedings of the 2018 World Wide Web Conference, Pages 1165–1174 vol 2
32. Wang H, Wang N (2015) Collaborative deep learning for recommender systems. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 1235–1244. https://doi.org/10.1145/2783258.2783273
33. Wang W, Zhang G, Lu J (2016) Member contribution-based group recommender system. Decis Support Syst 87:80–93
34. Wu H, Zhang Z, Yue K, Zhang B, He J, Sun L (2018) Knowledge-based systems dual-regularized matrix factorization with deep neural networks for recommender systems R. Knowledge-Based Syst. 145:46–58
35. Zhang L, Luo T, Zhang F, Wu Y (2018) A recommendation model based on deep neural network. IEEE Access 6:9454–9463. https://doi.org/10.1109/ACCESS.2018.2789866
36. Zhang S, Yao L, Sun A, Tay Y (2019) Deep Learning Based Recommender System: A Survey and New Perspectives. ACM Comput. Surv. 52:5:1–5:38. https://doi.org/10.1145/3285029
37. Zheng L, Noroozi V, Yu PS (2017) Joint deep modeling of users and items using reviews for recommendation. In proceedings of the tenth ACM international conference on web search and data mining (WSDM '17). ACM, New York, NY, USA, 425–434. https://doi.org/10.1145/3018661.3018665.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Rahul Katarya[1] · **Yamini Arora**[2]

[1] Department of Computer Science and Engineering, Delhi Technological University, Delhi, India

[2] Department of Information Technology, Delhi Technological, Delhi, India

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Citation-Based Scientific Paper Summarization Using Game Theory

Anil Singh Parihar[1], Anjali Jain[2], Archit Gupta[3]
Department of Computer Science and Engineering
Delhi Technological University
New Delhi, India
[1]parihar.anil@gmail.com, [2]anjali7230@gmail.com, [3]archit.g1997@gmail.com

***Abstract*—Given the vast amount of research-oriented literature, any new research work asks for a comprehensive literature review. This process is typically very time-consuming. Citation information of a paper is a reflection of its contributions to the researcher's group. We use the citations of the paper to generate a comprehensible and informative summary, which will reduce the time taken to review the paper. In this paper, we propose a novel evolutionary game-theoretic approach to summarize scientific papers. A set of two-player games is played to choose the most relevant sentences in the paper that should be a part of the summary. Our method doesn't require an annotated dataset with gold summaries, thus levitating the need for a large, manually created dataset, which would be a primary requirement for any supervised approach. We have evaluated our results against human-expert generated summaries and state-of-the-art methods.**

Keywords—Extraction based text summarization, game theory

I. INTRODUCTION

Scientific articles are being published at a rate faster than ever. Whenever someone tries to enter a new field of research, they have to go through a large amount of literature to become familiar with the current progress achieved in that area. Reducing the complexity of this task is the main motive behind developing automatic summarizers for scientific papers. Scientific paper summarization is divergent from summarization of general articles [1] as these are divided into different sections, involve complex technical terms, and are much longer as compared to general articles. As of now, Scientific paper summarization follows two strategies. One of them uses the paper's abstract. Now, while the authors' perspective of the contents of a paper is concisely reflected in its abstract, it is also essential to include the views of the scientific community and the impact the paper has had on it. As a remedy to this problem, the second strategy, i.e., citation-based summaries, was proposed [2]. The citation summaries are formed by using a set of citations of an article.

We propose an approach that combines both these strategies. Also, instead of using citation sentences directly (as they may not contain enough relevant information or are even inaccurate sometimes), we follow the methodology of [3] , by making use of cited text spans, i.e., A collection of sentences in the cited paper that is referred to, by the citing sentences from other papers, for generating summaries. Our model takes a reference paper to be summarised and constructs its summary by aggregating and integrating contents from the abstract, and the cited text spans using game theoretical techniques. Hence it includes both the author's vision and the scientific communities' perspective.

We use the SciSummNet corpus as described in [4] as our dataset. It provides over 1,000 papers in the ACL anthology network with their citation networks (e.g., citation sentences, citation counts) and their comprehensive, manual summaries. For evaluation, we use the benchmark gold summaries provided in the CL-SciSumm shared task [5] . The evaluations show that our method outperforms the existing state-of-the-art methods and provides summaries which inculcate the outlook of the author and the research community.

II. RELATED WORK

Early works in scientific article summarization [6, 7] , produced the summaries of target papers based only on their contents. [1] formulated these summaries by identifying suitable content from the papers using a naïve Bayes classifier. But, the discourse of the paper didn't provide any information about the impact of the paper on the scientific community and their views on it. As a solution to this problem, citation-based summarization was proposed. A summary of an article created by using a set of sentences in other articles that cite that article is called its citation summary. According to [8] , citation summaries, while overlapping with the abstract of an article, also focus on other aspects and provide additional information that is not contained in the abstract of an article.

Further research deemed the direct use of citation sentences for summarization as inaccurate as they may not provide enough information and sometimes may even be imprecise. So, [3, 9] proposed identifying the relevant context (usually a set of sentences) in the reference paper (original paper) for each citation and using that for summarization. Such a set of sentences is usually called a 'text

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:00:35 UTC from IEEE Xplore. Restrictions apply.

span,' and in this case, it is known as cited-text span. The summaries formed using these cited-text spans would include the insights of the research community while being composed of discourse from a stable source (only a single paper) and, thus, be more comprehensible and coherent. [10] performed experiments that show that summaries formed by using cited text spans outperformed those formed by using the citation sentences directly. CL-SciSumm shared task [5] also uses the cited text-spans based approach.

III. PRELIMINARIES

Skip thoughts, [11] is a model consisting of an encoder based on a recurrent neural network that is used to encode input sentences to a given size output feature vector. This output is used as an input to two decoders. One of the two generates the previous sentence from the context of this vector, and the other one generates the next sentence of the document given the context of the input vector. Outputs from the two decoders are used to train the encoder. Skip thoughts network is trained on the dataset of reference papers whose summaries need to be generated to convert a given sentence of the paper into a fixed-size feature vector.

In classic game theory, we define a set of players, player ={1,2…..n}and every player has a corresponding set of strategies, $strategy_i$ = {1,2…..m}. The payoff is the reward a player earns by playing a strategy in a particular game setting. Utility function relates strategies chosen by all players to the payoff, mathematically defined as, $U=strategy_i \times \ldots \times strategy_j$. To play the game, each player selects a strategy, and the utility function depends on the combination of strategies played simultaneously by all the players involved rather than one player. A crucial assumption being that all the players are rational, and their essential goal is to boost the estimation of U. In non-cooperative games, players first consider what the other people can play, and then choose an ideal strategy.

Nash equilibria can be defined as the stable state of a system where changing the strategy of only one player while keeping the strategies of other players the same gives incentives to other players. In other words, it is the state of a system when each player gets the highest payoff at the given strategy spaces of the complete game. Furthermore, most games involve mixed strategies for players, which are probability distributions over pure strategies. Consider a set of pure strategies, $strategy_i$ = {1,2….m}. In a mixed strategy setting, each pure strategy is assigned a probability, $X = \{x_1, x_2 \ldots .. x_m\}$ where each component x_i is the probability of a player selecting its ith strategy.

In a game between two players, the payoff of the first player from a strategy i whose probability in mixed strategy setting is p concerning the strategy j of the other player whose probability in mixed strategy setting is q will be $u_{ij} = p.A_1.q$ where A_1 is the payoff matrix of the first player.

Evolutionary game hypothesis veers off from the old-style idea as in a player can play the game with all the co-players winning a halfway result after each game and afterward acquires the complete total of every single fractional result as the last result. In this hypothesis, the players over and over take part in games and update their methodology space at every emphasis, considering the adjustments granted during past games. The payoff of a pure strategy h for player i in this setting is given as,

$$u_i\left(e_i^h\right)=\Sigma_{j=1}^{n}\left(A_{ij}x_j\right)_h \quad (1)$$

where x_k denotes the mixed strategy space of player k and A_{ij} is the payoff matrix between player i and j. The average payoff for this player is evaluated as,

$$u_i(x)=\Sigma_{j=1}^{n} x_i^T A_{ij} x_j \quad (2)$$

The probabilities in the mixed strategy space of each player are refreshed by the replicator dynamic condition [12] to accomplish Nash equilibria.

$$\dot{x}=\left[u\left(e^h,x\right)-u(x,x)\right].x^h \ \forall h \in x \quad (3)$$

We have utilized the discrete rendition of this condition, in our work [13] in our work. At each time step, the strategy space is updated following this equation until the system attains Nash equilibrium.

$$x^h(t+1)=x^h\frac{u\left(e^h,x\right)}{u(x,x)} \quad (4)$$

IV. PROPOSED WORK

We took the concept of evolutionary game theory and applied it to the task of text summarization. We take abstract sentences and the citation sentences of the reference paper as the players of the game. Each player plays a game against all the other players, similar to an evolutionary game setting. Weightage of a game in updating the strategy space of a player depends on the similarity between the two players (or sentences). We build a weight matrix W as,

$$w_{ij}=0.7*\text{similarity}\left(\text{player}_i, \text{player}_j\right)+0.3*\text{auth_score}_{(i,j)} \quad (5)$$

where $player_i$ & $player_j$ denote players i & j respectively and $auth_score_{(i,j)}$ is relative authority score of the paper which contains $player_i$ to the authority score of the paper which contains $player_j$ and similarity is cosine similarity between vectors of the respective players. The relative weightage of these two scores for which this method works best has been calculated through experimentation. This set of weights gives the best results. We then convolve the W matrix with a Gaussian filter to smooth it. The set of strategies of each player is a set of sentences similar to the player sentence in the reference paper. The size of the strategy space of each player is kept the same. We initialize the probabilities equal in the mixed strategy setting. We define the payoff matrix A_{ij}

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:00:35 UTC from IEEE Xplore. Restrictions apply.

between players i and j where similarity is the cosine distance between the vectors of $strategy_{ik}$ and $strategy_{jl}$ as,

$$a_{ij}^{kl} = similarity\left(strategy_{ik}, strategy_{jl}\right) \quad (5)$$

We follow the aforementioned discrete dynamic equation (4) for updating probability to attain Nash equilibria in the above system where the payoff of the h^{th} pure strategy of $player_i$ is defined as,

$$u_i\left(e_i^h\right) = \Sigma_{j=1}^{n}\left(w_{ij} A_{ij} x_j\right)_h \quad (6)$$

and the average payoff of $player_i$ is defined as,

$$u_i(x) = \Sigma_{j=1}^{n} x_i^T (w_{ij} A_{ij} x_j) \quad (7)$$

We can terminate updating the probabilities once Nash equilibrium is attained that is the state of the game becomes stable or after a fixed number of iterations.

Our method finds the most suitable representation of abstract and citation sentences in the paper and scores them in order of their relevance in a summary.

For each player, the best strategy that is the one having maximum probability in the player's mixed strategy is chosen as the best representative of that player. Then all these representative strategies are compared together and sorted in the decreasing order of the average payoff of their respective player. The sentences are chosen to be a part of the summary in the priority of their sorted order and taking care that similar sentences are not being added to the summary. The sentences chosen for the summary by the above method are then arranged in the order in which they appear in the reference paper, and then the summary can be reduced to 200 words by omitting last sentences as later parts of a paper are very specific for a summary compared to sentences at the beginning of the paper.

The intuition behind this being that for a player, the strategy which is coherent with all other strategies of all other players, will have a better pure strategy payoff than the rest of the other strategies of that player, which will lead to a gain in its probability, increasing its probability from the same initial value in the mixed strategy of that player. Thus, the strategy with maximum probability can be taken as the representative of that player. It is the most suitable sentence in the reference paper, which is similar to the player and also coherent with the other sentences of the summary. The importance of the representative strategy, in summary, will be indicated by the average payoff of the respective player as more will be the average payoff of a player more relevant will be its strategy to the summary.

V. Experimentation and Results

We have used human-annotated, Scisumm dataset [2], ScisummNet which constitutes of 1000 papers in the domain of ACL anthology network along with their human expert generated summaries to test our model.

We trained the skip thoughts network with the reference papers of the dataset and used the cosine distance between skip thought vectors as a similarity measure to create strategy space for each player. For building the weight matrix and payoff matrix, we used a weighted score of cosine distance between skip thoughts vectors and tf-idf vectors of the sentences as an altered cosine similarity measure. We used a weight of 0.8 for skip thought vectors similarity and 0.2 for tf-idf vectors similarity, the values of these weights were determined experimentally. We then convolved the weight matrix with a Gaussian filter to smooth it. We followed the system dynamics for the above-mentioned evolutionary game theory model to generate summaries.

An overall framework of our model is shown in Fig. 1. Abstract and citation sentences of the reference paper are players of the game. Tf-idf and skip-thoughts vector encodings are used to create the strategy space of each player and for building payoff and weight matrices. Evolutionary game dynamics are followed to get a suitable representative of every player and to rank those representatives in order of their importance in the summary. Dissimilar sentences are chosen in the order they have been ranked in. This set of sentences is arranged in the order in which they appear in paper to generate the summary.

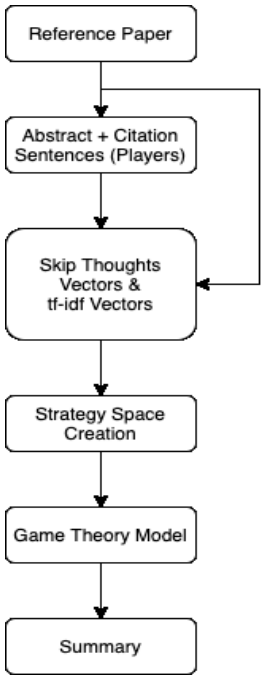


Fig. 1. Proposed Framework.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:00:35 UTC from IEEE Xplore. Restrictions apply.

In Fig. 2, we present an abstract of a reference paper from the above dataset on the topic "Unsupervised Semantic parsing". The summary generated by our method is shown in Fig. 4.

We took two players from the same reference paper whose abstract is cited in Fig. 3, one being a relevant sentence("We evaluate our approach by using it to extract a knowledge base from biomedical abstracts and answer questions.") to the summary and the other being a sentence("The sentence can be rephrased as "Utah is next to Idaho," "Utah shares a border with Idaho," etc.") that is too specific for a summary and doesn't add much information to it and show the contrast in the curve of average payoff between a coherent, informative player that is integral to a summary and a player that is not important concerning a summary. The obtained graph (Fig. 3) is in concurrence with our intuition.

"We present the first unsupervised approach to the problem of learning a semantic parser, using Markov logic. Our USP system transforms dependency trees into quasi-logical forms, recursively induces lambda forms from these, and clusters them to abstract away syntactic variations of the same meaning. The MAP semantic parse of a sentence is obtained by recursively assigning its parts to lambda-form clusters and composing them. We evaluate our approach by using it to extract a knowledge base from biomedical abstracts and answer questions. USP substantially outperforms Text Runner, DIRT and an informed baseline on both precision and recall on this task."

Fig. 2. Abstract of reference paper on the topic "Unsupervised Semantic parsing".

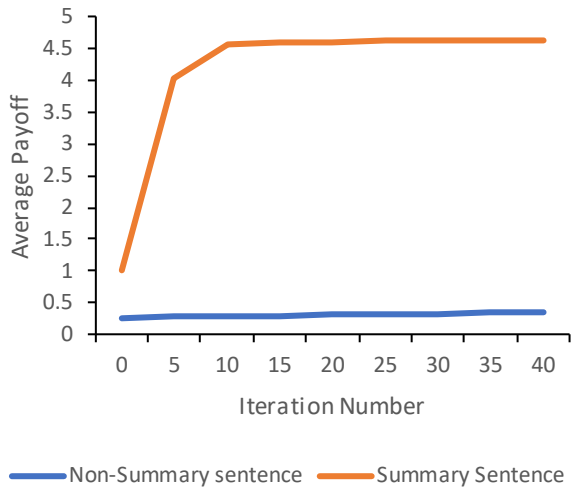


Fig. 3. Average payoff vs. iteration number graph between a summary and a non-summary sentence.

"Our USP system transforms dependency trees into quasi-logical forms, recursively induces lambda forms from these, and clusters them to abstract away syntactic variations of the same meaning. We evaluate our approach by using it to extract a knowledge base from biomedical abstracts and answer questions. In this paper we develop the first unsupervised approach to semantic parsing, using Markov logic. The semantic parse of a sentence is derived by starting with logical forms in the lexical entries and recursively composing the meaning of larger fragments from their parts. Markov logic makes it possible to compactly specify probability distributions over complex relational domains, and has been successfully applied to unsupervised coreference resolution and other tasks.To address this problem, we impose local normalization constraints on specific groups of formulas that are mutually exclusive and exhaustive, i.e., in each group, we require that Eki=1 ewi = 1, where wi are the weights of formulas in the group.6Regularizations, e.g., Gaussian priors on weights, alleviate this problem by penalizing large weights, but it remains true that weights within a short range are roughly equivalent. and update agenda and candidate operations until agenda is empty return the MLN with learned weights and the semantic parses Another major challenge in USP learning is the summation in the likelihood, which is over all possible semantic parses for a given dependency tree."

Fig. 4. Summary generated by our method.

We have evaluated the results, both qualitatively and quantitatively.

A. Quantitative Measure

We have used ROGUE metrics (Recall-Oriented Understudy for Gisting Evaluation) [14] for quantitative evaluation. ROGUE-2 estimates recall and F1 Measure for the bigram overlap between gold and generated summary. We gather from Table 1 that our approach performs as good as the GCN model of the ScisummNet [2] generated summaries, which is the benchmark for this dataset.

TABLE I. ROGUE-2 SCORES

Summarizer	Recall	F1 Measure
Abstract	28.52	29.40
Our Approach	34.61	32.73

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:00:35 UTC from IEEE Xplore. Restrictions apply.

Baseline	33.88	31.54

B. *Qualitative Measure*

We gathered a group of 50 people and asked them to score summaries generated by our method on a scale of 0-10 on the parameters of coherence and covering the paper's citation information. Our summaries got an average score of 8.2, which symbolizes that the method generates a comprehensible summary that represents citation information adequately.

VI. CONCLUSION

Citation information of a paper is an outlook of its influence on the researcher's community; thus, appending citation information to a paper's summary is crucial to a faster and comprehensive reviewing of the paper and deciding its relevance in one's work. Our method proposes an unsupervised game-theoretic approach to the task of citation-based text summarization, which levitates the need of a large human-annotated dataset. Our experiments show that our method generates better and comprehensive summaries than the abstract of the paper and the previous citation-based text summarization work. Additional features can be added to bias the weight and payoff matrices. Generic sentence embedding approaches with the semantic representation of the input will lead to better results. We hope that our game-theoretic approach to this task contributes to more novel work in this domain.

REFERENCES

[1] S. Teufel and M. Moens, "Summarizing Scientific Articles: Experiments with Relevance and Rhetorical Status," Computational Linguistics, vol. 28, no. 4, pp. 409–445, 2002.

[2] V. Qazvinian and D. Radev, "Scientific Paper Summarization Using Citation Summary Networks", in Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008), pp. 689–696, ACL, 2008

[3] A. Cohan and N. Goharian, "Scientific document summarization via citation contextualization and scientific discourse", International Journal on Digital Libraries, vol. 19, no. 2-3, pp. 287–303, Sep. 2017.

[4] M. Yasunaga, J. Kasai, R. Zhang, A. R. Fabbri, I. Li, D. Friedman, and D. R. Radev, "ScisummNet: A Large Annotated Corpus and Content-Impact Models for Scientific Paper Summarization with Citation Networks," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 7386–7393, 2019.

[5] K. Jaidka, M.K. Chandrasekaran, S. Rustagi, M. Kan, "Overview of the CL-SciSumm 2016 Shared Task", in Proceedings of the Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL), pp. 93-102, 2016.

[6] H. P. Luhn, "The Automatic Creation of Literature Abstracts," IBM Journal of Research and Development, vol. 2, no. 2, pp. 159–165, 1958.

[7] C. Paice and P. Jones, "The identification of important concepts in highly structured technical papers", in Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '93), pp. 69-78, ACM, 1993.

[8] A. Elkiss, S. Shen, A. Fader, G. Erkan, D. States and D. Radev, "Blind men and elephants: What do citation summaries tell us about a research article?", Journal of the American Society for Information Science and Technology, vol. 59, no. 1, pp. 51-62, 2008.

[9] A. Cohan and N. Goharian, "Contextualizing Citations for Scientific Summarization using Word Embeddings and Domain Knowledge", in Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17), pp. 1133-1136, ACM, 2017.

[10] Q. Mei and C. Zhai, "Generating impact-based summaries for scientific literature", in ACL-08: HLT - 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, pp. 816-824, ACM, 2008.

[11] R. Kiros, Y. Zhu, R. Salakhudtinov, R. Zemel, A. Torralba, R. Urtasun and S. Fidler, "Skip-thought vectors", in Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'15), pp. 3294-3302, MIT Press, 2015.

[12] P. D. Taylor and L. B. Jonker, "Evolutionary stable strategies and game dynamics," Mathematical Biosciences, vol. 40, no. 1-2, pp. 145–156, 1978.

[13] A. Erdem and M. Pelillo, "Graph Transduction as a Non-cooperative Game," Graph-Based Representations in Pattern Recognition Lecture Notes in Computer Science, pp. 195–204, 2011.

[14] C. Lin, "ROUGE: A Package for Automatic Evaluation of Summaries", in Text Summarization Branches Out: Proceedings of the ACL-04 Workshop, pp. 74-81, ACL,2004.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:00:35 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Comparative Analysis of PWM Techniques and SVPWM operated V/f control of Induction Motor having different power ratings

Divyam
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
fastrack2030@gmail.com

Ayushi Saxena
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
ayushi.saxena1407@gmail.com

Bharat Singh
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
bharatsinghuae@gmail.com

J. N. Rai
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
jnrai@dce.ac.in

Corresponding Author: Bharat Singh, bharatsinghuae@gmail.com

***Abstract*— The paper makes a comparative study among the different Pulse Width Modulation (PWM) techniques including – Space Vector Pulse Width Modulation (SVPWM), Sinusoidal PWM (SPWM) and Multiple PWM (MPWM). These techniques are defined, examined and compared to generate gate pulses for a three phase Voltage Source Inverter (VSI). The modulation index is varied for achieving the comparative analysis of these techniques. Pulse Width Modulation techniques are preferred because of low harmonic content obtained in the output voltage and current, fixed switching frequency and higher utilization of the DC bus. Also, SVPWM is then used for achieving V/f control of squirrel cage Induction Motor having different power ratings.**

Keywords—pulse width modulation, space vector pulse width modulation, sinusoidal pulse width modulation, multiple pulse width modulation, induction machine, V/f control

Acronyms

- ***PWM (Pulse Width Modulation)***
- ***SPWM (Sinusoidal Pulse Width Modulation)***
- ***SVPWM (Space Vector Pulse Width Modulation)***
- ***MPWM (Multiple Pulse Width Modulation)***
- ***VSI (Voltage Source Inverter)***
- ***IM (Induction Machine)***

I. INTRODUCTION

An inverter is a power electronic converter used for obtaining ac power from dc power at the required output voltage and given frequency. They can be mainly categorized as Current Source Inverters (CSI) and Voltage Source Inverters (VSI). Voltage Source Inverters have a stiff dc voltage source at the input whereas Current Source Inverters have a stiff current source at the input. The output voltage waveform of the inverter can be sinusoidal, quasi-square wave or square wave with some distortions and can be controlled with the help of switches. The harmonic distortions in the output voltage and current can be considerably reduced by using proper switching schemes. Pulse Width Modulation is one such technique which can be implemented to achieve an efficient switching scheme.

Pulse Width Modulation is a method which uses a pulse wave whose pulse width is modulated to achieve a variation of the mean value of the output waveform. It is a way to control the switching sequence and time period for the power electronic switches used in an inverter. If the switch is ON for longer durations in a cycle, it will result in higher average output of the inverter. [1] The proportion of the ON duration of the switch to the total time period is known as the duty cycle. Also, the switching power loss is very low as negligible current flows in the OFF state of the device and the voltage drop is also negligible. Fig. 1 represents the block diagram for the PWM technique.

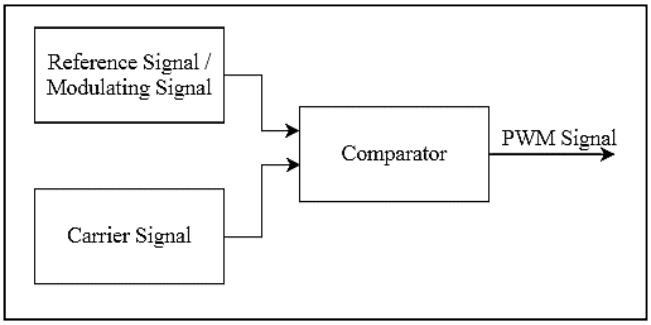


Fig. 1. Block Diagram depicting PWM Technique

The distortion in the output waveform is due to the presence of harmonics. Harmonics are the frequency component of the voltage present at integral multiples of the frequency of the main signal. This can be measured by the Total Harmonic Distortion (THD):

$$THD = \frac{\sqrt{(\sum_{n=2}^{\infty} V_{n_rms}^2)}}{V_{fund_freq}} \quad (1)$$

The THD is affected by the Modulation Index (M_A), which can be defined as the ratio of the peak value of the reference signal to the maximum value of the carrier signal.

IM are the most commonly used machines, especially as motors, because of excellent running conditions and low maintenance. Induction motors aggregate to 70-75% of total electrical load. Though the speed control techniques of IM are not much effective, V/f control is considered as the most popular method. In this paper, V/f control of Induction Motor having different power ratings is achieved by using SVPWM.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:53:43 UTC from IEEE Xplore. Restrictions apply.

II. TYPES OF PWM TECHNIQUES

A. Multiple PWM

In this method, multiple pulses are given to the switches in each half cycle. The reference signal is measured against a carrier wave in order to obtain the gate pulses of equal width for the switches of the VSI. It is also known as Uniform Pulse Width Modulation.

B. Sinusoidal PWM

In this method, the sinusoidal reference or modulating wave is measured against a large frequency carrier wave in order to generate gate pulses to be fed to the VSI.

C. Space Vector PWM

This is a technique which uses an algorithm for modifying the pulse width as per the given reference vector. In this method, vectors are used to represent all possible switching in a space having two dimensional voltage. This is obtained by the transformation to three phase independent vectors from the three phase dependent vectors. The assumption includes a constant reference vector for the corresponding sampling time. The sampling frequency is dependent on the ON and OFF times of the device used, which should be ideally infinite. Thus, for minimum total harmonic distortion and minimum switching losses, the maximum possible sampling frequency is selected. [3]

The line side frequency and magnitude of the fundamental component are regulated using the magnitude and frequency of the reference voltage vector. [1] In SVPWM, there is a unique correspondence between a three-phase system and a space vector in complex plane, which is represented in a mathematical manner.

III. METHODOLOGY

A. Implementation of SVPWM

SVPWM Technique involves transformation of three phase sinusoidal signal into revolving voltage vector in complex plane (abc to dq axis transformation) with constant magnitude revolving with constant angular frequency.

The six switches in a three phase inverter correspond to 8 switching states (all combinations of ON and OFF of the six switches) which are called voltage vectors in the three phase VSI structure. The switching sequences of the six switches should abide by these two restrictions - out of the six switches, three must always be in the ON state and the remaining switches must be in the OFF state. Also, both the switches of the same leg of the inverter should always be driven by two complementary pulse signals.

Fig. 2 shows the vectors in complex plane in different sectors. There is a 60° angle amidst any two neighboring non-zero vectors. In the meantime, zero voltage to the load is applied by V0 and V7 (two zero vectors) at the origin as there is no transfer of power in these switching states due to short circuit. The eight vectors- V0 to V7 are called basic space vectors. The required voltage reference vector V_{ref} can be obtained by applying the same transformation to desired output voltage in the dq plane. [5]

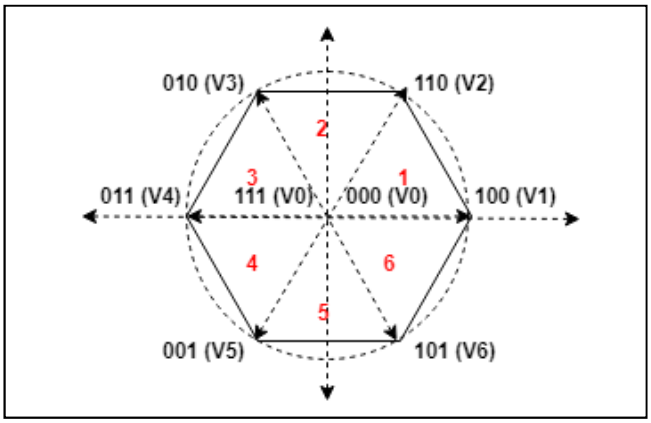


Fig. 2. Possible switching sequence for the VSI

Main function of SVPWM is to estimate voltage reference (V_{ref}) vector with the help of inverters eight switching pattern and this approximation is done by generating output voltage of inverter in small sampling period which is equivalent to V_{ref} in the same period. This V_{ref} voltage vector has discrete rotation in complex plane between sectors portioned at 60° hence forms hexagon trajectory. Decomposition of this vector on real and imaginary axis indicates that it coincides with the generating switching vector. Switching time at any sector by using V_{ref} as a combination of adjacent vectors can be determined by using Fig. 4 as:

$$T_1 = \frac{\sqrt{3}*T*|\overrightarrow{V_{ref}}|}{V_{dc}} sin(\frac{n\pi}{3} - \alpha), \quad (2)$$

$$T_2 = \frac{\sqrt{3}*T*|\overrightarrow{V_{ref}}|}{V_{dc}} sin(\alpha - \frac{(n-1)\pi}{3}), \quad (3)$$

$$T_0 = T - T_1 - T_2. \quad (4)$$

where T is the time period and n denotes number of sector ranging from 1 to 6. Also, α should lie between 0° and 60°. [5]

The switching time for the different sectors using (2), (3) and (4) for the VSI can be summarized as in Table I.

TABLE I. SWITCHING TIMES FOR VSI

Sector	Upper Switches (S_1, S_3, S_5)	Lower Switches (S_4, S_6, S_2)
1	$S_1=T_1+T_2+T_0/2$ $S_3=T_2+T_0/2$ $S_5=T_0/2$	$S_4= T_0/2$ $S_6= T_1+T_0/2$ $S_2= T_1+T_2+T_0/2$
2	$S_1=T_1+T_0/2$ $S_3= T_1+T_2+T_0/2$ $S_5= T_0/2$	$S_4= T_2+T_0/2$ $S_6= T_0/2$ $S_2= T_1+T_2+T_0/2$
3	$S_1= T_0/2$ $S_3= T_1+T_2+T_0/2$ $S_5= T_2+T_0/2$	$S_4= T_1+T_2+T_0/2$ $S_6= T_0/2$ $S_2= T_1+T_0/2$
4	$S_1= T_0/2$ $S_3= T_1+T_0/2$ $S_5= T_1+T_2+T_0/2$	$S_4= T_1+T_2+T_0/2$ $S_6= T_2+T_0/2$ $S_2= T_0/2$
5	$S_1= T_2+T_0/2$ $S_3= T_0/2$ $S_5= T_1+T_2+T_0/2$	$S_4= T_1+T_0/2$ $S_6= T_1+T_2+T_0/2$ $S_2= T_0/2$
6	$S_1= T_1+T_2+T_0/2$ $S_3= T_0/2$ $S_5= T_1+T_0/2$	$S_4= T_0/2$ $S_6= T_1+T_2+T_0/2$ $S_2= T_2+T_0/2$

The switching times as per Table I have been used in the MATLAB Simulink simulation for the Space Vector Pulse Width Modulation for VSI. Fig. 3, Fig. 4, Fig. 5 and Fig. 6 represent the implementation of this technique using MATLAB Simulink.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:53:43 UTC from IEEE Xplore. Restrictions apply.

B. *Implementation of Sinusoidal PWM*

In SPWM inverter the pole-voltage pulse widths for a time period varies in a sinusoidal manner. The sinusoidal pulse width modulation technique revolves around comparison of a triangular carrier signal having high frequency with a sinusoidal modulating signal which gives the fundamental frequency component of the output current and voltage waveform. The maximum value of the modulating signal should not be more than the maximum value of the carrier signal for linear operation. The lower and upper switches of the VSI are controlled using this comparator output. Fig. 7, Fig. 8, Fig. 9 and Fig. 10 represent the implementation of this technique using MATLAB Simulink.

C. *Implementation of Multiple PWM*

In MPWM inverter the pole-voltage pulse widths for a time period varies in a square wave manner. The multiple pulse width modulation technique revolves around comparison of a triangular carrier signal having high frequency with a square modulating signal which gives the fundamental frequency component of the output current and voltage waveform. The maximum value of the modulating signal should not be more than the maximum value of the carrier signal for linear operation. The lower and upper switches of the VSI are controlled using this comparator output. Fig. 11, Fig. 12, Fig. 13 and Fig. 14 represent the implementation of this technique using MATLAB Simulink.

D. *Implementation of V/f control of IM using SVPWM*

For the speed control of IM, though there is no best or effective way to control the speeds, the V/f control is still the most popular method. Since,

$$V = \sqrt{2}\pi f \emptyset_m N \quad (5)$$

$$B_m \propto V/f \quad (6)$$

We operate at the knee-point of the B-H curve, we do not want to vary the maximum magnetic field density (B_m). Hence, voltage and frequency are varied in such a manner that the V/f ratio is always constant. Since the maximum torque (τ_m)

$$\tau_m \propto (V/f)^2 \quad (7)$$

Hence, in V/f control, τ_m remains same and thus the power increases linearly with the rotor speed. Here, the V/f ratio is taken to be 8. This has been plotted for various induction motor drives having different power ratings using SVPWM and unity modulation factor. [9] Fig. 15 represents the implementation of this technique using MATLAB Simulink.

IV. SIMULATION RESULTS

The paper discusses about the software implementation of various PWM techniques namely SVPWM, SPWM and MPWM through MATLAB Simulink. The ultimate objective of either of these techniques is to obtain alternating output voltage with minimum harmonics and maximum fundamental component. By analyzing and comparing their THD Analysis with varying Modulation Indices, the observations can be summarized as follows:

A. *SVPWM*

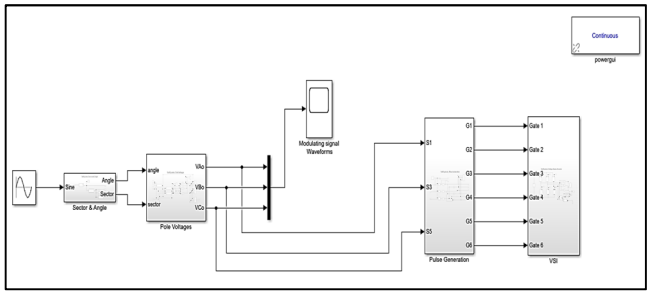


Fig. 3. MATLAB Simulink model for SVPWM with parallel RL load of 50Ω and 5 mH

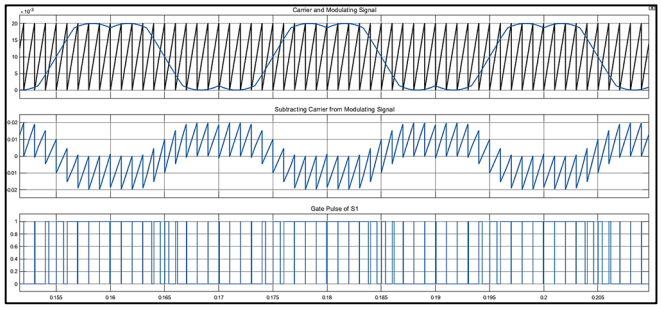


Fig. 4. Waveforms for carrier signal, modulating signal, comparator output and gate pulse for switch S1 at $M_A = 1$

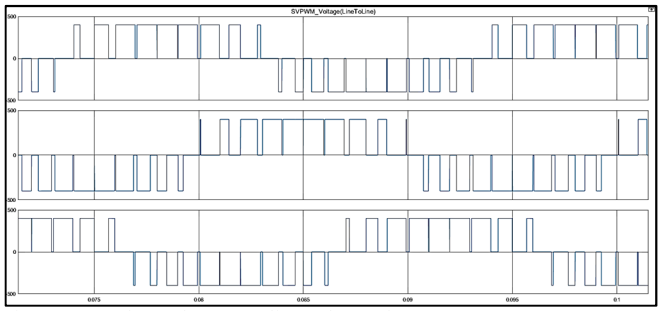


Fig. 5. Waveforms for output line voltages for SVPWM

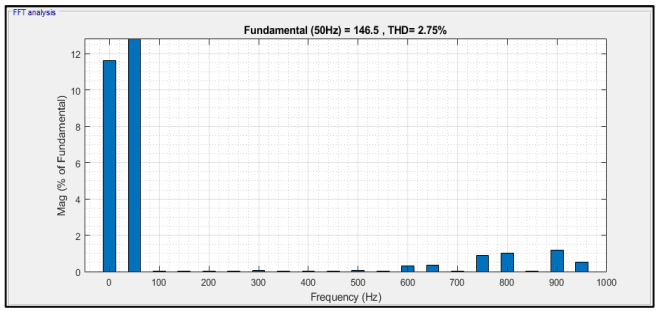


Fig. 6. THD for output current of VSI for SVPWM

B. *SPWM*

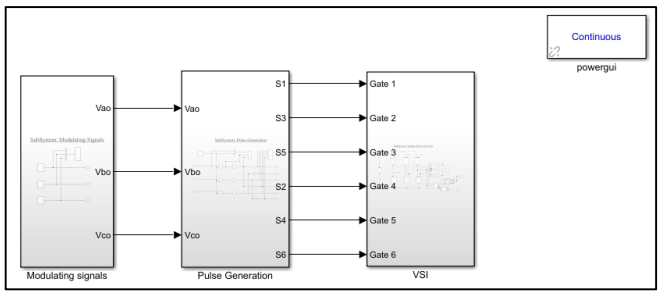


Fig. 7. MATLAB Simulation Simulink model for SPWM with parallel RL load of 50Ω and 5 mH

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:53:43 UTC from IEEE Xplore. Restrictions apply.

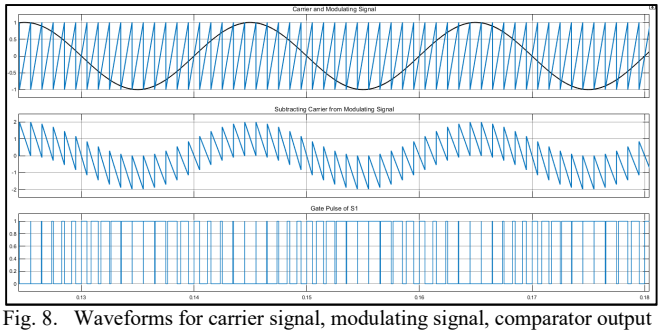


Fig. 8. Waveforms for carrier signal, modulating signal, comparator output and gate pulse for switch S1 at $M_A = 1$

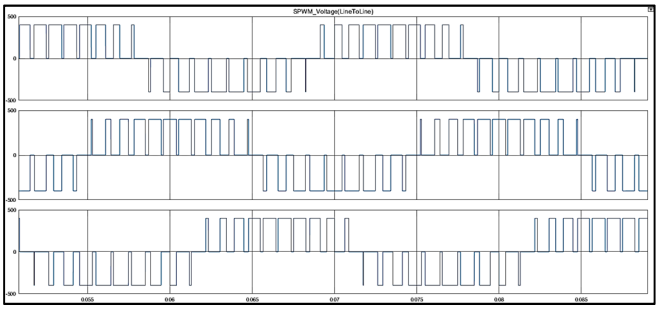


Fig. 9. Waveforms for output line voltages for SPWM

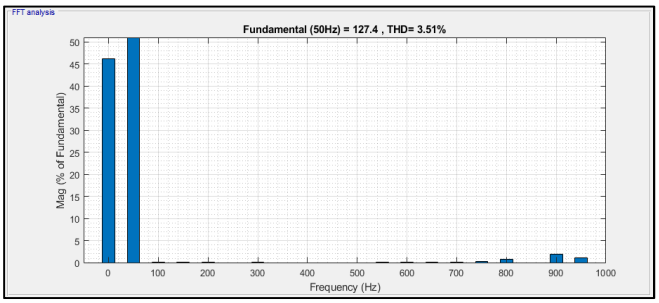


Fig. 10. THD for output current of VSI for SPWM

C. MPWM

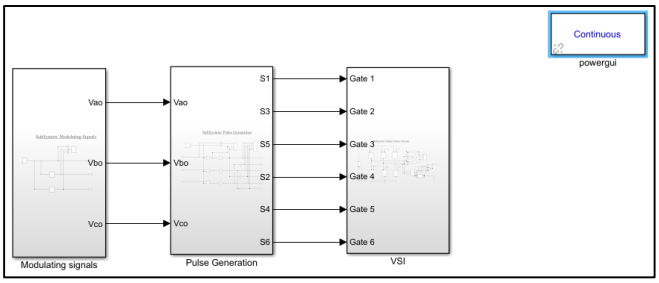


Fig. 11. MATLAB Simulink model for MPWM with parallel RL load of 50Ω and 5 mH

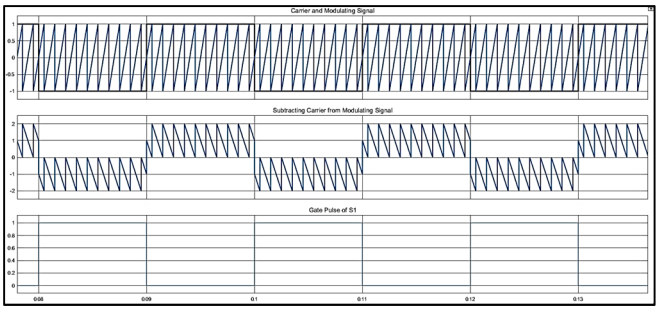


Fig. 12. Waveforms for carrier signal, modulating signal, comparator output and gate pulse for switch S1 at $M_A = 1$

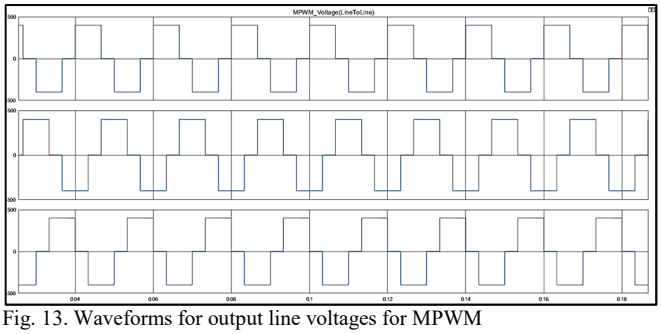


Fig. 13. Waveforms for output line voltages for MPWM

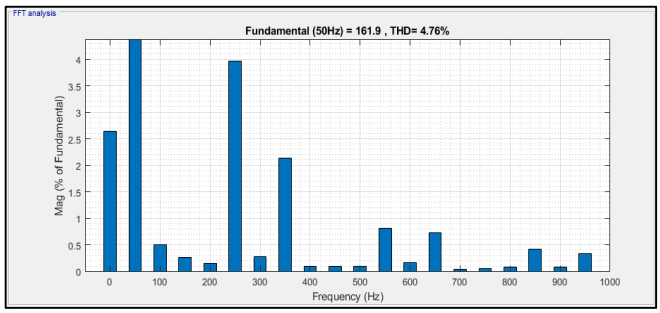


Fig. 14. THD for output current of VSI for MPWM

D. V/f control of IM using SVPWM

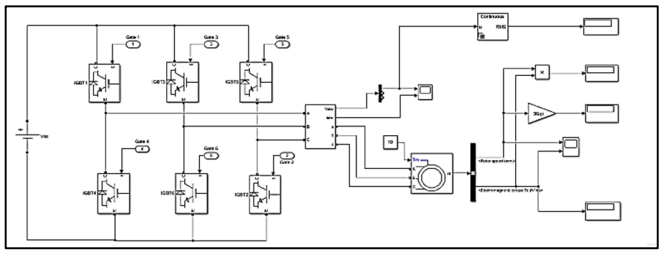


Fig. 15. MATLAB Simulink Subsystem for V/f control of IM using SVPWM

V. OBSERVATIONS

The three PWM techniques namely- MPWM, SPWM and SVPWM were compared for different modulation indices (M_A) ranging from 0 to 6 and the ratio of corresponding value of peak fundamental output line voltage (V_{01}) to constant dc supply voltage (V_{dc}=400V) was noted for all the three techniques. The observations have been summarized in Table II.

TABLE II. COMPARISON OF MPWM, SPWM AND SVPWM

MA	V_{01}/V_{dc}		
	MPWM	SPWM	SVPWM
0	0	0	0
0.4	0.44	0.3475	0.4005
0.6	0.6565	0.50725	0.598
0.8	0.8645	0.67025	0.79575
1.0	1.09575	0.848	1.0005
1.1	1.10625	0.9245	1.048
1.2	1.10625	0.95475	1.05775
1.3	1.10625	0.9645	1.058
1.4	1.10625	1.009	1.0795
1.5	1.10625	1.01025	1.08
1.6	1.10625	1.0235	1.08625
1.8	1.10625	1.04475	1.09225
2.0	1.10625	1.0525	1.092
2.2	1.10625	1.07225	1.094
2.4	1.10625	1.08325	1.10275
2.6	1.10625	1.0725	1.11175
2.8	1.10625	1.1	1.117
3.0	1.10625	1.0805	1.118

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:53:43 UTC from IEEE Xplore. Restrictions apply.

3.2	1.10625	1.094	1.115
3.5	1.10625	1.0805	1.11425
4.0	1.10625	1.10225	1.1115
4.5	1.10625	1.07475	1.11325
5	1.10625	1.08	1.1105
5.5	1.10625	1.08325	1.113
6	1.10625	1.077	1.11025

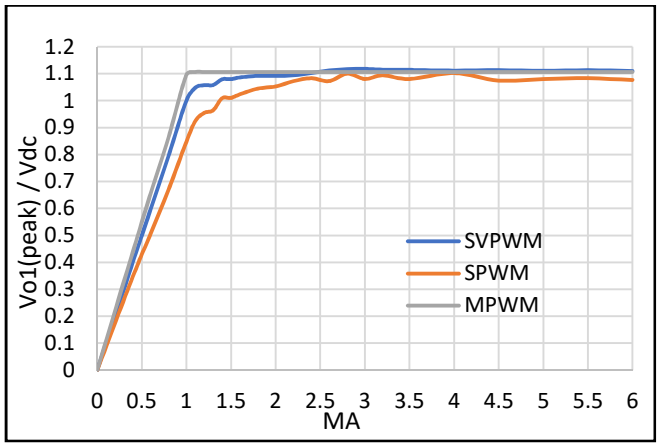


Fig. 16. Plot for V_{01}/V_{dc} Vs M_A for MPWM, SPWM and SVPWM

From Fig. 16 it is observed that the fundamental output voltage for SVPWM is 17.98% higher than that for SPWM at unity modulation index. Also, for a range of modulation index from 0.4 to 1.6, the mean fundamental output voltage for SVPWM is 12.37% higher than that for SPWM.

It is also observed from Fig. 6, Fig. 10 and Fig. 14 that the THD in output current is least for SVPWM (THD=2.75%), followed by that for SPWM (THD=3.51%) and the highest for MPWM (THD=4.76%).

Also, the peak line to line output VSI voltage is highest for MPWM as the pulse widths are almost same but the current produced by MPWM has highest distortion and thus is not preferred for practical use.

SVPWM technique, at unity modulation index and constant load torque of 10 Nm, was then used for achieving the V/f control of IM of different power ratings, the simulation results of which have been summarized in Table III.

TABLE III. COMPARISON OF V/F CONTROL OF DIFFERENT POWER RATING OF IM USING SVPWM

IM WITH POWER RATING OF 4 KW		
SPEED (RPM)	TORQUE (NM)	POWER (W)
1452	10.2	1561
1263	10.6	1409
1075	10.11	1139
886.8	9.95	923.1
698.9	9.94	727.3
513.2	9.83	527.9
337	10.8	387.2
77.55	10.19	82.74
IM WITH POWER RATING OF 7.5 KW		
SPEED (RPM)	**TORQUE (NM)**	**POWER (W)**
1477	9.69	1500
1289	10.08	1361
1101	9.71	1119
913	9.69	927
724.1	9.63	729.7
505.1	11.89	628.7
419.3	19.07	837.4
125.2	4.58	60.1

IM WITH POWER RATING OF 15 KW		
SPEED (RPM)	TORQUE (NM)	POWER (W)
1494	9.95	1557
1305	11.17	1527
1118	9.97	1167
929.6	10.21	994
740.6	12.31	954.7
530	53	3300

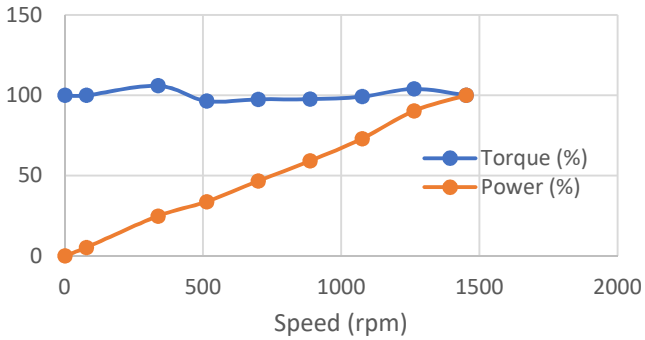


Fig. 17. 4 kW IM characteristics for V/f control

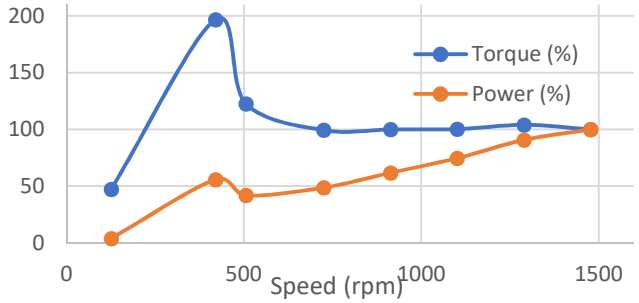


Fig. 18. 7.5 kW IM characteristics for V/f control

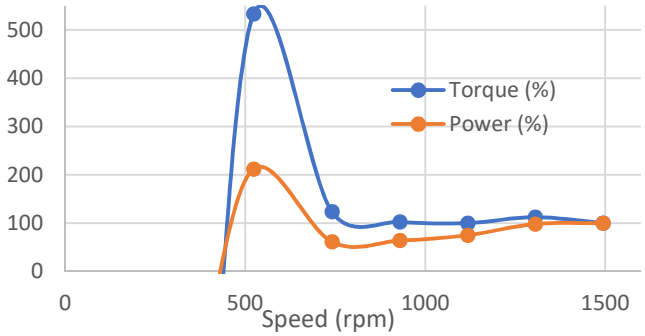


Fig. 19. 15 kW IM characteristics for V/f control

From Fig. 16, we observe that the torque remains almost constant throughout the V/f operation and the power varied linearly with the rotor speed.

From Fig. 17, Fig. 18 and Fig. 19, we observe that in the low frequency region, since the inductance becomes very low, the circuit becomes resistive in nature and hence the impedance does not vary with frequency, so with decreasing voltage and frequency, the current in the machine decreases and hence torque decreases.[12] Therefore, in low frequency region (< 37.5% of the rated frequency), the performance of IM deteriorates. Also, as the machine power rating increases, there is a decrease in the impedance of the machine and hence it becomes constant with respect to frequency even in the relatively higher frequency range (in the low frequency region) and accordingly we observe a greater dip in the torque with the increase in the power rating of the IM.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:53:43 UTC from IEEE Xplore. Restrictions apply.

VI. CONCLUSION

Here, SVPWM Technique was used to reduce harmonics in output current of the VSI. Also, it was compared with other PWM techniques, namely – SPWM and MPWM, it can be concluded that SVPWM provides the least THD in the output line current when compared to SPWM and MPWM and MPWM provides the highest THD. Also, the maximum line output VSI voltage is higher for SVPWM as compared to SPWM. Although the maximum line output VSI voltage is high for MPWM, it is not preferred due to presence of high harmonic distortions in current. For MPWM, V_{01} saturates for M_A greater than unity, i.e. over modulation region. However, in case of SPWM and SVPWM, V_{01} can be increased even in the over modulation region, though non-linearly, and does not saturate as in MPWM. Thus, based on above conclusions, SVPWM is a more sophisticated and effective technique for inverter operation and hence is used for the V/f control.

We also conclude that V/f control should not be preferred in the lower frequency range and the performance of IM deteriorates in this region.

Further, in future, the concept of Random PWM (RPWM) implementing random switching frequency method can be employed along with the above implementation of the pulse width modulation techniques for achieving a harmonic content spread over the entire harmonic spectrum and not particularly on one harmonic frequency.

ACKNOWLEDGMENT

We would like to thank Prof J.N. Rai, for his constant motivation and guidance during the project. Without his timely help and support, the completion of this paper wouldn't have been possible. We would also like to thank the publisher of the research papers from where we have taken reference.

REFERENCES

1. Avinash Mishra, Swaraj Save, Rohit Sen ," Space Vector Pulse Width Modulation" , International Journal of Scientific and Engineering Research, Feb 2014.
2. Wan Noraishah, Mohd. Firdaus Ismail, Ahmed Farid Abidin, "Multi-Phase Inverter Space Vector Modulation", 2013 IEEE 7th International Power Engineering and Optimization Conference (PEOCO2013),Malaysia,June 2013.
3. B.Arulvani, V.Padmathilagam, "Space Vector Pulse Width Modulation Controlled Multilevel CSI", International Research Journal of Computer Science (IRJCS), May 2015.
4. Lingling Fan "Realizing Space Vector Modulation in MATLAB and PSCAD", IEEE, 2013.
5. Renu Sharma, Deepak Kumar Goyal , Harpreet Singh, "Analysis of Space Vector PWM for 3 Phase Inverter and comparison with SPWM",International Journal of Advanced Research in Electrical , Electronics and Instrumentation Engineering (IJAREEIE), Jan 2016.
6. Ayman Y. Yusuf, "Space Vector Pulse Width Modulation Technique" International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE) ISSN: 0976-1353 Volume 15 Issue 1 –MAY 2015.
7. Ahmed Muhsin Abdul Majeed, Enas Hamid Ibrahim, Nadia Qassim, "Investigation of SPWM and SVPWM VSI for Induction Motor Drive" ,The Fourth Post Graduate Engineering Conference.
8. Umashankar S, Arun Shankar V.K., Geet Jain, Mohan L Kolhe, "Comparative Evaluation of Pulse Width Modulation Techniques on Effective DC Link Voltage Utilization of Grid Connected Inverter", International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) – 2016.
9. Ranju Bharti, Madhav Kumar, B.M Prasad, "V/F Control of Three Phase Induction Motor", International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN), 2019.
10. Di Wu, CS Chang, Y Fang, H Bai, "Performance improvement of V/f induction -motor control in the low-frequency range", 8th International Conference on Advances in Power System Control, Operation and Management, 2009.
11. Bose, B. K., Modern Power Electronics and AC drive, Prentice Hall, 2001.
12. Zhe Zhang, Yiqi Liu, Ali M. Bazzi, "An Improved High-performance Open-loop V/f Control Method for Induction Machines", IEEE Applied Power Electronics Conference and Exposition (APEC), 2017.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:53:43 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Comparative Analysis of Route Planning Algorithms on Road Networks

R. K. Yadav
Department of Computer Science and Engineering,
Delhi Technological University,
New Delhi, India
rkyadav@dtu.ac.in

Giriraj Kishor
Department of Computer Science and Engineering,
Delhi Technological University,
New Delhi, India
girirajrpvv@gmail.com

Himanshu
Department of Computer Science and Engineering,
Delhi Technological University,
New Delhi, India
himanshu21498@gmail.com

Kishan Kashyap
Department of Computer Science and Engineering,
Delhi Technological University,
New Delhi, India
kishankashyap5march@gmail.com

***Abstract* — It is merely impossible to know the history of Route Planning Algorithms. It can be imagined that from the very primitive era, animals have dedicated their time to find the shortest path to reach their food. This kind of algorithms is one of the most practised ones and forms the basis for solving many real-world problems. But like most of the graph algorithms, they were proposed in the mid-twentieth century. The applications of these algorithms are many, the most important one is its usage in Routing and Navigation based applications like Google Maps, Ola and, Uber. Here, the road network is resolved as a large graph with nodes as coordinates holding some meaning and edges as ways/paths connecting adjacent coordinates, then route planning algorithms are applied for finding the shortest path between two nodes in the graph created. A noteworthy application is in Routing data packets on the internet; the Open Shortest Path First (OSPF) routing protocol is based on the Dijkstra algorithm and is used in autonomous systems like Local Area Network (LAN). It is also being used in routing especially long-distance calls. Another interesting usage can be seen in the Currency Arbitrage strategy where currencies are bought and sold instantaneously from one market to another to make a profit from the exchange rate divergences. Due to applications in so many important fields, these algorithms are studied rigorously yet to come up with a more optimized version targeting certain scenarios or in general. It is in this understanding that this paper delves into the study and implementation of various route planning algorithms and portrays their performances based on efficiency, in comparison with each other when running on a graph obtained from real road network data. This paper will also deduce a set of parameters that affects the traffic flow and the evaluation metrics that should be considered when applying those algorithms.**

Keywords — Graph Theory, Route Planning Algorithms, Shortest Path, Dijkstra, A* search, Landmark Heuristic, Contraction Hierarchy, Transit Node Routing, Fitness Function, Genetic Algorithm.

I. INTRODUCTION

One of the first papers on the shortest path algorithms was given by Dantzig in the year 1958 to sort the list of edges based on the distance from the source node and to ignore the arcs of the list involving nodes which have been reached earlier. The complexity of the proposed algorithm was $O(n^2 logn)$, where n is the number of nodes in the graph. In the following year, Dijkstra proposed an improved method where a list of size n was maintained and for each vertex, the best path found so far was stored. The complexity was $O(n^2)$ and later it was realized that an efficient data structure like heaps can even reduce the running time to $O(m + nlogn)$. In 1964, Neel Nilson invented a heuristic-based approach to speed up the Dijkstra algorithm and named it A1; then in 1968 with the help of Bertram Raphael, Peter E. Hart gave the concept of consistent heuristic and named versions of the algorithms using it as A* [1]. In the case of road-networks, the size of the graph is huge and the Dijkstra algorithm is not time-efficient for the shortest path search queries due to enlarged search space [10]. Therefore, many improvements to the shortest path algorithm were proposed in 2007-2010 by the professors of KIT, Germany including D. Schultes, P. Sanders, and others. These algorithms include ALT, Contraction Hierarchies, and Transit Node Routing to speed up Dijkstra and are exhaustively being used by many routing based applications. The research is still going on in this field.

Estimating the optimal path to reach from one location to another on roadway networks is one of the most concerning tasks in the conveyance, distribution, and logistics applications. The routing algorithms should be highly scalable to accommodate various distributions amongst road networks that highly depend on the geography and demography of a given place.

This research paper aims to address the widely known algorithms for route planning that is being used on the motorway systems. Based on the implementation results of the route planning algorithms, a comparison based on the efficiency of these algorithms applied on a graph obtained from a real instance of a road network.

Afterwards, various parameters are inferred that create an impact over deciding the cost of a path. Various Evaluation metrics that should be taken into consideration for estimating the efficacy of a routing algorithm are also mentioned.

II. LITERATURE OVERVIEW

Route planning algorithms

There are several route planning algorithms applied in road networks, these are:

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:58:23 UTC from IEEE Xplore. Restrictions apply.

1) Dijkstra algorithm: This algorithm calculates the shortest paths from a given node (the start node) to all the other nodes in the graph [6]. One of the popular implementations of this algorithm is by maintaining a priority queue where priority is the degree of closeness with the start node. Hence, it comes under the famous Single Source Shortest Pathfinding algorithms. The nodes are visited in a sequence based on proximity with the start node and the algorithm terminates when either the target node is popped from the queue (i.e. settled) or all the nodes are settled depending upon the terminating condition being used. The correctness of the algorithm fails for negative edge weights and negative edge cycles. Bellman-Ford algorithm is a dynamic programming-based approach that works for such graphs and overcomes the limitation of the Dijkstra algorithm [3].

2) A* algorithm: A* is a successor of the Dijkstra algorithm and uses a heuristic function that approximates the distance to the target node which helps in leaning towards the target with an idea of the direction to expand the nodes rather than expanding them based on the closeness with source node. To perform an efficiently heuristic function is expected to be consistent or monotone and hence admissible [4]. A heuristic function is *consistent* if, for every node *n* and its neighbor *s*, its value for *n* is never greater than the sum of step cost of getting to *s* and the heuristic value for *s*. Hence:

$$h(n) \leq h(s) + dist(n, s) \quad (1)$$

3) A*, Landmarks, and Triangle Inequality algorithm: ALT (A* Landmark Technique) is a preprocessing-based, goal-directed and more efficient technique than the Dijkstra algorithm. The heuristic used by this algorithm uses one basic principle called reverse triangle inequality which states that the length of a side of any triangle is always greater than the difference in the length of the other two sides. Using this theorem, a set of landmarks are chosen and then the heuristic value for a particular node is taken as the maximum difference in the pre-computed shortest path-length of any landmark from that particular node and the target node [4]. The pre-computation for this heuristic to become a good lower bound is massive and depends upon the number of landmarks chosen as for each landmark the 'to' and 'from' distance for every other node in the graph needs to be stored or at least one of the two if the graph is undirected. It further depends on how isolated these landmarks are from each other; creating a set of the landmarks situated at the maximum distances from each other is an N-P Hard problem [7] as finding the maximum length path in a directed cyclic graph is an N-P Hard problem too.

4) Contraction Hierarchy: Here a contraction graph is created from the original graph where every node has a level assigned based on the sequence of contraction and this sequence is based on ordering the nodes by some measure of importance [8]. Contraction of a node deals with adding shortcuts if the shortest path between its neighbors goes via it; this step makes it the most efficient route planning technique but increases the pre-computation time. During the query, an upward graph is formed starting from the source with edges directed from the lower level to a higher level, and a backward graph from the target is also formed in the same manner. Dijkstra algorithm from the source and the target is applied to the upward and downward graph respectively.

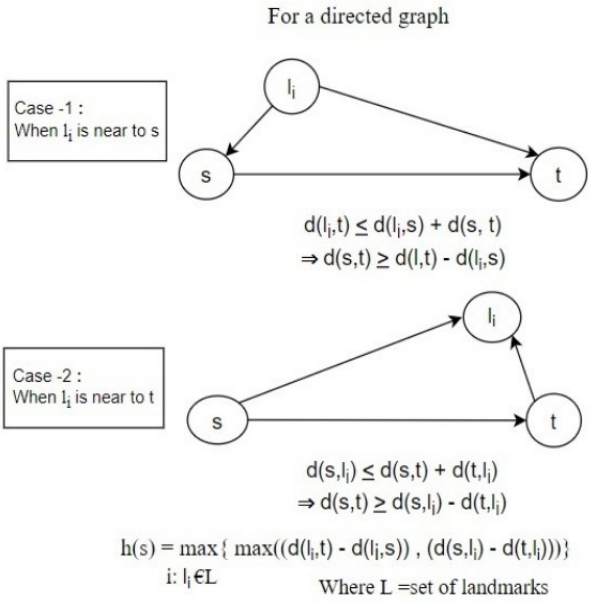


Fig. 1: Calculating the Heuristic Value for Node's' using ALT

From the set of common nodes settled in both the graphs '*S*':

$$dist(s, t) = min\{dist(s, v)+dist(v, t): v \in S; s, t \in V\} \quad (2)$$

The basis of this algorithm is the optimal node ordering (achieved using simple heuristics) [8], based on that forward and backward graphs are formed. Unidirectional Dijkstra turns out to be useless for the hierarchical approach.

The above methodology of applying search algorithms from both the source and target nodes comes under Bidirectional Search and is seen to be much efficient when compared to unidirectional search as the number of nodes expanded is significantly lesser.

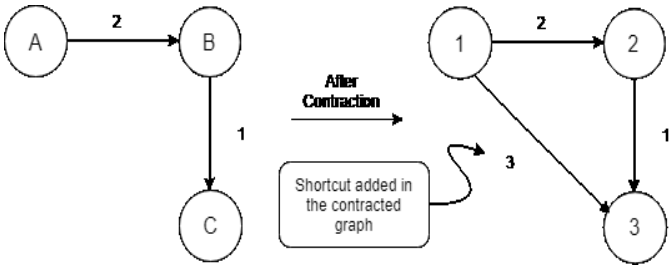


Fig. 2: Creating Contraction Graph (Here, Node Ordering is based on the Sequence of Node Contraction.)

5) Transit Node Routing: Transit Node Routing (TNR) is a technique and not a whole, independent algorithm in itself; for example one of the famous TNR variants uses Contraction Hierarchy pre-computations in its implementation. Its major drawback is the time it spends in

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:58:23 UTC from IEEE Xplore. Restrictions apply.

preprocessing which is significantly larger than even Contraction Hierarchy but this makes it the most efficient algorithm of its type. The underlying observation in TNR is: For multiple long distant routes with common start point and varying destination points, the initial portion of the shortest path route will always be one of the 'few' standard routes [9]. The nodes present at the periphery of this initial portion are called access nodes and a separate list of these nodes is maintained for both the source (say *X*) and the destination (say *Y*) also storing the distance from them respectively. Say $x \in X$ and $y \in Y$ then during query processing, the minimum of *dist(x, source) + dist(x, y) + dist(y, destination)* from every combination of *x* and *y* (total: $|X| + |Y|$) is taken as the shortest path from source to destination.

The most widely known variant of TNR uses Contraction Hierarchy for pre-computation [9]. CH orders the node in such a way that a node occurring in many shortest paths is assigned a higher level in the hierarchy. Hence a set of nodes above a certain threshold level is chosen for selecting the access nodes (called transit nodes). From the source node, a forward search in the upward graph is run and all the transit nodes settled during the run become access nodes. During this Dijkstra run, edges leaving access nodes remain unrelaxed. Similarly, a backward search in the downward graph from the destination node is run to collect its access nodes.

6) Multi-label Algorithms: The above routing algorithms only apply for a scalar number as the edge-weight that is the travel-time but there are also other important criteria that should be taken into consideration while calculating the most optimal path. These factors are discussed in the later section. All the above-mentioned algorithms can be transformed to consider these kinds of criteria but due to the dynamic nature of the road-network, the graphs are also dynamic and the weight also changes with time hence a mean value of these criteria is taken making these algorithms still static in nature.

7) Genetic Algorithm: Genetic algorithm (GA) falls in the range of meta-heuristic inspired by the process of natural selection. It belongs to the larger class of Biologically-inspired algorithms. In this paper, GA Algorithm is primarily used to generate high-quality solutions, enhance and explore the solution to route optimization problems [13].

Working: GA initiates its procedure with a population of candidates (also referred to as a chromosome). Each individual represents that it can be a feasible solution to the problem. Then, each individual is associated with some fitness value according to the fitness function. The fitness function is designed following the solution's quality [15]. The reproduction process generates the next generation of individuals by two distinct processes named 'repair/crossover and 'mutation'.

Need for GA: In maximum real-existence eventualities, the shortest path might not be a great course for the user. In fact, distinct customers may additionally opt for exclusive routes depending on the traffic congestion level, road incidents, and different time various quantities. Therefore, want to explore a multi-objective optimization algorithm that can deal with many distinctive targets simultaneously.

Crossover: It is a process that starts with two fittest individuals (viz. Parents). Also, crossover requires that they should have at least one different gene in both parents and afterward, the traditional method of one point crossover operation is simulated, and finally, any variant of the mutation function is applied on generated offsprings to remove feasible loops and add randomness in the solution [15].

Parent Selection: One of these three specific choice schemes for parent selection:

- Roulette wheel choice
- Rank-primarily based choice
- Match selection (with a tournament size of 2).

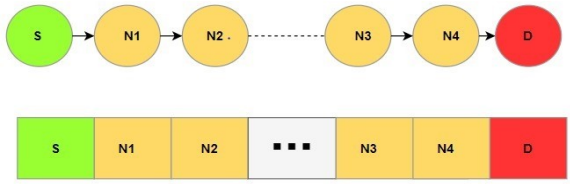


Fig. 3: Representation of a Chromosome

III. Enhancement Techniques for the Route Planning Algorithms

Through this part, the techniques that can be used to enhance the proficiency of the prevailing route planning algorithms is introduced for real and large network of roads based on the information obtained so far. Firstly, discussed the most crucial parameters that must always be considered while designing the algorithms for route planning and then mentioned some other parameters that can be implemented in their designing process to improve their performance on real road networks. After discussing these parameters, found and present some performance evaluation metrics to be used for evaluating the performance of various algorithms used in route planning for real and large road networks.

Now the input parameters are discussed, criteria for determination of the best route and evaluation metrics for route planning algorithms one by one in the next sections.

A. Crucial Parameters for route planning algorithms

Here, the most crucial parameters are discussed that must be always considered as parameters for designing the route planning algorithms. The parameters to be considered are as follows:

- Knowledge about the starting location: The beginning point for travel or the point on the map from where the journey commences.
- Knowledge about the destination location: The point on the map where the journey terminates or the end point for travel. Information regarding the objective for travelling is also useful.
- Features of road network: This information contains the details such as the number of lanes on that particular road, the present traffic scenario of the road, weather

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:58:23 UTC from IEEE Xplore. Restrictions apply.

conditions of the area, information about incidents such as accidents and road blockages etc.

- Vehicle characteristics and other insights: This information contains the details such as vehicle-related conditions including the size and speed of vehicle, the fuel remaining in the fuel tank, any specific condition related to traveller or the path to be chosen, the operating hours of vehicle etc.

Along with the aforementioned parameters, there are some other parameters that can be utilized in the process of designing the route planning algorithms such as the features of vehicle, i.e., its condition and physical specifications and also the information regarding the present traffic and future predictions of traffic on the road network [11]. These mentioned parameters assists in the process of choosing the best route.

For the implementation of routing algorithms on real and large road networks, there is a need to provide the current traffic conditions, the graph of road networks, the traffic prediction data, and the characteristics of vehicle as input parameters. These input parameters are processed in the routing algorithm and continuously updated with respect to the changing traffic conditions and occurrence of specific incidents.

The goal of routing algorithm is to determine the most optimal route to reach the target location (destination) from vehicle's current location (origin). This route is suitably changed in order to prevent such updates from affecting the selected travel path as well as making the route unavailable for variable duration and have to note that the features of vehicle is a very crucial input for route planning as vehicle's weight, its height and its vehicle class (e.g. bus, car, train) are crucial parameters to determine the best route in accordance to the physical capacities of the road networks (like lanes, flyovers, underpasses, type of highway etc.) and the driving regulations of our country.

B. Criteria for determination of the best route for these algorithms

In case needed to determine the best route from the start location to the destination location as per the request of driver, then first deduce the metric or the set of metrics which must be utilized in order to compare the different routes from a set of routes for that specific vehicle [11]. In order to measure and identify an optimal route by the objective function, the objective function to be composed of a single metric or a composed of a set of metrics applicable to any of the aforementioned algorithms is needed. In some studies, researchers have applied a wide variety of parameters in their routing algorithms. Now in the further section, presented the most valuable metrics that the route planning algorithms may utilize to find out the requested route as per our need.

- Time taken for travel: According to this metric, tend to look for the quickest route for journey instead of the smallest one (in terms of distance) among all the available routes. The quickest route implies a solution path that converges to the destination within the shortest time. Due to the dynamic nature of road and its associated parameters, the quickest route might not be the same as the smallest route. Furthermore, needed to update the quickest path continuously as the vehicle moves towards its destination and that too in accordance with the traffic conditions on road networks which change rapidly.

- Travel distance: In a road network, the values of the length of all roads have a specific value. When talk about finding the shortest path, it means interested in a path for a given vehicle through which it travels with the least distance from the start point to the destination point on its journey.

- Cost of travel: The cost of travel refers to all the expenses that have been incurred during the journey. It includes the cost of fuel used in addition to the total amount of toll taxes paid by the vehicle's driver on the route. This cost is affected by several factors such as the length of path to cover, the number of halting points on the route, the vehicle-type and the dimension and condition of road chosen for travel. So, to get a cost efficient route, routing algorithms can use this criterion, and provide a low cost route for reaching the destination location to the driver.

- Traffic conditions: The traffic conditions of the road network refer to the traffic congestion levels. If the traffic congestion on one route is less as compared to the other route, the vehicle must take the path having less traffic congestion, if the travel time for this path is also less, so as to reach the destination location in minimal time.

- Ease of driving: In a road network, the driver tends to choose a route which is smooth, safe and comfortable. This metric can be thought of a metric solely based on the driver's preference and several other parameters such as the dimensions of road, the number of traffic signals on road, the number of lanes and turns on road etc.

- Integration of two or more metrics: As per this criterion, two or more of the aforementioned criteria can be integrated with one another to act as a single metric to handle the preferences of a specific driver.

C. Evaluation metrics for route planning algorithms

Evaluation metrics are the certain parameters that are used for evaluating the performance of a particular algorithm. These metrics assist in differentiating between the different routing algorithms based on their performance measure. The four significant evaluation metrics that can be used for evaluating route planning algorithms are computational complexity, execution time, quality of the solution, and scalability.

- Computational complexity: Computational complexity refers to the computational performance of a particular route planning algorithms and the resources it requires for its execution. For instance, an algorithm may not be the best routing algorithm if it finds the most optimal route with a very high complexity, even if has comparatively low journey time, especially when the route needs to be continuously updated during the vehicle's journey.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:58:23 UTC from IEEE Xplore. Restrictions apply.

- Execution time: The execution time of these algorithms also serves as an evaluation criteria. The algorithms having overall less execution time for a variety of route is considered to be better than the others.
- Quality of the solution: In case to find multiple best routes, need to use a metric to find the best out of all the available routes. For this purpose, use this particular metric to compare the different solutions obtained by using different algorithms based on various parameters mentioned earlier, and then determining which of these algorithms provides the most optimal route by giving a value closest to the solution.
- Scalability: Scalability for a routing algorithm refers to the change in the performance and efficiency of a routing algorithm when the real network of roads is scaled up or down in terms of size. Hence, an algorithm which is efficient for a particular area might be inefficient or inapplicable when the area under consideration is increased in size.

IV. Experimental Results And Analysis

To fulfil the aim i.e. to present a comparison of various route planning algorithms, the above-stated algorithms namely-: Dijkstra, ALT, Contraction-Hierarchy, and Transit Node Routing algorithms (plus bidirectional variants) were implemented in Java on Eclipse IDE. On a simple weighted directed (cyclic) graph with no self-loop having a total of 422047 nodes and 1011274 edges obtained from parsing the OSM road-network data for the union territory of India 'Delhi', following line graphs (Execution Time (in ms) vs. Query number) were obtained when aforementioned algorithms were run for 420 shortest path queries between randomly chosen source and target nodes:

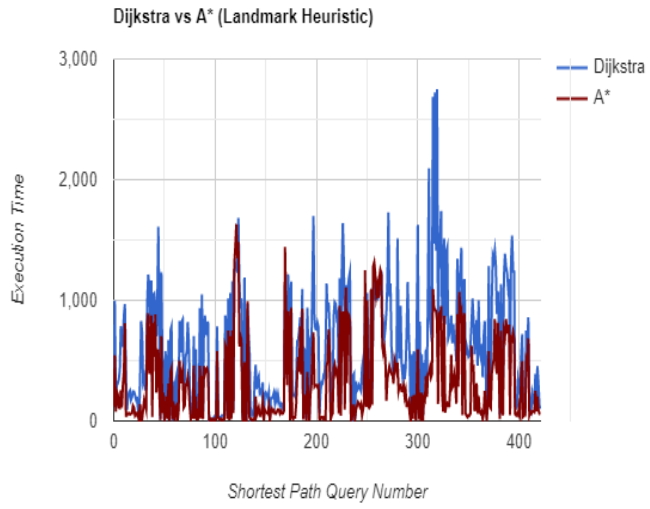


Fig. 4: A Comparison between Dijkstra and ALT based on Execution Time.

- Apparently, for most of the queries, ALT algorithm proved to be more efficient than the Dijkstra algorithm.
- On testing against various values of landmark-set size, the threshold for the given graph is 8 as beyond this value, no significant improvement in efficiency of the ALT algorithm was observed.

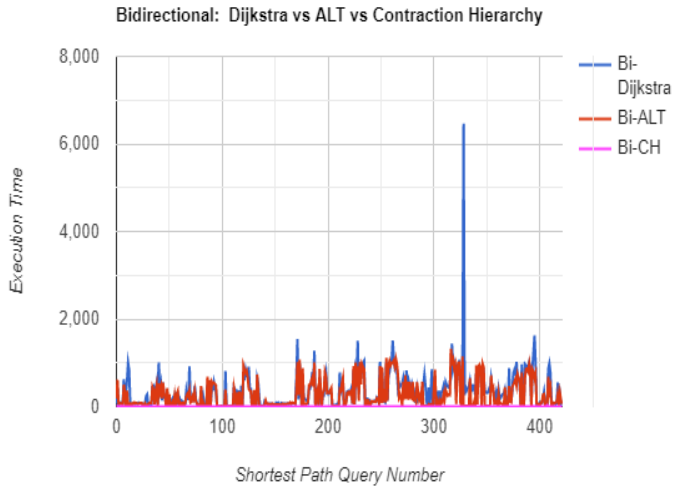


Fig. 5: A Comparison between Bidirectional Dijkstra, Bidirectional ALT and Bidirectional Contraction Hierarchy.

- Here, the blue peaks in the graph depict that the Bidirectional Dijkstra algorithm is least efficient amongst the three.
- As expected, Contraction Hierarchy took negligible time to calculate the optimal path for any of the queries.
- The difference between the execution time of Bidirectional Dijkstra and ALT was thin in many queries due to smaller path-length.

Fig. 5 shows the bidirectional variant of the three algorithms namely Dijkstra, ALT, and CH. When Bidirectional Dijkstra and ALT were run on the same 420 queries, results were difficult to interpret. Sometimes ALT performed better and other times Bidirectional Dijkstra had upper hand. But in the below graph (Fig. 6), the efficiency of Bidirectional Dijkstra emerged clearly for relatively longer path lengths where ALT took much longer times to process those queries.

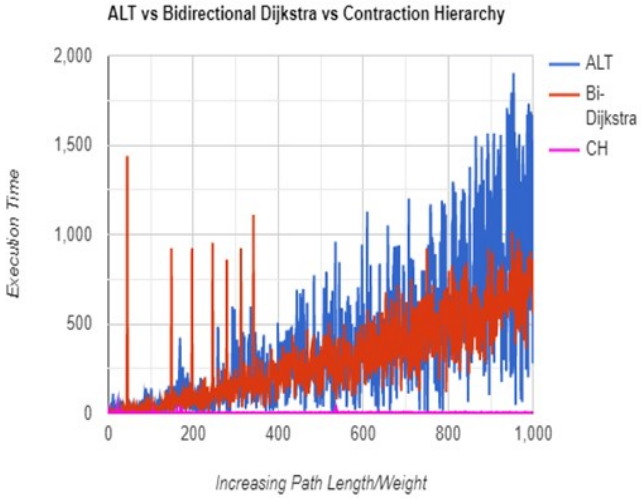


Fig. 6: A Comparison between Bidirectional Dijkstra, ALT and CH based on Increasing Path Length.

- The longest path length was estimated to be around 1300 km (by running on 10 different queries).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:58:23 UTC from IEEE Xplore. Restrictions apply.

- Bidirectional Dijkstra proved to be better than ALT when the path length was relatively larger with respect to the longest path length found in the graph [2] while CH outperforms all of them.

The below Figs. (Fig. 7 and 8) were obtained after running Transit Node Routing and Contraction Hierarchy on the simple directed weighted graph obtained from Andorra.osm file. Due to heavy preprocessing required by both the algorithms Andorra.osm was chosen to create a smaller graph. Andorra is a European Country with an area of only 468 km^2. From the bar graphs, it can be clearly seen that the preprocessing and query time are completely opposite of each other for TNR and CH. The time spent during preprocessing is compensated with the improvement seen during the shortest path query processing in the case of Transit Node Routing.

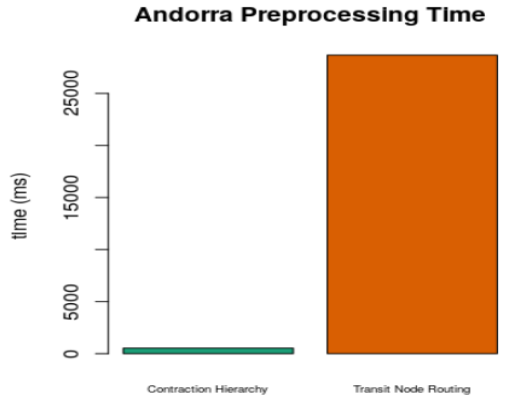


Fig. 7: Bar graph for comparing pre-computation time between CH (in green) and TNR (in orange)

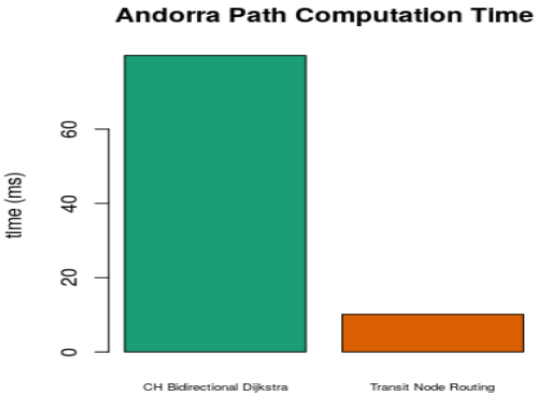


Fig. 8: Bar graph for comparing query execution time between CH (in green) and TNR (in orange)

V. Conclusion

It can be concluded that in order to create a more efficient route planning algorithm a compromise has to be made by doing more pre-processing. Transit Node Routing is the fastest route planning technique with approximately constant query processing time but with the highest pre-computation time. Contraction Hierarchy is one of the simplest ways to calculate the shortest path distance between two nodes in a graph and A* algorithms can also be fast with a good admissible heuristics while finding the shortest path is not always effective in real-time scenarios of the motorway systems as there are several parameters, the most important being the traffic that also contribute majorly in determining the optimal path.

References

[1] P. E. HART, N. J. NILSSON, AND B. RAPHAEL., "A FORMAL BASIS FOR THE HEURISTIC DETERMINATION OF MINIMUM COST PATHS IN GRAPHS." IEEE TRANS. SYST. SCI. AND CYBERNETICS, SSC-4(2):100–107, 1968.

[2] SAWLANI, SNEHA, "EXPLAINING THE PERFORMANCE OF BIDIRECTIONAL DIJKSTRA AND A* ON ROAD NETWORKS" (2017). ELECTRONIC THESES AND DISSERTATIONS. 1303.

[3] THIPPESWAMY.K,HANUMANTHAPPA.J., DR.MANJAIAH D.H. "A STUDY ON CONTRAST AND COMPARISON BETWEEN BELLMAN-FORD ALGORITHM AND DIJKSTRA'S ALGORITHM" (2015).

[4] A. V. GOLDBERG AND C. HARRELSON., "COMPUTING THE SHORTEST PATH: A* SEARCH MEETS GRAPH THEORY." PROCEEDINGS OF THE SIXTEENTH ANNUAL ACM-SIAM SYMPOSIUM ON DISCRETE ALGORITHMS, PAGES 156–165, 2005.

[5] K.ROHILA, P.GOUTHAMI, PRIYA M, "DIJKSTRA'S SHORTEST PATH ALGORITHM FOR ROAD NETWORK",IJIRCCE,VOL. 2, ISSUE 10, OCTOBER 2014,ISSN(ONLINE): 2320-9801

[6] DOMINIK SCHULTES, "ROUTE PLANNING IN ROAD NETWORKS",PAGE 22-29

[7] M. POTIMIAS, F. BONCHI, C. CASTILLO, A. GIONIS, "FAST SHORTEST PATH DISTANCE ESTIMATION IN LARGE NETWORKS" (2009)

[8] R. GEISBERGER, P. SANDERS, D. SCHULTES, D. DELLING, "CONTRACTION HIERARCHIES: FASTER AND SIMPLER HIERARCHICAL ROUTING IN ROAD NETWORKS" (2008)

[9] J. ARZ, D. LUXEN, P. SANDERS, "TRANSIT NODE ROUTING RECONSIDERED." (2013)

[10] LU FL, XIAO JX. , "AN IMPROVEMENT OF THE SHORTEST PATH ALGORITHM BASED ON DIJKSTRA ALGORITHM." NEW YORK: IEEE; 2010.

[11] VI NGOC-NHA TRAN,JOHN MURPHY, "A COMPARATIVE STUDY OF VEHICLES' ROUTING ALGORITHMS FOR ROUTE PLANNING IN SMART CITIES." CONFERENCE: 2012 FIRST INTERNATIONAL WORKSHOP ON VEHICULAR TRAFFIC MANAGEMENT FOR SMART CITIES (VTM).

[12] SURANGA NANAYAKKARA,D. SRINIVASAN,LAI WEI LUP, "A COMPARATIVE ANALYSIS OF MULTI CRITERA ROAD NETWORK, MUHAMMAD S. F. QURESHI, SYED MUJTABA SHAH, MR. GHARIB ISMAIL GHARIB AL-MATROUSHI" 2013.

[13] LINGARAJ, HALDURAI. (2016), "A STUDY ON GENETIC ALGORITHM AND ITS APPLICATIONS.INTERNATIONAL JOURNAL OF COMPUTER SCIENCES AND ENGINEERING." 4. 139-143.

[14] "ROUTE OPTIMIZATION TECHNIQUE: AN OVERVIEW", BALE, D. L. T. UGWU, C. NWACHUKWU, E. O., DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF NIGERIA.

[15] XAVIER GERMAN, "GENETIC ALGORITHM BASED ROUTE PLANNER FOR LARGE URBAN STREET NETWORKS." DOI:10.1109/CEC.2007.4425056 • SOURCE: IEEE XPLORE CONFERENCE: EVOLUTIONARY COMPUTATION, 2007

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:58:23 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Comparative Study of Static VAR Compensation Techniques – Thyristor Switched Reactor and Thyristor Switched Capacitor

Divyam
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
fastrack2030@gmail.com

Ayushi Saxena
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
ayushi.saxena1407@gmail.com

Bharat Singh
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
bharatsinghuae@gmail.com

J. N. Rai
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
jnrai@dce.ac.in

Corresponding Author: Bharat Singh, bharatsinghuae@gmail.com

***Abstract*— The paper makes a comparative study of the two Static VAR Compensation (SVC) techniques - Thyristor Switched Capacitor (TSC) and Thyristor Switched Reactor (TSR). These techniques are defined, examined and compared. The major constituent of the power flow should be active power and a sufficient and minimal amount of reactive power should be allowed to flow in the transmission line. For power system stability and reliability, there arises a need to compensate the reactive power flow. This Reactive Power Compensation has been done on a medium transmission line, which was modelled by scaling down to practically achievable parameters in a laboratory setup, for different load values. Thus, the paper also enables appropriate and relevant selection of the FACTS Controller based on the type and rating of load and also provides scope for implementation in the laboratory.**

Keywords—flexible AC transmission system, static VAR compensation, modelling of transmission line, thyristor switched reactor, thyristor switched capacitor

Acronyms

- ***FACTS (Flexible AC Transmission System)***
- ***SVC (Static VAR Compensation)***
- ***TSR (Thyristor Switched Reactor)***
- ***TSC (Thyristor Switched Capacitor)***

I. INTRODUCTION

FACTS is a system made up of power electronic components and power system components to improve the stability and controllability of the power transmission line system and improve its power transferring capability. The introduction of the power electronic devices such as the thyristor created new avenues for the invention of devices such as FACTS controllers. In short, the FACT system is utilised to improve the controlling capability of the power system transmission network by using power electronic devices and by introducing capacitive or inductive power in the power system.

In conventional AC power transmission networks, the capability of transferring AC power is restricted due to the effect of several factors such as voltage limit, short circuit current limit, transient stability limit, thermal limits etc. These factors provide the maximum limit for the electrical power that can be effectively transferred via the transmission lines without making any harm to the transmission lines and the electrical equipment involved. This is done typically by bringing about modifications in the power transmission system structure. Although, this is not an easy task in the modern power system. Thus, alternative method of getting highest power transmission capacity without any modifications in the power system structure is preferred. With the use of inconstant impedance electrical components such as inductors and capacitors, the entire energy or power is not able to reach the load from the source. Instead, a part of this is stored in such devices in the form of reactive power and reaches back to the source. This enables us to meet the reactive power demand of the load locally at the receiving end itself, and thus, the substantial quantity of power transmitted to the load or the active power is always lower than the apparent power. This also helps in maintaining the receiving end voltages at desirable values and increases the reliability of the power system. For optimal power transmission, the real power should always be equal to the total power or the power factor should be unity. This is where the role of reactive power compensation comes in achieving FACTS. [3]

II. STATIC VAR COMPENSATION

SVC is considered as the one of the earliest generations of FACTS controllers. The compensator is made up of a rapid thyristor regulating a shunt capacitive bank and an inductor to impart compensation in shunt. They are made up of shunt associated inconstant impedance components such as capacitor or inductor whose output can be modified via thyristor switches, to incorporate these inductive or capacitive reactance into the power line. They may also be located in the centre of the transmission line to enhance the maximal power transmission capacity and at the terminal of the transmission line to make up for the fluctuations arising out of different loading. The two types of SVCs are-

A. TSR

It is basically made up of a inductor which is shunt connected and the impedance of which is controlled using a Thyristor switch in a stepwise manner. TSR comprises of a reactor in series with bidirectional thyristor switches. This method is used to adjust the received reactive power by the load in a system with varying conditions. The difference in the sending and receiving reactive power leads to the difference in sending and receiving end voltages, TSR helps to reduce the difference between these two voltages and improves the voltage regulation of a power system. The inductor absorbs

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:54:43 UTC from IEEE Xplore. Restrictions apply.

the excess lagging VARs present in a capacitive system and maintains a balance of the reactive power, thus equalizing voltages at both the ends of the transmission line. [4]

B. TSC

It is basically made up of a capacitor which is shunt connected and the impedance of which is controlled using a Thyristor switch in a stepwise manner. TSC comprises of a capacitor in series with bidirectional thyristor switches. This method is used to adjust the received reactive power by the load in a system with varying conditions. The difference in the sending and receiving reactive power leads to the difference in sending and receiving end voltages, TSC helps to reduce the difference between these two voltages and improves the voltage regulation of a power system. The capacitor present in TSC provides the excess lagging VARs present in an inductive system and maintains a balance of the reactive power, thus equalizing voltages at both the ends of the transmission line. [5]

III. METHODOLOGY

A. Modelling of Transmission Line [1]

A Π-model of transposed medium transmission line with length 200 km, voltage level of 220 kV and the rating of 500 MVA is chosen for the analysis in order to inculcate the effect of both TSC and TSR and to help in their comparison. The transmission line model used is shown in Fig. 1 and Fig. 2. [8]

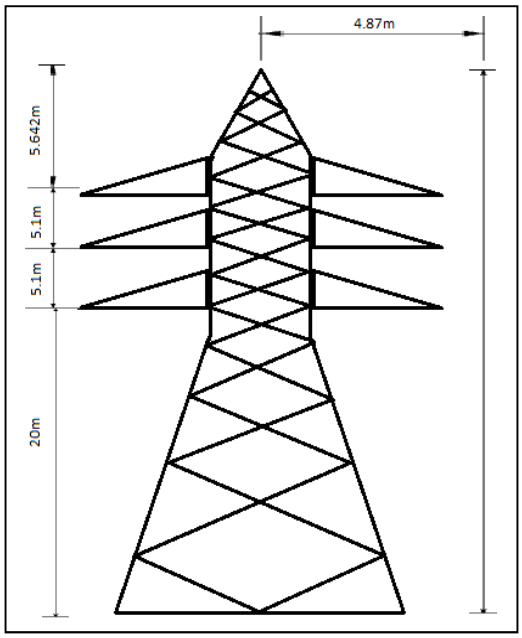


Fig. 1. Transmission Line Tower Design

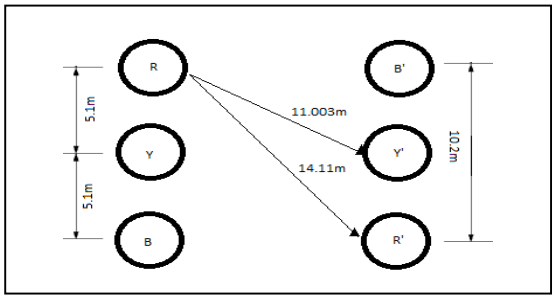


Fig. 2. Transposed Model for Transmission Line

Calculating Geometric Mean Distance (GMD), [7]

$D_{m1} = D_{m3} = [(5.1)(10.2)(9.75)](11.003)]^{1/4}$ m

$D_{m2} = [(11.003)(11.003)(5.1)(5.1)]^{1/4}$ m

Thus, $GMD = [D_{m1}D_{m2}D_{m3}]^{1/3} = 8.24$ m (1)

where D_{m1}, D_{m2} and D_{m3} are the mutual geometric mean distances for the three phases.

Calculating Geometric Mean Radius (GMR), [7]

$D_{s1} = D_{s3} = [(14.11)(0.7788 \times 7.5 \times 10^{-3})]^{1/2}$ m

$D_{s2} = [(9.75)(0.7788 \times 7.5 \times 10^{-3})]^{1/2}$ m

Thus, $GMR = [D_{s1}D_{s2}D_{s3}]^{1/3} = 0.27$ m (2)

where D_{s1}, D_{s2} and D_{s3} are the self-geometric mean distances for the three phases.

Inductance per phase of transmission line from (1) and (2),

$L_{ph} = \frac{\mu}{2\pi}\ln(\frac{GMD}{GMR})$ H/m = 0.683mH/km [7]

Capacitance per phase of transmission line from (1) and (2),

$C_{ph} = \frac{2\pi\varepsilon}{\ln(\frac{GMD}{GMR})}$ F/m = 16.26 nF/km [7]

B. Transmission Line Ratings

Length, L = 200 km
Voltage, V = 220 kV
Frequency, f = 50 Hz
Power, S = 500 MVA
L_{ph} = 0.683 mH/km
C_{ph} = 16.26 nF/km

C. Per Unit Calculations for Transmission Line Parameters

Current in line, $I = \frac{S}{V} = 2.2727$ kA

Base impedance, $Z = \frac{V}{I} = 96.8\ \Omega$

Total inductance for 200 km,

L = (0.683 x 200) mH = 136.6 mH

Total capacitance for 200 km,

C = (16.26 x 200) nF = 3.252 µF

$X_L = \omega L = 2\pi fL = 42.91415\ \Omega$

$X_C = \frac{1}{\omega C} = 978.81268\ \Omega$

X_L (p.u.) $= \frac{XL}{Z} = 0.44332$ p.u.

X_C (p.u.) $= \frac{Xc}{Z} = 10.1117$ p.u.

Y_C (p.u.) = 0.098895 p.u.

D. Scaling down Transmission Line Parameters for 5kVA, 220V model

Current in line, $I' = \frac{S\prime}{V\prime} = 2.2727$ kA

Base impedance, $Z' = \frac{V\prime}{I\prime} = 9.68\ \Omega$

X_L (actual) = X_L (p.u.) x Z' = 4.2913 Ω

L (actual) $= \frac{XL}{\omega} = 13.66$ mH

X_C (actual) = X_C (p.u.) x Z' = 97.88125 Ω

C (actual) = ω X_C = 32.52 µF

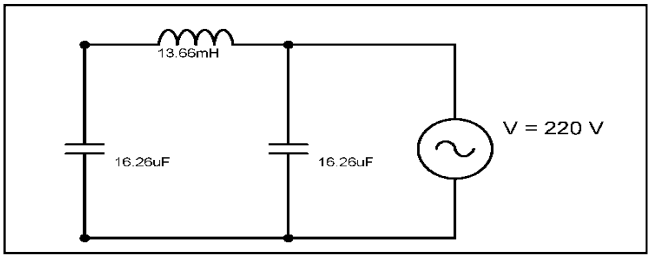


Fig. 3. Transmission Line Model at 5kVA, 220V

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:54:43 UTC from IEEE Xplore. Restrictions apply.

E. ABCD Parameters of Transmission Line [7]

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} \frac{YZ}{2}+1 & Z \\ Y\left(\frac{YZ}{4}+1\right) & \frac{YZ}{2}+1 \end{bmatrix}$$

$$= \begin{bmatrix} 1.02192 & j0.44332 \\ j0.099978 & 1.02192 \end{bmatrix}; \text{A} = 0^0, \text{B} = 90^0$$

F. Load Values

In order to achieve a complete and effective comparative analysis, both the leading and lagging loads are considered having a fixed power rating, i.e. 200 W, which is practically feasible to implement, for various power factor values.

TABLE I. LOAD VALUES FOR LAGGING LOADS

S.No.	Load Values	R (in Ω)	L (in H)
1.	0.1 pf lag, 200 W	2.42	0.0766
2.	0.2 pf lag, 200 W	9.68	0.1509
3.	0.3 pf lag, 200 W	21.78	0.2204
4.	0.4 pf lag, 200 W	38.72	0.2824
5.	0.5 pf lag, 200 W	60.5	0.3355
6.	0.6 pf lag, 200 W	87.11	0.3697

TABLE II. LOAD VALUES FOR LEADING LOADS

S.No.	Load Values	R (in Ω)	C (in µF)
1.	0.1 pf lead, 200 W	2.42	132.195
2.	0.2 pf lead, 200 W	9.68	67.123
3.	0.3 pf lead, 200 W	21.78	45.9613
.	0.4 pf lead, 200 W	38.72	35.8785
5.	0.5 pf lead, 200W	60.5	30.376
6.	0.6 pf lead, 200 W	87.11	27.4028

G. Equations Involved [7]

The real and reactive power at receiving end of the transmission line can be given as:

$$\text{P}_\text{R} = \left|\frac{VsVr}{B}\right| \cos(\beta - \delta) - \left|\frac{AVr^2}{B}\right| \cos(\beta - \alpha) \quad (1)$$

$$\text{Q}_\text{R} = \left|\frac{VsVr}{B}\right| \sin(\beta - \delta) - \left|\frac{AVr^2}{B}\right| \sin(\beta - \alpha) \quad (2)$$

For the transmission line under analysis, $\beta = 90^0$, $\alpha = 0^0$, $V_R = 1$ and $V_S = 1$, and P_R can be calculated from the load values used. Thus, δ can be calculated from (1). This value can be substituted in (2) to find Q_R.

Also, lagging VARs requirement of the transmission line with load is $|Q_R|$ and

$$Q_{cll} = - Q_R$$

Reactive power of load,

$$|Q_{load}| = P_R \tan \varphi$$

where cos φ is the power factor of the load.

For lagging loads, there is reactive power requirement and thus, a shunt capacitor is required for providing this reactive power and thus to compensate the reactive power.

Lagging VARs (in p.u.) provided by the shunt capacitor,

$$Q_c = |Q_{load}| - |Q_{cll}|$$

Thus, value of required capacitor,

$$\text{C}_\text{reqd} = \frac{|Qc| \times 5000}{\omega V^2} \text{ F} \quad (3)$$

For leading loads, there is a reactive power supply and thus these surplus lag VARs have to be absorbed by the shunt reactor and thus to compensate the reactive power.

Lagging VARs (in p.u.) provided by the shunt reactor,

$$Q_L = |Q_{load}| + |Q_{cll}|$$

Thus, value of required reactor,

$$\text{L}_\text{reqd} = \frac{\omega |Q_L| \times 5000}{V^2} \text{ H} \quad (4)$$

IV. SIMULATIONS

The model was simulated for all lagging, leading and no load values as in Table I and Table II. In the simulations, a triac has been used instead of back to back thyristor, which gives similar effect as two back to back thyristor with common gate to provide the bidirectional flow of power. Fig. 4, Fig. 5 and Fig. 6 show the snapshot of the simulation diagram for lagging load, no lead and leading load respectively.

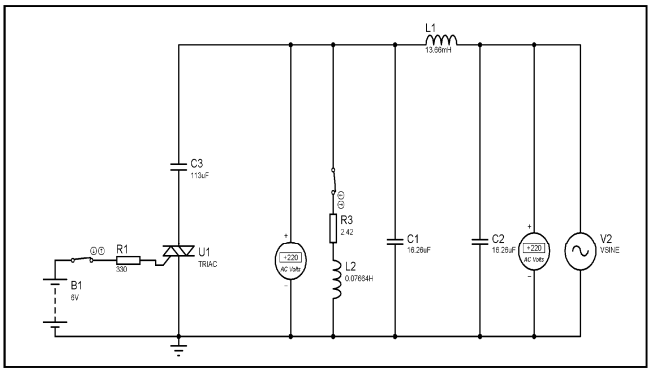


Fig. 4. Simulation snapshot for 0.1 pf lag, 200 W load using TSC

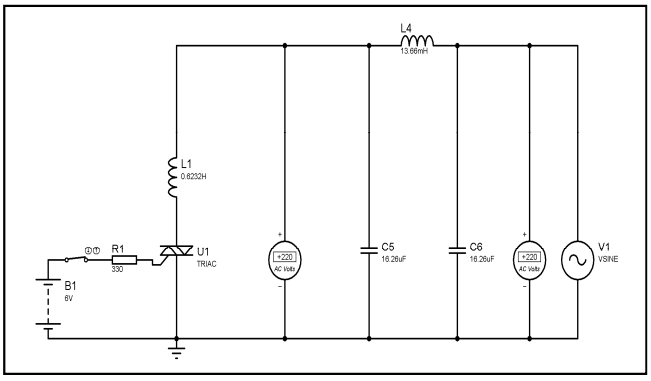


Fig. 5. Simulation snapshot for no load using TSR

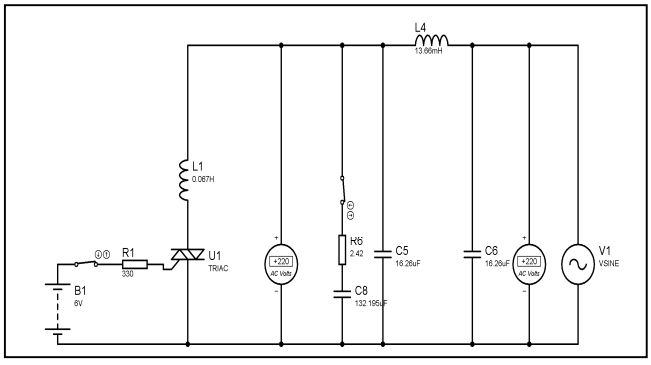


Fig. 6. Simulation snapshot for 0.1 pf lead, 200 W load using TSR

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:54:43 UTC from IEEE Xplore. Restrictions apply.

V. CALCULATIONS

The required $|Q_c|$ and $|Q_L|$ has been calculated for the load values and corresponding C_{reqd} and L_{reqd} has been calculated using (3) and (4) for each of the load values. After obtaining the required values of capacitor and inductor for compensating each of the loads, the voltage before and after compensation is noted through simulation.

TABLE III. CALCULATIONS FOR C_{REQD} FOR LAGGING LOADS

Load	Load Values		$\|Q_C\|$ (p.u.)	C_{reqd} (µF)	Compensation	
					Before	After
	R (Ω)	L (H)			V_R	V_R
0.1 pf lag, 200 W	2.42	0.0766	0.3481	113.0	191	220
0.2 pf lag, 200 W	9.68	0.1509	0.1461	48.062	207	220
0.3 pf lag, 200 W	21.78	0.2204	0.0773	25.449	213	220
0.4 pf lag, 200 W	38.72	0.2824	0.0418	13.762	216	220
0.5 pf lag, 200 W	60.5	0.3355	0.0194	6.4067	218	220
0.6 pf lag, 200 W	87.11	0.3697	0.0035	1.1622	220	220

TABLE IV. CALCULATIONS FOR L_{REQD} FOR NO LOAD

Load	$\|Q_L\|$ (p.u.)	L_{reqd} (H)	Voltage (V)			
			Before Compensation		After Compensation	
			VS	VR	VS	V_R
No Load	0.049445	0.6232	220	225	220	220

TABLE V. CALCULATIONS FOR L_{REQD} LEADING LOADS

Load	Load Values		$\|Q_L\|$ (p.u.)	L_{reqd} (H)	Compensation	
					Before	After
	R (Ω)	C (µF)			V_R	V_R
0.1 pf lead, 200 W	2.42	132.195	0.4477	0.067	276	220
0.2 pf lead, 200 W	9.68	67.123	0.2457	0.1253	248	220
0.3 pf lead, 200 W	21.78	45.9613	0.1769	0.1740	239	220
0.4 pf lead, 200 W	38.72	35.8785	0.1414	0.178	235	220
0.5 pf lead, 200 W	60.5	30.376	0.1190	0.2587	232	220
0.6 pf lead, 200 W	87.11	27.4028	0.1031	0.2987	230	220

VI. OBSERVATIONS

A. Lagging Loads

TABLE VI. OBSERVATIONS FOR LAGGING LOADS

pf	Q_{load} (VARs)	Q_C (VARs)
0.1 lag	1989.97	1740.975
0.2 lag	979.79	730.800
0.3 lag	635.96	386.965
0.4 lag	458.25	209.265
0.5 lag	346.41	97.415
0.6 lag	266.67	17.67

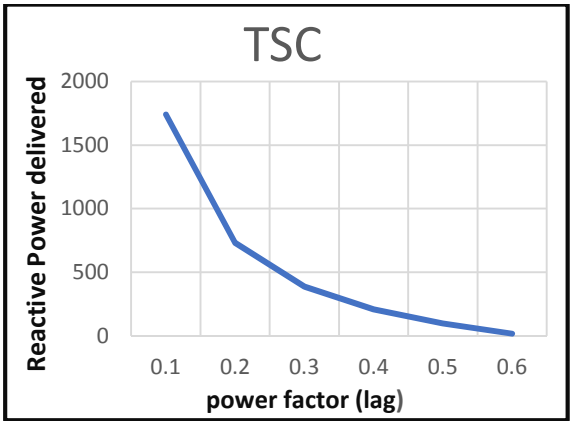


Fig. 7. Reactive Power Delivered Vs Power Factor for lagging loads

From Table VI and Fig. 7, as shown above, it is observed that for a given real power demand (200W), as the power factor of the load is increased from 0.1 to 0.2, the reactive power demand of the load decreases by 1010.17 VARs. Further, as the power factor increases from 0.5 to 0.6, the reactive power demand decreases by 79.74 VARs. Therefore, the requirement of the compensating reactive VARs to be supplied by the TSC decreases as the power factor increases.

As the power factor improved to 0.6 pf lag then the reactive power demand of the lagging load became almost equal to the lagging power supplied by the capacitive transmission line and hence the requirement of the compensation is almost negligible (17.67 VARs), and is depicted by the graph in Fig. 7, tending to zero.

B. No Load

Under no load conditions, it is observed from Table IV, that for proper compensation, an inductor is required. This proves that a medium transmission line, having a length of 200km has a dominating capacitive nature, and due to this capacitance effect, surplus reactive power is injected in the transmission line and is responsible for the increase in the voltage level of receiving end. This is phenomenon of increase in the receiving end voltage due to capacitance effect of the transmission line is known as Ferranti Effect.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:54:43 UTC from IEEE Xplore. Restrictions apply.

C. Leading Loads

TABLE VII. OBSERVATIONS FOR LEADING LOADS

pf	Q_{load} (VARs)	Q_L (VARs)
0.1 lead	1989.97	2238.97
0.2 lead	979.79	1228.75
0.3 lead	635.96	884.995
0.4 lead	458.25	707.252
0.5 lead	346.41	595.405
0.6 lead	266.67	515.66

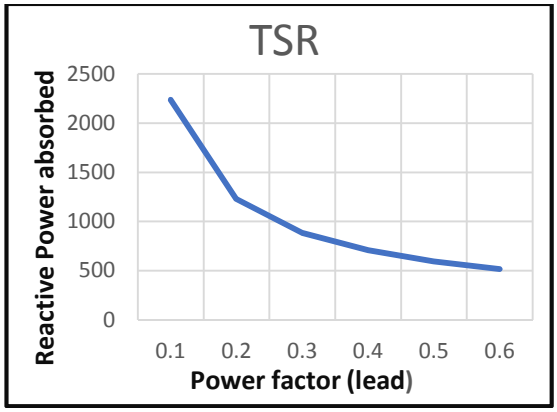


Fig. 8. Reactive Power Absorbed Vs Power Factor for leading loads

From Table VII and Fig. 8 as shown above, it is observed that for a given real power demand (200W), as the power factor of the load is increased from 0.1 to 0.2, the reactive power absorbed decreases by 1010.2 VARs. Further, as the power factor increases from 0.5 to 0.6, the reactive power absorbed decreases by 79.74 VARs. Therefore the requirement of compensating reactive VARs to be absorbed by the TSR decreases as the power factor increases.

VII. CONCLUSION

The paper presented the comparison of FACTS Static VAR Controllers, i.e. TSC and TSR on a medium transmission line for varying loads (No Load, Lagging and Leading loads) and with the help of modelling and mathematical calculations, in order to compensate the reactive power, thereby attaining equal sending end and receiving end voltage levels. We can also conclude that the above chosen transmission line depicted Ferranti Effect and hence TSR is required to compensate this effect. The method employed to achieve the improvements outlined are, introduction of a TSR to absorb the reactive power and TSC to deliver the reactive power demand. This improves the capacity of an already built transmission line and is used to meet the increasing load demands, by locally meeting the reactive power demands at the receiving end itself, and hence achieving the goal of flexible transmission and voltage regulation.

The above comparative study can be experimentally implemented and verified in a laboratory setup. The accuracy of the experimental results in the laboratory setup may vary on the availability of the exact capacitor and inductor rating. Though the above study has been implemented for a medium transmission line, in future, the study can be extended for short and long transmission lines as well.

ACKNOWLEDGMENT

We would like to thank Prof J.N. Rai, for his constant motivation and guidance during the project. Without his timely help and support, the completion of this paper wouldn't have been possible. We would also like to thank the publisher of the research papers from where we have taken reference.

REFERENCES

[1] Mr.Vinayak Gaikwad, Suyash Tripathi, Yash Wanjari, Ankesh Thakre, Diksha Dakare, Jayant Shendre, "Laboratory Setup For Long Transmission Line", International Research Journal of Engineering and Technology (IRJET), Volume: 04 Issue: 03, Mar -2017.

[2] Jai Damania, Karuna Nikum, Shoeb Shaikh, Hardik Save, "Flexible AC Transmission System using TCR", International Journal of Engineering Research & Technology (IJERT) 2017.

[3] Amaranath Sanyal, Archita Ghosh, Raju Basak, Sourish Sanyal, Subhasree Chatterjee, Deepayan Biswas "An Article On Flexible AC Transmission System" IRJET Volume 3, Issue 5, May 2016.

[4] R.T Kalapad, "Review on FACTS devices integrated with TSR" International Journal of Engineering Research and Applications (IJERA), October 2015.

[5] Aashish Goyal, Neharika Kapil, Sheila Mahapatra, "Implementation of Thyristor Switched Capacitors for Power Factor Improvement" Advance in Electronic and Electric Engineering. ISSN 2231-1297, Volume 4, Number 3 (2014).

[6] Rudra Rameshwar, "Efficient, Reliable and Cost Effective Transmission of Electrical Energy using FACTs", National Power System Conference (NPSC), 2004.

[7] "Power System Analysis" , John Grainger, William Stevenson, Jr.

[8] R. Saravanan, "Transmission Line Design Concept", SlideShare.

[9] Rasal Madhuri B., Tagad Bharati T., Surwase Komal R., Prof. Tapre.P.C, "Ferranti Effect Overcome By Using Thyristor Switched Reactor (TSR) In FACTS" IJARIIE-ISSN(O)-2395-4396, Vol-5 Issue-3 2019.

[10] "Modern Power System Analysis", Fourth Edition, D. P. Kothari, I. J. Nagrath.

[11] Jyotideep Deka, Kamaldeep Sharma, Krishan Arora, "Study Of 132 Kv Transmission Line Design And Calculation Of Its Parameters", International Journal for Electrical and Electronics Engineers, Vol. 8, Issue-01, January-June, 2016.

[12] Rishabh Jain , Sunil Kumar Sahu , C Venkata Srikar , Rohit , Suvit Chitnis , Mr. Mousam Sharma, "FACTS (Flexible AC transmission system) by using TSR (Thyristor Switch Reactance)", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 6, Issue 4, April 2017.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:54:43 UTC from IEEE Xplore. Restrictions apply.

Air Quality, Atmosphere & Health
https://doi.org/10.1007/s11869-020-00881-z

Comparative study on air quality status in Indian and Chinese cities before and during the COVID-19 lockdown period

Aviral Agarwal[1] · Aman Kaushik[1] · Sankalp Kumar[1] · Rajeev Kumar Mishra[1]

Received: 26 May 2020 / Accepted: 15 July 2020
© Springer Nature B.V. 2020

Abstract

Amidst COVID-19 pandemic, extreme steps have been taken by countries globally. Lockdown enforcement has emerged as one of the mitigating measures to reduce the community spread of the virus. With a reduction in major anthropogenic activities, a visible improvement in air quality has been recorded in urban centres. Hazardous air quality in countries like India and China leads to high mortality rates from cardiovascular diseases. The present article deals with 6 megacities in India and 6 cities in Hubei province, China, where strict lockdown measures were imposed. The real-time concentration of $PM_{2.5}$ and NO_2 were recorded at different monitoring stations in the cities for 3 months, i.e. January, February, and March for China and February, March, and April for India. The concentration data is converted into AQI according to US EPA parameters and the monthly and weekly averages are calculated for all the cities. Cities in China and India after 1 week of lockdown recorded an average drop in $AQI_{PM2.5}$ and AQI_{NO2} of 11.32% and 48.61% and 20.21% and 59.26%, respectively. The results indicate that the drop in AQI_{NO2} was instantaneous as compared with the gradual drop in $AQI_{PM2.5}$. The lockdown in China and India led to a final drop in $AQI_{PM2.5}$ of 45.25% and 64.65% and in AQI_{NO2} of 37.42% and 65.80%, respectively. This study will assist the policymakers in devising a pathway to curb down air pollutant concentration in various urban cities by utilising the benchmark levels of air pollution.

Keywords Air quality index · China · COVID 19 · India · NO_2 · $PM_{2.5}$

Introduction

In the present time, with the emergence of rapid globalisation and urbanisation, megacities in developing nations are facing severe health issues due to ambient air pollution. According to WHO (World Health Organization), seven million people die each year because of exposure to polluted air (UN Environment Programme 2018). Numerous epidemiological studies in the past two decades have highlighted outdoor air pollution as a cause of various respiratory diseases such as asthma, premature deaths and cardiovascular diseases. These have been identified as primary causes of mortality. In such cases, the population living in the vicinity of major roadways in metropolitan cities suffers the most (Park et al. 2020). In urban areas, 80% of people live in concentrations exceeding the WHO limits (Błaszczyk et al. 2017). Motor-vehicle emitted compounds in urban areas which include carbon monoxide (CO); nitrogen oxides (NO_x); coarse (PM_{10}), fine ($PM_{2.5}$), and ultrafine ($PM_{0.1}$) particle mass, black carbon, polycyclic aromatic hydrocarbons and benzene which are found in elevated concentrations as reported by Venkatram and Schulte (2018). Moreover, studies suggest that particulate matter and NO_2 levels are higher in cities with greater transportation activity and urban backgrounds (Rodríguez et al. 2016). PM_{10} and $PM_{2.5}$ are the two primary particulate matters monitored all over the world. However, $PM_{2.5}$ possesses a higher health risk as compared with PM_{10} because of its high retention time and ability to penetrate deep into the lungs and enter the bloodstream (US EPA 2018). The WHO ambient air quality guidelines suggest an annual mean $PM_{2.5}$ concentration limit of 10 μg/m^3 and 25 μg/m^3 for the 24-hourly mean. The NO_2

✉ Rajeev Kumar Mishra
rajeevkumarmishra@dtu.ac.in

Aviral Agarwal
agarwalaviral1999@gmail.com

Aman Kaushik
kakukaushik@gmail.com

Sankalp Kumar
sankalp.vibhu@gmail.com

[1] Department of Environmental Engineering, Delhi Technological University, Delhi 110042, India

Published online: 23 July 2020

limit is 40 μg/m^3, and 200 μg/m^3 for the annual and 1-h mean, respectively (World Health Organization 2005).

At the dawn of twenty-first century, developing economies like India and China are undergoing rapid industrialisation and modernisation, which are leading to hazardous levels of air pollution similar to the Industrial Revolution in Europe. It is well-understood that megacities in both countries like Beijing, Shenyang, Taiyuan, New Delhi, Mumbai, and Chennai are the world's most polluted cities (Zhu 2005). The primary sources of air pollution in India have been identified as vehicular emissions, industrial emissions, coal combustion, biomass burning, road dust, and refuse burning (Pant and Harrison 2012). Likewise, the poor air quality in China is a matter of global concern. The air pollution caused by transportation and industries is a serious environmental issue in urban settlements, and 50% of the PM in the urban air comes from traffic emissions (Li et al. 2017). Kumar and Joseph (2006) analysed ambient and kerb site air pollution correlation of PM_{10}, $PM_{2.5}$, and NO_2 in Mumbai, India. The results indicated a strong correlation between $PM_{2.5}$ and NO_2 at the ambient site due to vehicular emission as a result of high traffic density. A 2016 report estimated that only 3% of the Chinese population and less than 1% of the Indian population have exposure to $PM_{2.5}$ concentrations complying with WHO guidelines (IEA OECD 2016), although, an overall monotonic decrease in air pollutants was recorded in China from 2015 to 2018 (Fan et al. 2020). The existing levels of ambient $PM_{2.5}$ and NO_2 are above the safe limit. A 2017 report on global air pollution stated that China and India contribute to 52% of global $PM_{2.5}$—attributable deaths (1.525 million deaths) (Health Effects Institute 2019). In the past decades, both countries have been pro-active towards the efforts reducing air pollution. However, no long-term solution has been identified yet. Numerous academic studies have been conducted in both nations regarding the growing air pollution and its health effects. Kumar and Mishra (2018) conducted an assessment of major air pollutants at 36 transport corridors in Delhi, India, and the results of the study concluded that 31 corridors had "severe" and "very poor" AQI, and high traffic volume in most corridors is characterised by traffic-induced human health risks. It has been found that high levels of ambient $PM_{2.5}$ and NO_2 increase the risk of cardiovascular diseases and lung cancer in humans (Liu et al. 2018; Siddique et al. 2010).

In late December 2019, there was an outbreak of a highly contagious disease caused by the novel coronavirus, SARS-CoV-2. The first case emerged from Wuhan City, Hubei Province, China. The disease has been identified as Coronavirus disease (COVID-19). It's an outbreak, recognised as a "Pandemic" by WHO, has been extensively worldwide and exponential with more than 200 countries and territories reporting 3,267,936 cases and 234,703 deaths (7.18%) as of April 30, 2020 (WHO 2020). Individuals with underlying health problems, weak immunity, and the elderly are most likely to become extreme cases (Chen et al. 2020a). The critical sources of infection are patients infected with the novel coronavirus and those with asymptomatic infection (Wang et al. 2020a). Studies indicate correlation between the long-term exposure to air pollutant and COVID 19 death rate. Cities with hazardous air quality face a serious threat from the pandemic (Wu et al. 2020b; Conticini et al. 2020). Therefore, in the absence of a vaccine or treatment available for COVID-19, there has been a coordinated global response of imposing "lockdown" measures on citizens. As of now, more than a third of the worldwide population is under restriction. India recorded 34,867 COVID-19 cases (as of 30-04-2020) and a nationwide lockdown was imposed in India on March 25, 2020 for 21 days (MoHFW 2020). The lockdown constrained people from stepping out of their homes. Transport services, road, air, and rail, were suspended along with institutions and industrial establishments except for essential goods and services (Jain and Sharma 2020) and has been extended up to May 3, 2020. In parallel, 82,862 cases (as of 30–04-20) were recorded in China and 82% of these cases were recorded in Hubei Province (National Health Commission of the People's Republic of China 2020).

In Hubei province, Wuhan was regarded as the epicentre (Zhang et al. 2020) of the virus. On January 23, 2020, Hubei province, China, was kept under community quarantine with the shutdown of public transport, educational institutes, business centres, parks, and other social contacts to slow down the spread of COVID-19 (Wilder-Smith and Freedman 2020).

Further restrictions in Hubei province were lifted on March 23, 2020.With the implementation of lockdown and other federal restrictions in various countries around the globe, a visible reduction in air pollution is found in megacities. This study is aimed at quantifying and analysing the reduction in air pollution due to the lockdown imposed in two overpopulated and highly polluted countries of the world, viz. China and India, to determine the effect of lockdown on the air quality in an urban environment. The results of this study will help in gauging the ability of a full lockdown on reducing air pollution. Further, it will help in devising a response plan for unforeseen episodes of the high level of air pollution in urban environments.

Methodology of the study

Site selection

For the present study, 6 cities have been selected, each from India and Hubei Province, China. These locations are selected based on the availability of historical air pollution data, population density, monitoring station network, and the number of positive COVID-19 cases per million people. Selected cites

with their population, the number of monitoring stations taken into account, COVID-19 cases per million people, their geographical coordinates, and start and end date of lockdown are given in Table 1.

Parameters for analysis

For the analysis of the effect of the lockdown imposed by the governing authorities on the air quality, $PM_{2.5}$ and NO_2 are selected as parameters of the study. Both of these pollutants have a direct relationship with various anthropogenic activities that were restricted due to the lockdown (US EPA 2018; Ministry for the Environment New Zealand 2020). Hence, analysing these parameters assist in espying the effect of lockdown on the air quality of the selected locations.

Data collection and interpretation

During the data collection, 24-h average concentration (μg/m^3) data is taken for $PM_{2.5}$, and hourly average concentration data is taken for NO_2 (ppb) from respective EPAs of the locations selected. The data is collected in China for 13 weeks starting from January 1, 2020 except for Wuhan for which data was collected for 15 weeks since the lockdown was imposed till April 8, 2020. In India, data is collected for 13 weeks starting from February 1, 2020 to April 30, 2020. For the years 2016–2019, the data for Hubei Province, China has been collected from January 1 to March 31, and similarly, for India, the data has been collected from February 1 to April 30.

Weekly average data of $PM_{2.5}$ and NO_2 for the mentioned months has been calculated for the selected monitoring stations. The average value of the $PM_{2.5}$ and NO_2 concentration in a city is calculated by taking an average of all the monitoring stations selected, located at various distant locations throughout the city. The average value of the concentration of $PM_{2.5}$ and NO_2 is converted to individual AQI ($AQI_{PM\ 2.5}$ and AQI_{NO2}) by using the protocol suggested by US EPA for reporting the air quality data using the Air Quality Index (AQI) (Mintz 2012).

To analyse the changes in the $AQI_{PM2.5}$ and AQI_{NO2} levels, for each city, various drop percentages are calculated. The immediate drop percentage is calculated by the difference in average $AQI_{PM2.5}$ and AQI_{NO2} of the weeks before and after the lockdown was enforced. The final $AQI_{PM2.5}$ and AQI_{NO2} drop percentages are calculated by the difference in average $AQI_{PM2.5}$ and AQI_{NO2} of the week before lockdown and the last week when lockdown restrictions were lifted. Five-year and 1-year $AQI_{PM2.5}$ drop percentages in the year 2020 are calculated for January, February, and March in China and February, March, and April in India. It has been calculated by the difference in average $AQI_{PM2.5}$ of months of years 2016 and 2019 to the same months of 2020 for 5-year and 1-year drop percentages, respectively.

Results and discussion

With the parameters of immediate and final $AQI_{PM2.5}$ and AQI_{NO2} drop percentages, and 5 year and I year drop percentages of $AQI_{PM2.5}$, analysis for the cities in China and India has been done followed by a comparative assessment between the two countries.

Air quality analysis for the selected cities of China

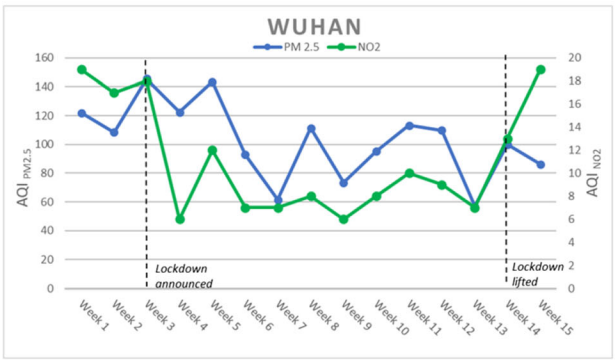
Due to a large number of reported cases, Wuhan and its neighbouring cities (Huanggang and Ezhou) implemented a lockdown on January 23, 2020 followed by several cities on

Table 1 General Information of the Selected Cities (Office of the Registrar General and Census Commissioner 2011; National Bureau of Statistics of China 2010; Central Intelligence Agency 2018; MoHFW 2020; National Health Commission of the People's Republic of China 2020)

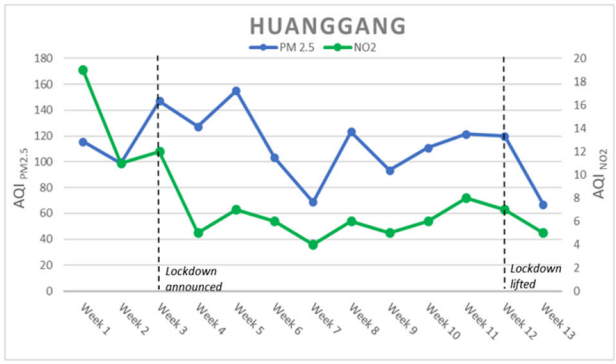
Country	City	Population (in million)	No. of stations monitored	COVID-19 cases per Million	Geographical coordinates
China	Xiangyang	5.89	5	199.22	30° 48′ 01″ N 110°23′11″ E
	Jingzhou	0.97	3	1624.47	30°13′35" N 111° 47′ 18″ E
	Huanggang	6.628	2	438.55	30° 24′ 16″ N 114° 42′ 49″ E
	Xiaogan	5.17	3	679.26	31° 03′ 41″ N 113° 25′ 37″ E
	Wuhan	8.11	5	6204.90	30° 42′ 07″ N 113° 46′ 52″ E
	Yichang	4.37	5	213.18	30° 41′ 49″ N 110° 48′ 01″ E
India	Delhi	18.62	5	501.03	29° 03′ 55″ N 76° 06′ 09″ E
	Lucknow	3.12	5	88.30	26° 52′ 53″ N 80° 41′ 49″ E
	Kolkata	4.98	5	240.51	23° 04′ 44″ N 87° 17′ 22″ E
	Mumbai	13.80	5	1279.98	19° 10′ 38″ N 72° 23′ 50″ E
	Chennai	5.16	5	1153.44	13° 07′ 10″ N 79° 44′ 05″ E
	Jaipur	6.42	4	234.51	27° 03′ 33″ N 75° 18′ 19″ E

January 24, 2020 (Wu et al. 2020a). Wuhan, having the highest number of cases than any other city in China enforced lockdown with strict federal orders restraining anthropogenic activity to minimal level (Lu 2020), which entailed the highest immediate AQI_{NO2} (69.35%) and $AQI_{PM2.5}$ drop (15.95%) among the selected cities. In contrast, Xiangyang recorded the lowest immediate and final AQI_{NO2} drop of − 3.22% and − 16.40%, respectively. This trend is observed since Xiangyang is one of the most industrialised cities in central China, and due to the high demand for PPE kits and testing equipment, industries were working at double shifts to meet the demand (Hubei Provincial People's Government 2020).

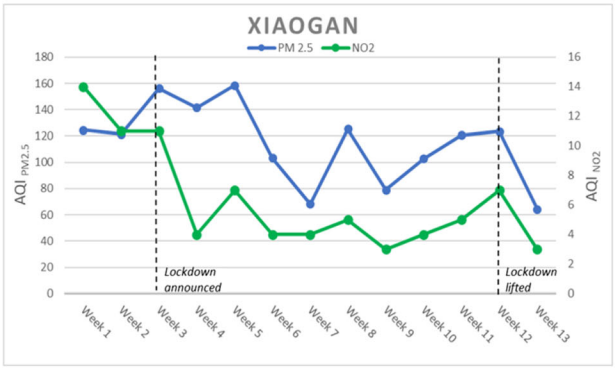
The average immediate and final AQI_{NO2} drop recorded is 48.61% and 26.64%, respectively. It can be concluded that the final AQI_{NO2} drop percentage is lower than the immediate AQI_{NO2} drop percentage; as the lockdown progressed, citizens had to get out of their homes for necessary essential commodities. Every city experienced a drop in $AQI_{PM2.5}$ (Fig. 1a, b, c, d, e, and f); on an average, the immediate $AQI_{PM2.5}$ drop is 11.28%; subsequently, the average final $AQI_{PM2.5}$ drop is 26.37%. Before lockdown was implemented, every city had $AQI_{PM2.5}$ within unhealthy for sensitive groups (101–150) and unhealthy (151–200) range according to US EPA standards. Nevertheless, the $AQI_{PM2.5}$ level



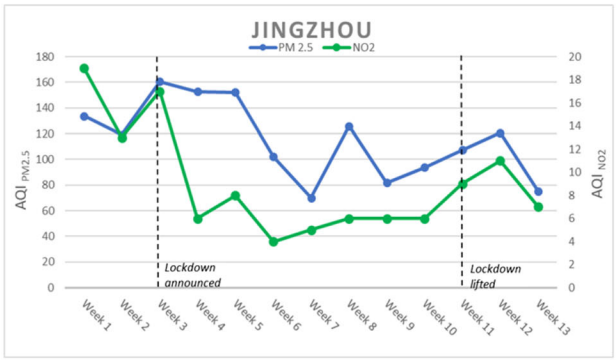
a



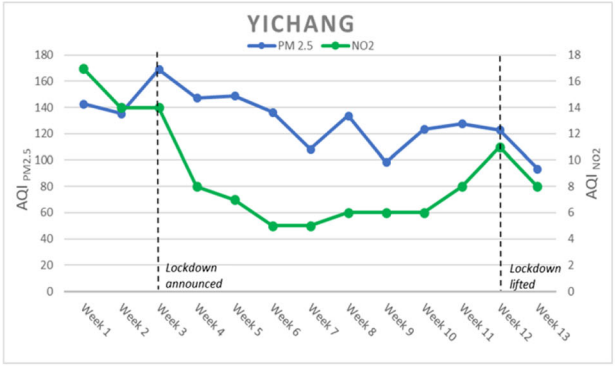
b



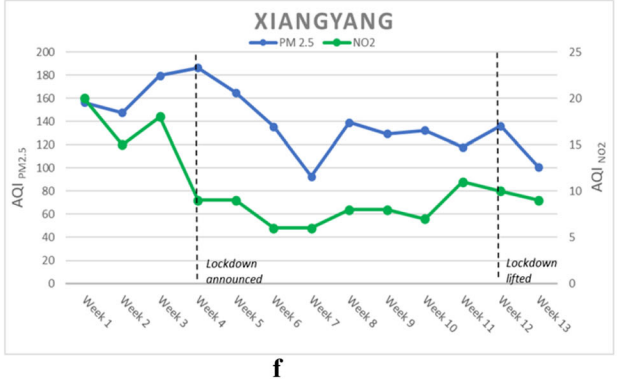
c



d



e



f

Fig. 1 Weekly averages of $AQI_{PM2.5}$ and AQI_{NO2} for cities of China

reduced and reached the moderate category during the lockdown period. Furthermore, for January, February, and March 2020, the cities recorded the lowest $AQI_{PM2.5}$ levels in 5 years (Fig. 2a, b, c, d, e, and f). As the virus started to spread in Hubei province in January, the citizens avoided leaving their homes as a self-precautionary measure. As the lockdown was implemented from January 23, 2020 in various cities, a drastic drop has been recorded in $AQI_{PM2.5}$ levels

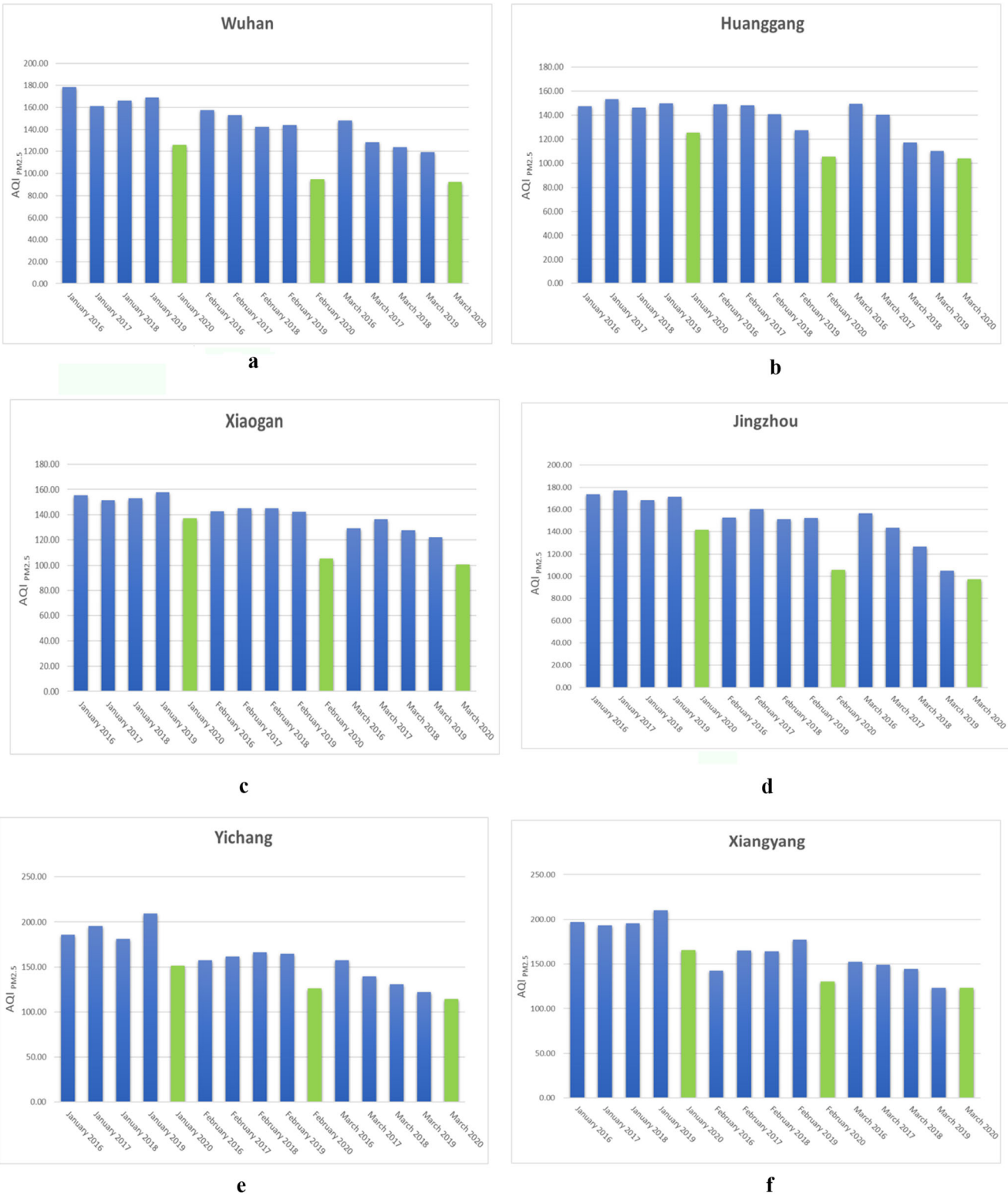


Fig. 2 Past 5 years' averages of $AQI_{PM2.5}$ for cities of China

from week 3 to week 4 (Fig. 1a, b, c, d, e, and f). The $AQI_{PM2.5}$ level remained constant through 2016–2019, to a certain degree for January and February. The highest 5-year and 1-year $AQI_{PM2.5}$ drop were recorded in Wuhan: 39.96% and 34.21%, respectively, while the lowest was found in Xiangyang: 8.53% and 26.55%, respectively. The average 5-year $AQI_{PM2.5}$ drop for January, February, and March, are 18.14%, 25.74%, and 29.18%, respectively. Subsequently, the average 1-year $AQI_{PM2.5}$ drop for January, February, and March are 20.17%, 26.31%, and 9.97%, respectively. It can be recorded that the 5-year and 1-year $AQI_{PM2.5}$ drop percentages are comparable in all months except March.

Additionally, meteorological conditions have an essential influence on the variations of $PM_{2.5}$ and NO_2 concentrations in the ambient environment (Agarwal et al. 2006). The cities Wuhan, Huanggang, and Xiaogan recorded a spike of $AQI_{PM2.5}$ and AQI_{NO2} levels (Fig. 1a, b, and c) in week 5 which led to a gradual decline in the forthcoming weeks of $AQI_{PM2.5}$ and AQI_{NO2}. The increase in the AQI levels was due to low precipitation recorded in week 5 in the three cities; total average precipitation for the three cities for week 4 and 5 were 42.6 mm and 0.34mm, respectively. In the subsequent weeks, rainfall intensity increased, which led to the drop in $AQI_{PM2.5}$ and AQI_{NO2} levels; the total average rainfall in weeks 6 and 7 for the three cities were 19.94 mm and 22.07 mm. In week 8, it has been observed from the Fig. 1a, b, c, d, e, and f that there is an increase in $AQI_{PM2.5}$ and AQI_{NO2} levels. Wuhan experienced 4.2mm precipitation in Week 8, as compared with 28.6mm and 39.4mm precipitation in Week 7 and Week 9, respectively. Hence, the abrupt increase in $PM_{2.5}$ and NO_2 in the cities is due to the low precipitation received in central China in Week 8 (19th–25th, February 2020). In the 13th week, a sharp decrease in $AQI_{PM2.5}$ and AQI_{NO2} levels is recorded. The mean rainfall in six cities in the 12th week is 14.23 mm. In contrast, the 13th week recorded heavy precipitation in all cities with mean average rainfall as 66.36 mm.

Air quality analysis for the selected cities of India

India enforced a nationwide lockdown from March 24, 2020, to May 3, 2020, after successive extensions as a preventive measure against COVID-19 pandemic. As the lockdown was implemented from March 24, 2020, a drastic drop has been recorded in $AQI_{PM2.5}$ levels of all selected cities from week 7 to week 8 (Fig. 3a, b, c, d, e, and f). Maharashtra, western peninsular state of India, has recorded the most cases of COVID-19 and deaths, 12,296 and 521, respectively (MoHFW 2020), and among the six megacities of India. Its capital, Mumbai, has shown the highest immediate drop of both AQI_{NO2} and $AQI_{PM2.5}$, i.e. 76.28% and 34.02%, respectively. Kolkata recorded the highest final $AQI_{PM2.5}$ drop (76.67%), and Lucknow recorded the least immediate drop in $AQI_{PM2.5}$ (6.47%) partly due to negligible precipitation in week 8. Chennai experienced the least immediate and final drops of AQI_{NO2}, which are 32.14% and 20.95%, respectively. It can be understood because of Chennai having an already low value of AQI_{NO2} (Fig. 3a) in week 7 and the weeks before the lockdown (Table 2).

The 6 cities experienced an average immediate $AQI_{PM2.5}$ drop of 20.21%, and an average final $AQI_{PM2.5}$ drop of 37.42%. Each one of the six Indian cities in the study recorded an immediate and final AQI_{NO2} drop with the average immediate AQI_{NO2} drop of 59.26% and an average final AQI_{NO2} drop of 65.80%. It has shown an overall drop in both $AQI_{PM2.5}$ and AQI_{NO2}. The average 5-year $AQI_{PM2.5}$ drop for the months of February, March, and April is recorded as 16.05%, 26.68%, and 37.51%, respectively; subsequently, average 1-year $AQI_{PM2.5}$ drop for the months of February, March, and April is 3.48%, 17.98%, and 27.06%, respectively. Chennai recorded the highest 5-year drop and 1-year drop in $AQI_{PM2.5}$ in April 2020 as 59.79% and 42.90%, respectively. All cities, except Mumbai in April 2020 and Chennai in March 2020, recorded the lowest $AQI_{PM2.5}$ levels in March and April 2020 as compared with the past 5 years. Mumbai is the only Indian city in the study to have shown a 1-year rise (1.09%) in an average $AQI_{PM2.5}$ in April 2020 (Fig. 4a, b, c, d, e, and f). A spike in $AQI_{PM2.5}$ was recorded (Table 3) between weeks 9 and 10 in New Delhi. It is as a result of reported fireworks incidents recorded on April 5, 2020, the day-wise $AQI_{PM2.5}$ levels of week 9 and 10 are given in Fig. 3b (The Indian Express 2020).

Furthermore, meteorological factors have an essential factor in the reduction and increase of $PM_{2.5}$ and NO_2 concentrations in the ambient environment. It can be recorded from Fig. 3c, in Kolkata, that $AQI_{PM2.5}$ and AQI_{NO2} have decreased augmented by heavy precipitation in the weeks 11, 12, and 13 with 29.20 mm, 80.90 mm, and 60.20 mm, respectively. However, in the preceding weeks 9 and 10, there was no precipitation. On the other hand, New Delhi, Lucknow, and Jaipur recorded an increase in $AQI_{PM2.5}$ in week 11 due to high surface winds in northern India due to dust storms from western India according to the Ministry of Environment and System of Air Quality and Weather Forecasting and Research (SAFAR) (Fig. 3b, e, and f) (ANI News 2020).

Comparative analysis and discussion between China vs India

An entire month lockdown was implemented in February 2020 in Hubei province, China; likewise in India, the month of lockdown was April 2020. The 1-year drop for February in China comes out to be 26.31%, whereas, for April in India, it is 26.06%. Hence, it can be deduced from the results that an entire month lockdown in urban centres results in a drop of around 26% in $AQI_{PM2.5}$ if compared with previous year



Fig. 3 Weekly averages of $AQI_{PM2.5}$ and AQI_{NO2} for cities of India

Table 2 Chennai weekly AQI_{NO2} averages (CPCB-CCR 2020; Mintz 2012)

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13
AQI_{NO2}	4	5	4	4	4	3	3	2	3	4	3	3	2

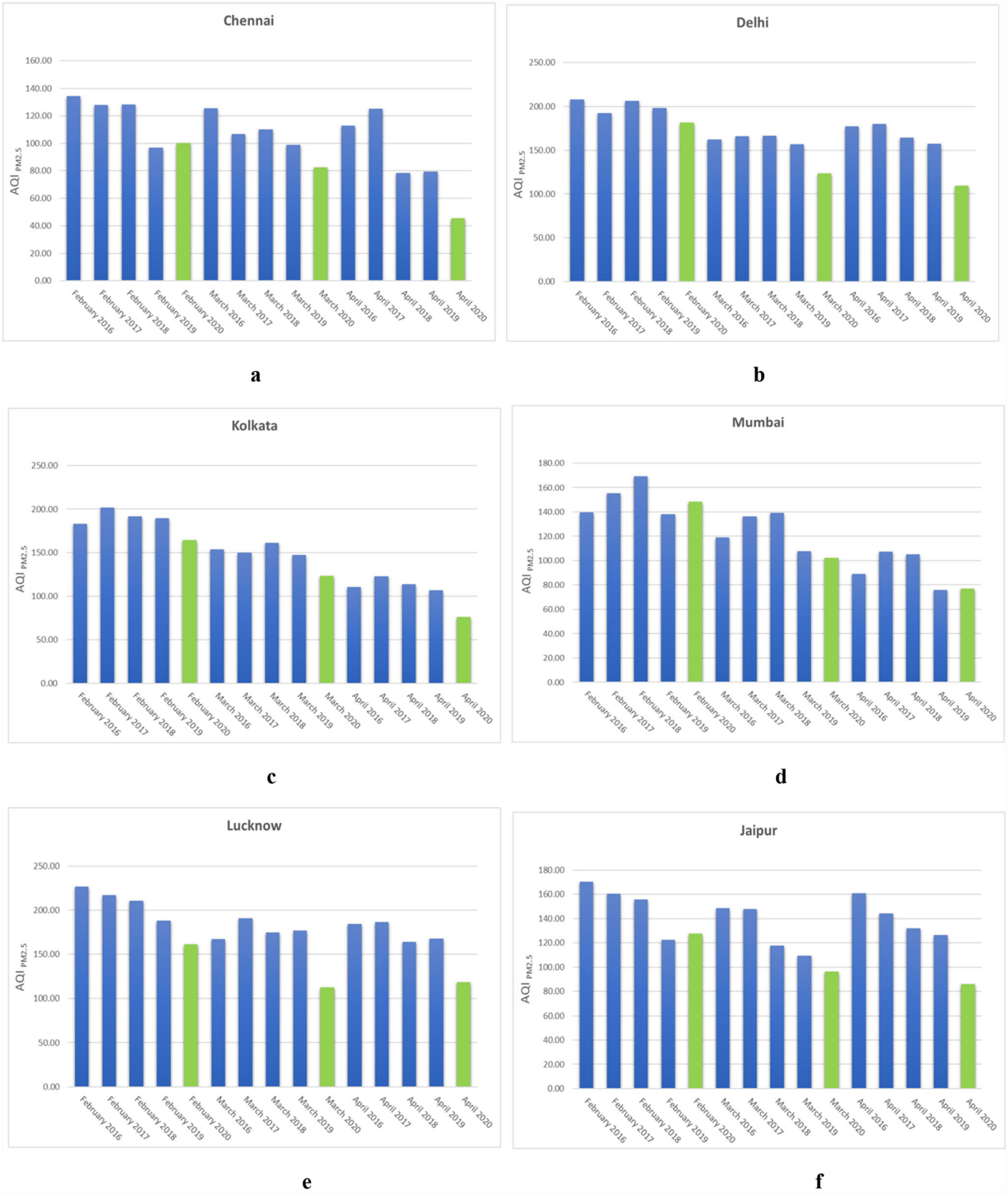


Fig. 4 Past 5 years' averages of $AQI_{PM2.5}$ for cities of India

value. Several academic studies have been conducted to study the relationship between the local meteorological factors and concentration of various pollutants (viz. $PM_{2.5}$ and NO_2) (Guo et al. 2017). The prime factors which influence the concentration of $PM_{2.5}$ and NO_2 have been identified as precipitation, ambient temperature, wind speed, and relative humidity.

Table 3 Delhi daily AQI_{NO2} and $AQI_{PM2.5}$ averages (CPCB-CCR 2020; Mintz 2012)

	April 1, 2020	April 2, 2020	April 3, 2020	April 4, 2020	April 5, 2020	April 6, 2020	April 7, 2020	April 8, 2020	April 9, 2020	April 10, 2020	April 11, 2020	April 12, 2020	April 13, 2020
AQI_{NO2}	7	9	10.75	12.33	11	8	9.5	11	13	11.67	8.33	12	12.33
$AQI_{PM2.5}$	92.75	77.2	88.2	97.75	116	137	94.67	90.67	93.5	124.25	122.25	113	128.75

However, the 12 selected cities tend to show a tremendous reduction in the concentration of $PM_{2.5}$ and NO_2 due to lockdown enforcement. It reflects that cutting down on anthropogenic sources of various pollutants can be useful in reducing the AQI.

Huanggang, China, has a population of 6.62 million, and the $AQI_{PM2.5}$ and AQI_{NO2} of Huanggang before the lockdown were 147.1 and 12, respectively. After the first week of implementation of lockdown, there was found a decrease of 13.40% and 58.54% in $AQI_{PM2.5}$ and AQI_{NO2}. Likewise, Jaipur, India, has a population of 6.42 million and the $AQI_{PM2.5}$ and AQI_{NO2} of Jaipur before the lockdown was implemented were 115.7 and 15, respectively. After the first week of the implementation of lockdown, a decrease of 27.50% and 60.37% in $AQI_{PM2.5}$ and AQI_{NO2} was recorded. Both cities have a comparable population and immediate AQI_{NO2} drop, but the immediate $AQI_{PM2.5}$ drop differs by 14.1%. The lowest week average $AQI_{PM2.5}$ recorded in Huanggang and Jaipur was recorded 68.7 and 73.4, respectively. Both the cities' $AQI_{PM2.5}$ has dropped down from unhealthy for sensitive groups to moderate air quality. Within 4 weeks of implementation of lockdown, Huanggang experienced a drop of 53.30% and 65.85% in week average $AQI_{PM2.5}$ and AQI_{NO2}. Furthermore, Jaipur saw a drop of 36.56% and 67.49% in week average $AQI_{PM2.5}$ and AQI_{NO2} in just 2 weeks.

For a holistic view, the cities selected in India recorded an average immediate $AQI_{PM2.5}$ and AQI_{NO2} drop of 20.21% and 59.26%, respectively. In contrast, on the other hand, the cities in China recorded an average drop of 11.32% and 48.61%, respectively. After 6 weeks of implementing the lockdown, cities in India recorded an average drop in $AQI_{PM2.5}$ and AQI_{NO2} of 37.42% and 65.80%, respectively, while cities in China recorded a drop of 42.54% and 56.67% respectively. From these results, it can be inferred that the drop in $PM_{2.5}$ is rather gradual as compared with the sudden drop in NO_2 concentrations throughout the cities.

As shown in Fig. 5, the drop in $AQI_{PM2.5}$ of coastal cities (viz. Chennai, Mumbai, and Kolkata) is relatively more significant than inland cities. The exceptional drop in $AQI_{PM2.5}$ in the coastal cities is vastly due to the coastal winds which are very prominent in these cities. Previous studies conducted in these coastal cities show that coastal regions show a significant drop in $PM_{2.5}$ in the morning as compared with inland regions (Chen et al. 2020b; Gupta et al. 2004). The three coastal cities incorporated in the present study record an average immediate $AQI_{PM2.5}$ drop of 24.96%, and the final average $AQI_{PM2.5}$ drop was found as 54.90%. On the other hand,

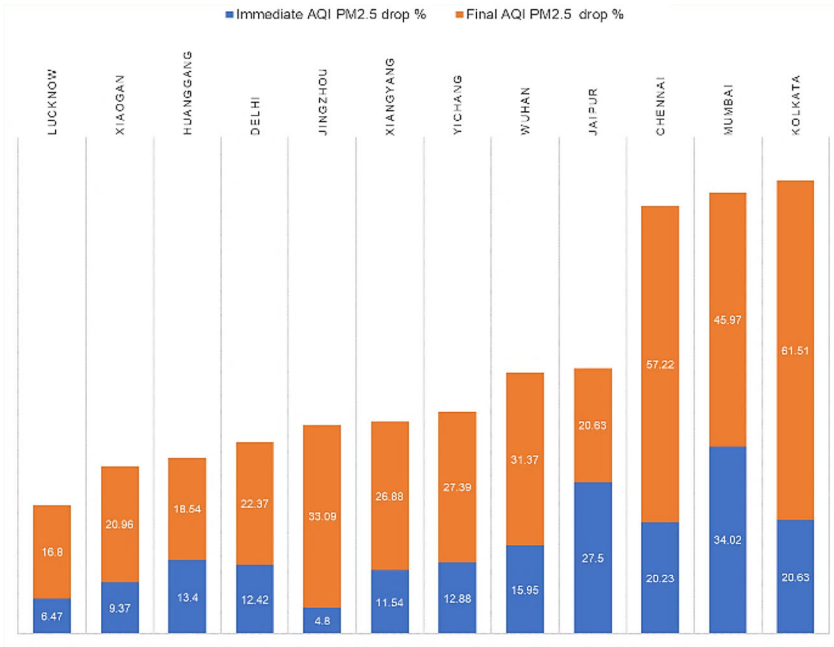


Fig. 5 Immediate and final Drop percentages of $AQI_{PM2.5}$ for all cities

the inland region cities recorded an average immediate $AQI_{PM2.5}$ drop of 12.70% and the final average $AQI_{PM2.5}$ drop of 24.23%. The drop in average immediate and final $AQI_{PM2.5}$ of cities in inland regions is moderate as compared with the drop in coastal regions.

Drop-in AQI_{NO2} also shows similar trends as $AQI_{PM2.5}$ (Fig. 6). The coastal cities recorded a much higher percentage of drop in AQI_{NO2} as compared with cities in inland regions. The coastal cities, except for Chennai, show an exceptionally high drop in AQI_{NO2}. Mumbai and Kolkata recorded an immediate AQI_{NO2} drop of 76.28% and 55.70%, respectively, whereas the final AQI_{NO2} drop is 92.58% and 76.67%, respectively. However, Chennai is a coastal city that recorded a much lower percentage drop in AQI_{NO2} as compared with the other two coastal cities, AQI_{NO2} of Chennai was already at a record low values between 3 to 5 before even lockdown was implemented. Xiangyang recorded an increase in AQI_{NO2} levels after the implementation of lockdown. Xiangyang, being a heavily industrialised city, had industries that were operating during the lockdown to produce essential medical equipment.

Conclusion

Both the nations followed different protocols for implementing lockdown in each country, although the lockdown in both of the nations was found effective in declining the rate of spread of COVID-19 cases (Wang et al. 2020b; Barkur et al. 2020), and it played a significant role in reducing the air pollution to record low values. The significant findings of the study are as follows:

- In China, the week before the lockdown was enforced, 4 out of 6 cities had an $AQI_{PM2.5}$ in the unhealthy category. Wuhan and Huanggang were found in unhealthy for sensitive group category. In the last week of lockdown, 5 out of 6 cities were found to be in unhealthy for sensitive group category except for Wuhan, which was found in the moderate category.
- In India, the week before the lockdown was enforced, 5 out of 6 cities had an $AQI_{PM2.5}$ that is unhealthy for sensitive group category except Chennai that was found under the moderate category. In the sixth week of lockdown, all cities were found in the good and moderate category except Delhi and Lucknow, which were found to be in unhealthy for sensitive group category.
- Meteorological factors are an essential factor in order to address pollutant concentration in ambient environment. Henceforth, meteorological should be taken into account before the execution of a response plan to mitigate pollution in urban cities around the world.
- For all 12 cities, a gradual decline has been recorded in $AQI_{PM2.5}$ levels in subsequent lockdown weeks. The mean immediate and final $AQI_{PM2.5}$ drops are 15.76% and 31.89%, respectively. However, in the case, AQI_{NO2} levels, a sharp decline has been recorded in the first week of lockdown. The mean immediate and final AQI_{NO2} drops are 53.93% and 46.22%, respectively.
- The coastal cities (viz. Chennai, Mumbai, and Kolkata) recorded a more significant decline in $AQI_{PM2.5}$ and AQI_{NO2} as compared with the other inland region cities.

The lockdown implemented in various regions around the world provided us with a unique opportunity to identify the

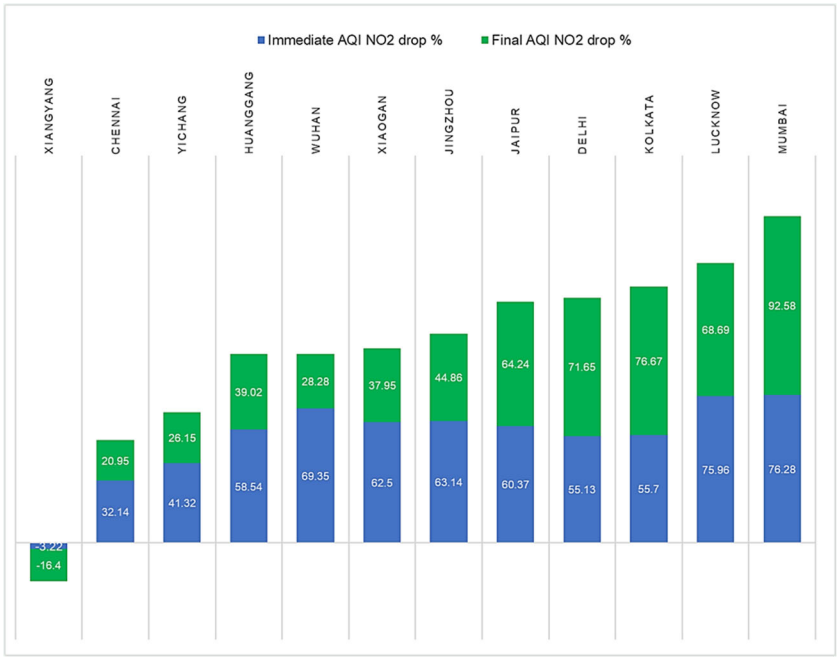


Fig. 6 Immediate and final Drop percentages of AQI_{NO2} for all cities

benchmark levels of pollutants in various urban cities around the world. The findings of the study will assist the governing authorities and policymakers to calibrate a proper response plan to bring down the ever-increasing pollution levels in various developing urban regions across the globe, especially China and India.

Acknowledgements Authors are incredibly grateful to Advance Air & Acoustics Research Laboratory, Delhi Technological University, Delhi (India), for encouraging to conduct this research and providing all the facility to compile this work.

Funding information This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of data and material The data used in the current study have been taken from respective EPA and is available in public domain.

Compliance with ethical standards

Competing interests The authors declare that they have no known competing financial interest or personal relationships that would have appeared to influence the work reported in this paper.

References

Agarwal R, Jayaraman G, Anand S, Marimuthu P (2006) Assessing respiratory morbidity through pollution status and meteorological conditions for Delhi. Environ Monit Assess 114:489–504. https://doi.org/10.1007/s10661-006-4935-3

ANI News (2020) Delhi's air quality may deteriorate due to western dust storms: SAFAR. https://www.aninews.in/news/environment/delhis-air-quality-may-deteriorate-due-to-western-dust-storms-safar20200415085141/. Accessed 1 Apr 2020

Barkur G, Vibha, Kamath GB (2020) Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: evidence from India. Asian J Psychiatr 51:102089. https://doi.org/10.1016/j.ajp.2020.102089

Błaszczyk E, Rogula-Kozłowska W, Klejnowski K, Kubiesa P, Fulara I, Mielżyńska-Švach D (2017) Indoor air quality in urban and rural kindergartens: short-term studies in Silesia, Poland. Air Qual Atmos Health 10:1207–1220. https://doi.org/10.1007/s11869-017-0505-9

Central Intelligence Agency (2018) The World Factbook. https://www.cia.gov/library/publications/the-world-factbook/fields/344rank.html. Accessed 18 Apr 2020

Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y, Xia J, Yu T, Zhang X, Zhang L (2020a) Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 395:507–513. https://doi.org/10.1016/s0140-6736(20)30211-7

Chen Y, Wild O, Conibear L, Ran L, He J, Wang L, Wang Y (2020b) Local characteristics of and exposure to fine particulate matter (PM2.5) in four Indian megacities. Atmos Environ X 5:100052. https://doi.org/10.1016/j.aeaoa.2019.100052

Conticini E, Frediani B, Caro D (2020) Can atmospheric pollution be considered a co-factor in extremely high level of SARS-CoV-2 lethality in northern Italy? Environ Pollut 261:114465. https://doi.org/10.1016/j.envpol.2020.114465

CPCB-CCR (2020) Central control room for air quality management – all India. https://app.cpcbccr.com/ccr/#/caaqm-dashboard-all/caaqm-landing/data. Accessed 16 Apr 2020

Fan H, Zhao C, Yang Y (2020) A comprehensive analysis of the spatio-temporal variation of urban air pollution in China during 2014–2018. Atmos Environ 220:117066. https://doi.org/10.1016/j.atmosenv.2019.117066

Guo J, Xia F, Zhang Y, Liu H, Li J, Lou M, He J, Yan Y, Wang F, Min M, Zhai P (2017) Impact of diurnal variability and meteorological factors on the PM2.5 - AOD relationship: implications for PM2.5 remote sensing. Environ Pollut 221:94–104. https://doi.org/10.1016/j.envpol.2016.11.043

Gupta A, Patil RS, Gupta S (2004) Influence of meteorological factors on air pollution concentration for a coastal region in India. Int J Environ Pollut 21:253. https://doi.org/10.1504/ijep.2004.004193

Health Effects Institute (2019) State of Global Air 2019. Special Report. Health Effects Institute, Boston

Hubei Provincial People's Government (2020) Hubei Enterprises. http://en.hubei.gov.cn/business/enterprises/. Accessed 3 Apr 2020

IEA OECD (2016) Energy and Air Pollution: World Energy Outlook Special Report 2016

Jain S, Sharma T (2020) Social and travel lockdown impact considering coronavirus disease (COVID-19) on air quality in megacities of India: present benefits, future challenges and way forward. Aerosol Air Qual Res 20:1222–1236. https://doi.org/10.4209/aaqr.2020.04.0171

Kumar R, Joseph AE (2006) Air pollution concentrations of PM2.5, PM10 and NO2 at ambient and Kerbsite and their correlation in Metro City – Mumbai. Environ Monit Assess 119:191–199. https://doi.org/10.1007/s10661-005-9022-7

Kumar A, Mishra RK (2018) Human health risk assessment of major air pollutants at transport corridors of Delhi, India. J Transp Health 10:132–143. https://doi.org/10.1016/j.jth.2018.05.013

Li S, Feng K, Li M (2017) Identifying the main contributors of air pollution in Beijing. J Clean Prod 163:S359–S365. https://doi.org/10.1016/j.jclepro.2015.10.127

Liu H, Tian Y, Song J, Cao Y, Xiang X, Huang C, Li M, Hu Y (2018) Effect of ambient air pollution on hospitalization for heart failure in 26 of Chinas largest cities. Am J Cardiol 121:628–633. https://doi.org/10.1016/j.amjcard.2017.11.039

Lu D (2020) Inside Wuhan's lockdown. New Sci 245:7. https://doi.org/10.1016/s0262-4079(20)30234-7

Ministry for the Environment New Zealand (2020) Nitrogen dioxide. https://www.mfe.govt.nz/air/specific-air-pollutants/nitrogen-dioxide. Accessed 5 Apr 2020

Mintz D (2012) Technical assistance document for the reporting of daily air quality – the air quality index. EPA-454/B-12-001

MoHFW, Ministry of Health and Family Welfare (2020) COVID-19 INDIA. https://www.mohfw.gov.in/. Accessed 2 Apr 2020

National Bureau of Statistics of China (2010) Census Data http://www.stats.gov.cn/english/Statisticaldata/CensusData/. Accessed 5 May 2020

National Health Commission of the People's Republic of China (2020) April 30: Daily briefing on novel coronavirus cases in China. http://en.nhc.gov.cn/2020-05/16/c_80207.htm. Accessed 30 Apr 2020

Office of the Registrar General & Census Commissioner, India (2011) 2011 Census Data. http://censusindia.gov.in/2011-Common/CensusData2011.html. Accessed 22 Apr 2020

Pant P, Harrison RM (2012) Critical review of receptor modelling for particulate matter: a case study of India. Atmos Environ 49:1–12. https://doi.org/10.1016/j.atmosenv.2011.11.060

Park S, Allen RJ, Lim CH (2020) A likely increase in fine particulate matter and premature mortality under future climate change. Air Qual Atmos Health 13:143–151. https://doi.org/10.1007/s11869-019-00785-7

Rodríguez MC, Dupont-Courtade L, Oueslati W (2016) Air pollution and urban structure linkages: evidence from European cities. Renew Sust Energ Rev 53:1–9. https://doi.org/10.1016/j.rser.2015.07.190

Siddique S, Ray MR, Lahiri T (2010) Effects of air pollution on the respiratory health of children: a study in the capital city of India. Air Qual Atmos Health 4:95–102. https://doi.org/10.1007/s11869-010-0079-2

The Indian Express (2020) Indian outrage on social media after people light fireworks, lead processions after lights off. https://indianexpress.com/article/trending/trending-in-india/primisniter-narendra-modi-9pm9minutes-coronavirus-6349647/. Accessed 5 May 2020

UN Environment Programme (2018) Air pollution: know your enemy. https://www.unenvironment.org/news-and-stories/story/air-pollution-know-your-enemy. Accessed 2 Apr 2020

US EPA (2018) Particulate Matter (PM) Basics. https://www.epa.gov/pm-pollution/particulate-matter-pm-basics. Accessed 12 Mar 2020

Venkatram A, Schulte N (2018) Urban transportation and air pollution. Elsevier Science Publishing Co Inc, s.l.

Wang P, Chen K, Zhu S, Wang P, Zhang H (2020a) Severe air pollution events not avoided by reduced anthropogenic activities during COVID-19 outbreak. Resour Conserv Recycl 158:104814. https://doi.org/10.1016/j.resconrec.2020.104814

Wang S-X, Wang Y, Lu Y-B, Li J-Y, Song Y-J, Nyamgerelt M, Wang X-X (2020b) Diagnosis and treatment of novel coronavirus pneumonia based on the theory of traditional Chinese medicine. Journal of Integrative Medicine 18:275–283. https://doi.org/10.1016/j.joim.2020.04.001

WHO (2020) Coronavirus Disease (COVID-19) Situation Reports. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports. Accessed 30 Apr 2020

Wilder-Smith A, Freedman DO (2020) Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak. J Travel Med 27. https://doi.org/10.1093/jtm/taaa020

World Health Organization (2005) WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulphur dioxide: global update 2005: summary of risk assessment (No. WHO/SDE/PHE/OEH/06.02)

Wu JT, Leung K, Leung GM (2020a) Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 395:689–697. https://doi.org/10.1016/s0140-6736(20)30260-9

Wu X, Nethery RC, Sabath BM, Braun D, Dominici F (2020b) Exposure to air pollution and COVID-19 mortality in the United States: a nationwide cross-sectional study. https://doi.org/10.1101/2020.04.05.20054502

Zhang L, Wang B, Zhou J, Kirkpatrick J, Xie M, Johri AM (2020) Bedside focused cardiac ultrasound in COVID-19 infection from the Wuhan epicenter: the role of cardiac point of care ultrasound (POCUS), Limited Transthoracic Echocardiography and Critical Care Echocardiography. J Am Soc Echocardiogr 33:676–682. https://doi.org/10.1016/j.echo.2020.04.004

Zhu Z (2005) Political economy of China and India: dealing with air pollution in the two booming economies. Historia Actual Online 7:123–132

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

J. Inst. Eng. India Ser. B
https://doi.org/10.1007/s40031-020-00472-7

ORIGINAL CONTRIBUTION

Control and Design of Seven-Level Cascaded Multilevel Converter for Transformerless Large-Scale Photovoltaic Integration

Jyoti Kulkarni[1] · **Narendra Kumar**[1] · **Bhim Singh**[2]

Received: 15 December 2019 / Accepted: 14 July 2020
© The Institution of Engineers (India) 2020

Abstract In this work, the design and control of seven-level H-bridge multilevel converter are carried out for the grid integration of large photovoltaic (PV) system without bulky line frequency transformer, as the multilevel converter provides a direct integration to the medium voltage grid. This system is a single-stage system to reduce cost and control complexity. Its control uses a decoupled current control with synchronous reference frame (SRF) and phase locked loop (PLL) with low switching frequency phase-shifted pulse width modulation (PWM) technique. It reduces the switching losses and hence improves the efficiency. The outer loop of control comprises of a separate maximum power point technique (MPPT) controller for each PV array to overcome the unpredictable irradiance level variations, ambient temperature, shading effect and other relevant factors. The improved perturb and observe (P&O) MPPT algorithm is used to track the maximum power from the PV array to give the accurate results. A significant reduction in total harmonic distortion (THD) in the grid current because of the increased number of steps in output AC voltage waveform of multilevel converter and consequently reduction in size of the filter component are achieved. Simulative experiments are carried out in MATLAB/Simulink, and the performance is validated in accordance with the IEEE-519 standard.

Keywords MPPT · PV generation · SRF-PLL · Phase-shifted PWM · VSC · THD

Introduction

The fuels used in conventional power plant are depleting and polluting in the nature. Hence, to get clean, ever available, reliable and environment-friendly electricity generation, the usage of renewable energy sources, like solar and wind, is increasing since last two decades. Lots of research are being done on power converters to interface these renewable sources to the grid, and multilevel converter has been perceived as gen-next technique to connect renewable energy sources with the grid. Multilevel converters are advantageous over the traditional two-level converter due to increased scalability, lower harmonics content, reduced filter components, low switching frequency, less voltage stress across switching device and an increased efficiency. Many multilevel topologies like diode clamped, flying capacitor and cascaded H-bridge are reported in the literature. Out of which, cascaded H-bridge (CHB) multilevel converter is the best suitable configuration for photovoltaic (PV) generation as multistring PV plant naturally provides the isolated DC source for each bridge of the CHB converter [1–5].

For grid interfacing of PV system, single-stage (requires only DC–AC converter) as well as double-stage PV systems (require both DC–DC converter and DC–AC converter) is reported in the literature [6–18]. This system is a single-stage H-bridge multilevel converter to eliminate DC–DC stage, for reducing the cost and also control complexity in the system.

As multilevel converter provides the direct connection to the medium voltage grid, hence eliminating the need

✉ Jyoti Kulkarni
jyotikulkarni0803@yahoo.com

[1] Department of Electrical Engineering, Delhi Technological University, Main Bawana Road, New Delhi, Delhi 110042, India

[2] Department of Electrical Engineering, Indian Institute of Technology, New Delhi, Delhi, India

Published online: 27 July 2020

of large size and heavyweight transformer for grid interfacing, only a low frequency transformer is sufficient for providing galvanic isolation to limit leakage current [13].

Extracting maximal power from the solar PV array is also a challenge. Researchers have tried to track the maximum power using various approaches known as "MPPT techniques." Different techniques for implementing the concept of MPPT have been developed and reported in the literature [19]. In this system, separate improved P&O algorithm is conceived due to good accuracy and deployed for tracking the maximum power from each PV panel [20, 21] and each controlled by separate controller to overcome the unpredictable irradiance level variations, ambient temperature, shading effect and other relevant factors [18].

Its control comprises the decoupled current controller with SRF-PLL and PWM control for the modulation. A number of multicarrier PWM techniques are discussed in [22]. The low switching frequency phase-shifted PWM technique is employed in this work, which reduces the switching losses and hence improves the efficiency. With an increase in the number of steps (corresponding to the number of levels achieved) in the AC voltage output waveform in the multilevel converter, the THD in output grid current waveform is reduced, which also contributes to reduced size filter component. Moreover, a low switching frequency used in given work produces low acoustic noise, which eliminates the need of high value of an inductive filter; hence, only a low-value inductive filter is sufficient for the filtrating. The reduction in size of filter further results in reduction in the cost of system [18, 22, 23].

Many methods for synchronized connection, i.e., "integration" of large solar PV plant to the grid, are reported in the literature [9–11]. In [11], multipulse converters are used in large solar PV grid-connected system to minimize harmonics in the voltage waveform resulting in good power quality, lower switching losses and good reliability in comparison with the two-level topologies, but the good designing of transformer and interfacing magnetic is required for proper operation of multipulse-based system. The authors in [10] have reported "a novel scheme suitable for high-voltage and large-capacity photovoltaic power stations."

Some topologies for grid integration of single-phase converter are discussed in [12–15]. Here, a 3-Φ, 1.48 MW, 3.3 kV and 50 Hz power–frequency grid-integrated solar photovoltaic plant interfaced through H-bridge cascaded seven-level multilevel converter with phase-shifted PWM operated at 500 Hz switching frequency is proposed. The performance of the system is analyzed on MATLAB/Simulink.

System Description

The system as shown in Fig. 1, is a 1.48 MW, 3.3 kV, 50 Hz, 3-Φ seven-level cascaded H-bridge (CHB) converter with low switching frequency of 500 Hz.

In cascaded multilevel converter, the number of levels is equal to (2 * number of H-bridge per phase + 1) [23], and hence, seven-level converter consists of nine H-bridges, three H-bridges in each phase and is fed by the separate PV array. A distinct MPPT algorithm is needed for each PV array.

Design of System

The system is designed for 1.48 MW power, 50 Hz frequency, 3.3 kV voltage, and it includes the design of PV array, the DC-link capacitor and DC-link voltage, other specifications of the system, which are given in Table 1.

Selection of V_{dc} Voltage

In case of L number of level converter, for each individual H-bridge connected to the PV array, the DC-link voltage (V_{dc}) is calculated as in [23],

$$V_{rms} = 0.612 * (L-1) * V_{dc} \qquad V_{dc} = 898.692\,\text{V}. \tag{1}$$

V_{dc} is taken as 905 V.

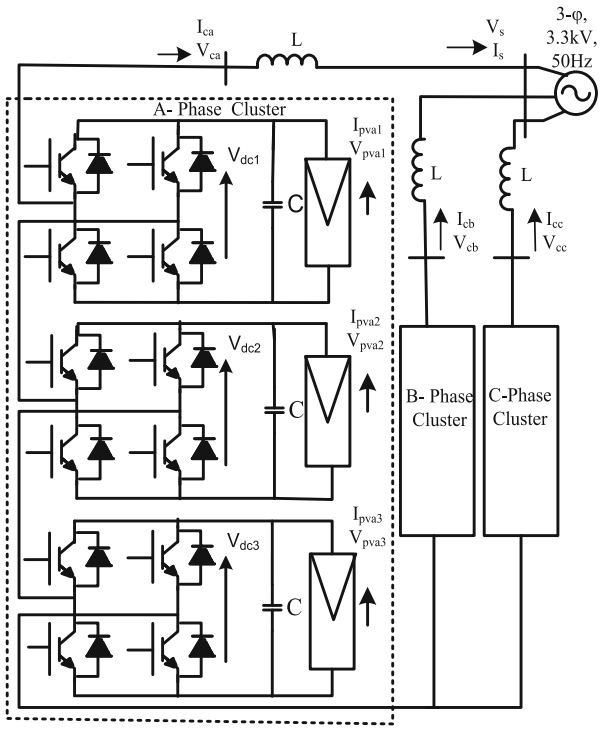


Fig. 1 Seven-level grid-interfaced converter PV system

Table 1 Component specifications

Component	Value
Power	1.48 MW
Voltage (V_{rms})	3.3 kV
Frequency	50 Hz
Switching frequency	500 Hz
Inductor (L)	5.4 mH
No of parallel paths in SPV array (N_p)	23
No of series paths in SPV array (N_s)	30
Voltage of PV array at maximum power (V_{mmp})	30.2 V
Current of PV array at maximum power (I_{mmp})	7.95 A
Open-circuit voltage of PV array (V_{oc})	37.5 V
Short-circuit current of PV array (I_{sc})	8.61 A
DC-link voltage (V_{dc})	905 V
DC-link capacitor (C_{dc})	14,500 μF

Design Calculations for Selection of PV Array

The required power rating of each PV array for a seven-level CHB converter consisting of nine PV array for 1.48 MW power grid is calculated as,

$$P_{mmp} = \frac{(\text{power rating of system})}{(\text{number of total PV array})} = 1.48\,\text{MW}/9 = 164.44\ \text{kW} \quad (2)$$

To obtain the PV array of 164.44 kW power (P_{mmp}) and V_{dc} of 905 V, a Sharp ND H230Q2 model is selected from PV array library, and numbers of PV modules in parallel and series are selected as,

$$N_s = V_{dc}/V_{mmp} = 30 \quad (3)$$

$$N_p = P_{mmp}/\left(N_s * I_{mmp} * V_{mmp}\right) = 23 \quad (4)$$

Selection of Capacitor for DC-Link

The DC-link voltage (V_{dc}) for each H-bridge connected to the PV array is obtained in accordance with the principle of conservation of energy for V_{dc} recovery in 5 ms and 1.2 overloading factor a, as [24],

$$\begin{aligned} {}^1\!/_2 * C_{dc}\left(V_{dc}^2 - V_{dc1}^2\right) &= K_1 * 3 * V * a * I * t \\ C_{dc} &= (0.1 * 164444 * 0.005)/ \\ &\quad (0.5 * (905 * 905 - 898.7 * 898.7)) \\ C_{dc} &= 14471.5\,\mu\text{F} \end{aligned} \quad (5)$$

It is selected as, $C_{dc}(C)$ = 14,500 μF.

Control Algorithm

The used algorithm for the control of the PV system connected to the three-phase AC grid is shown in Fig. 2. The maximum power developed by the PV array is tracked using an improved P&O algorithm. The decoupled current controller having a feedforward term, separate controller for each DC-link capacitor, SRF and PLL is used for controlling the voltage source converter (VSC) [25].

MPPT Algorithm

As shown in Fig. 3, by using an improved P&O MPPT algorithm, the maximum power is extracted from PV array and a suitable V_{dc_ref} is generated in this work. In this improved P&O MPPT algorithm, the detail of variation in current (ΔI_{pv}) is also considered along with the variation in voltage (ΔV_{pv}) and power (ΔP_{pv}) as in the P&O algorithm. In case of insolation change, normal P&O technique fails to differentiate between variation in the power on account of insolation variation and due to the perturbation, which is taken care of by improved P&O algorithm [20, 21].

Control of VSC

The voltage measured from each DC link is compared with the individual V_{dc_ref} given by each MPPT algorithm, as shown in Fig. 4. The error voltage (difference between the two) is processed by a proportional–integral (PI) controller to provide the active current loss component I^*_d, through the DC link, which is given as,

$$I_d^*(k) = I_d^*(k-1) + K_p\{V_e(k) - V_e(k-1)\} + K_i * V_e(k) \quad (6)$$

where V_e(k) and $V_e(k-1)$ are the error voltages at given sampling instants k & $(k-1)$ and K_p & K_i are proportional and integral gain constants. The sum of all reference DC-link currents along with the feedforward term is compared with the direct current component obtained by Park transform from the grid current as,

$$\begin{bmatrix} I_d \\ I_q \\ I_0 \end{bmatrix} = \frac{2}{3}\begin{bmatrix} \sin(\theta) & \sin\left(\theta - \frac{2\pi}{3}\right) & \sin\left(\theta + \frac{2\pi}{3}\right) \\ \cos & \cos\left(\theta - \frac{2\pi}{3}\right) & \cos\left(\theta + \frac{2\pi}{3}\right) \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}\begin{bmatrix} I_a \\ I_b \\ I_c \end{bmatrix} \quad (7)$$

The grid detection angle (θ) for Park transformation is obtained from three-phase PLL. I^*_q, the reference quadrature current (set to zero for feeding grid current at unity power factor) is compared with the quadrature current component obtained from Park transformation.

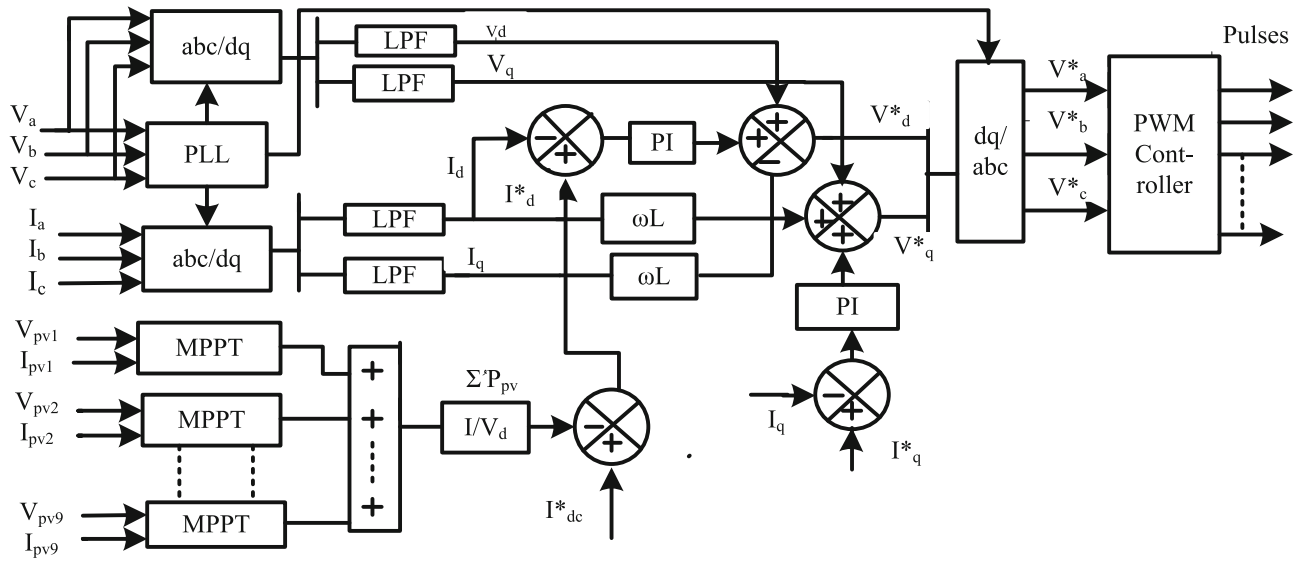


Fig. 2 Control algorithm

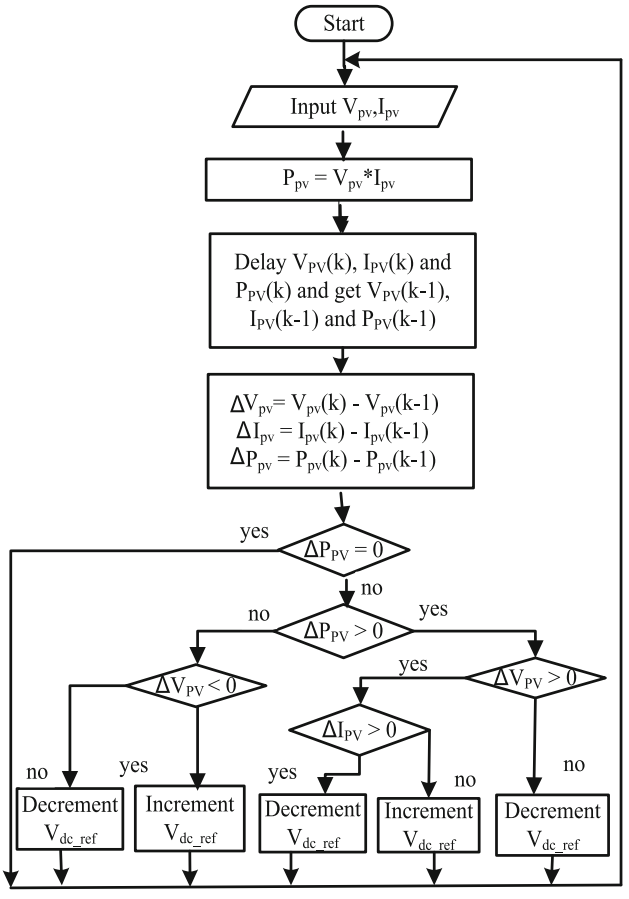


Fig. 3 Flowchart for improved P&O MPPT algorithm

Similarly, using Park transformation, direct and quadrature components of grid voltages (V_d & V_q) are calculated as follows.

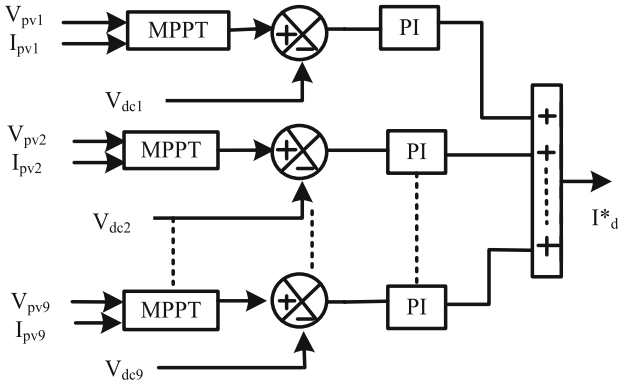


Fig. 4 Control algorithm for generating I^*_{dc}

$$\begin{bmatrix} V_d \\ V_q \\ V_0 \end{bmatrix} = \frac{2}{3} \begin{bmatrix} \sin(\theta) & \sin\left(\theta - \frac{2\pi}{3}\right) & \sin\left(\theta + \frac{2\pi}{3}\right) \\ \cos(\theta) & \cos\left(\theta - \frac{2\pi}{3}\right) & \cos\left(\theta + \frac{2\pi}{3}\right) \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} V_a \\ V_b \\ V_c \end{bmatrix} \tag{8}$$

The reference direct-axis voltage and quadrature-axis voltages are calculated as follows,

$$V_d^*(k) = V_d^*(k-1) + K_{p2}\{I_{de}(k) - I_{de}(k-1)\} + K_{i2} * I_{de}(k) + V_d - I_q^* \omega L \tag{9}$$

$$V_q^*(k) = V_q^*(k-1) + K_{p3}\{I_{qe}(k) - I_{qe}(k-1)\} + K_{i3} * I_{qe}(k) + V_q + I_d^* \omega L \tag{10}$$

where $I_{de} = (I^*_d - I_d)$, $I_{qe} = (I^*_q - I_q)$ and ωL is the coupling reactance. K_{p2}, K_{i2}, denote the proportional and

integral gains of direct-axis current controller and K_{p3} and K_{i3} are the proportional and integral gains of quadrature-axis current controller. The reference voltages (modulating signal) for the PWM controller are obtained using inverse Park transform.

In PWM controller, phase-shifted modulation technique is used, in which, for seven-level (L) converter, six (i.e., $(L-1)$), triangular carrier signals are generated [23]. The phase shift (φ_{cr}) between two adjacent carrier waves is given by

$$\begin{aligned}\varphi_{cr} &= 360^\circ/(L-1)\\ &= 360^\circ/6 = 60^\circ\end{aligned} \tag{11}$$

Each carrier (V_{cr1}–V_{cr6}) is compared with V_{ref} signal for generating pulses as shown in Fig. 5.

Results and Discussion

The system, shown in Fig. 1, is designed and simulated in MATLAB/Simulink, and the results are discussed as follows.

Response of System

The solar power incident on the PV array set equal to 1000 W/m^2 and temperature 25 °C for steady-state performance of the system and dynamic performance of the system is studied when insolation level is changed from 1000 to 400 W/m^2 at 0.6 s. The system performance is validated by the simulative results consisting of the grid voltages (V_{sabc}), grid current (I_{sabc}), AC phase voltage of VSC (V_{cbc}), PV voltage (V_{pv}), SPV current (I_{pv}), the power (P), PV voltage & current for array 1 of A-phase (V_{a1pv} & I_{a1pv}), PV voltage & current for array 1 of B-phase (V_{b1pv} & I_{b1pv}), PV voltage & current for array 1 of C-phase (V_{c1pv} & I_{c1pv}) as depicted in Figs. 6a and b and 7a–c. The maximum active power transfer is evident at a power factor of unity.

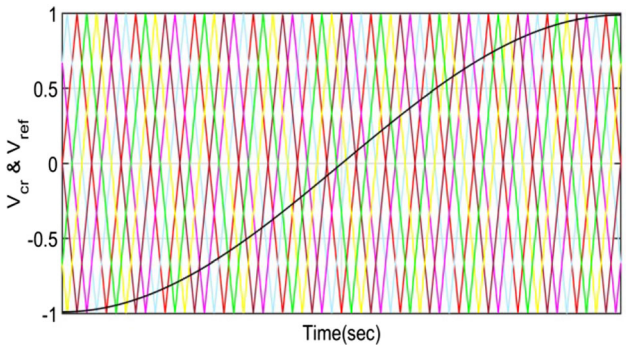


Fig. 5 Phase-shifted PWM scheme

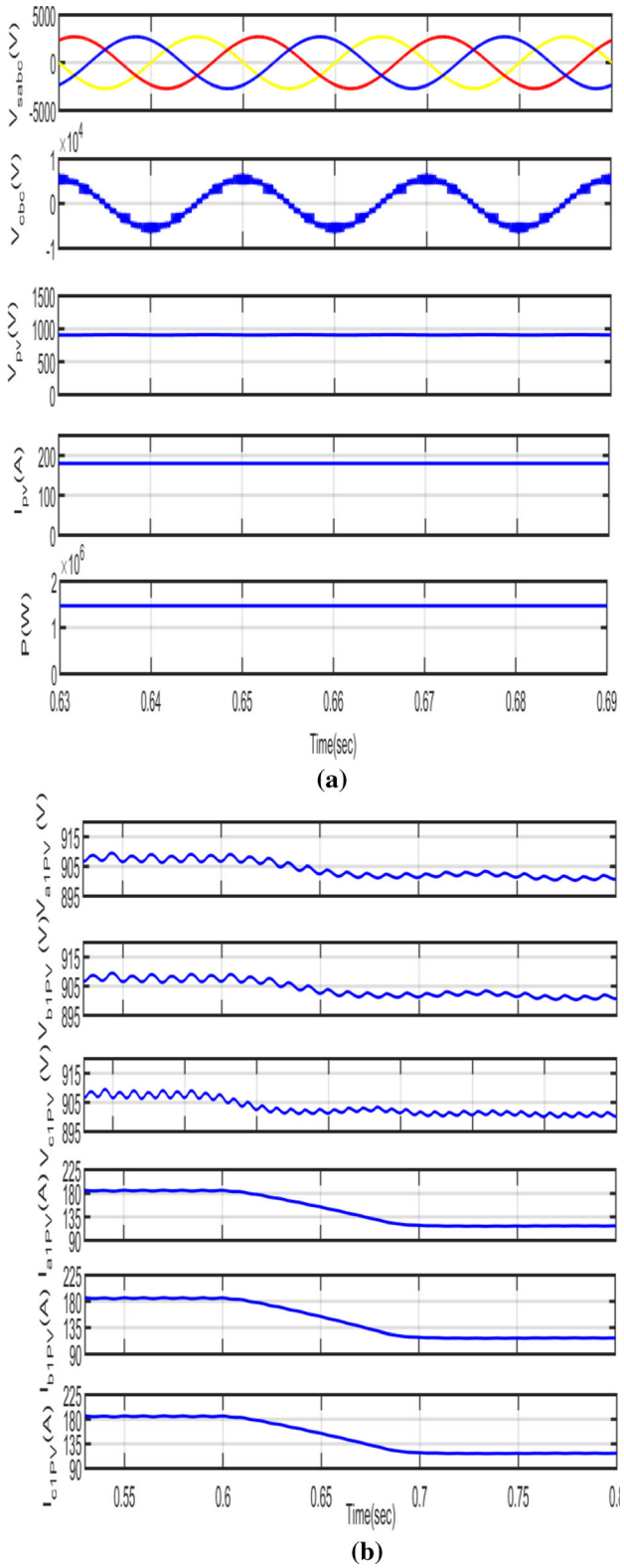
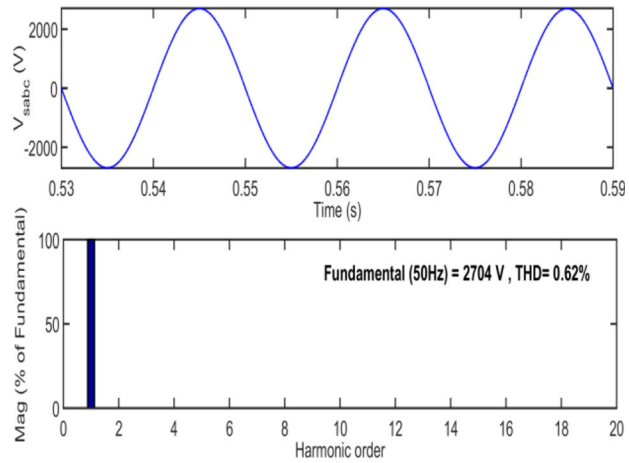
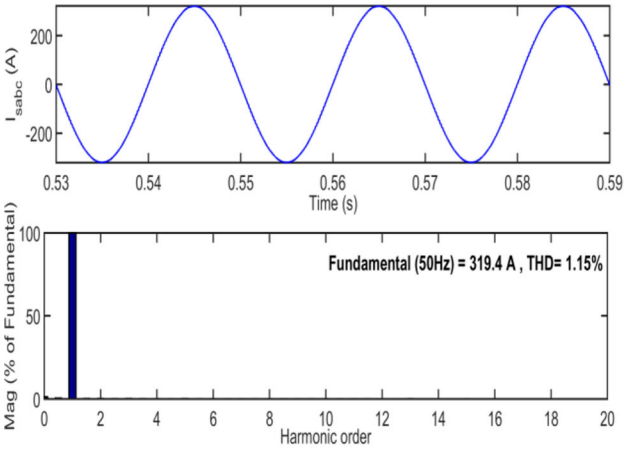


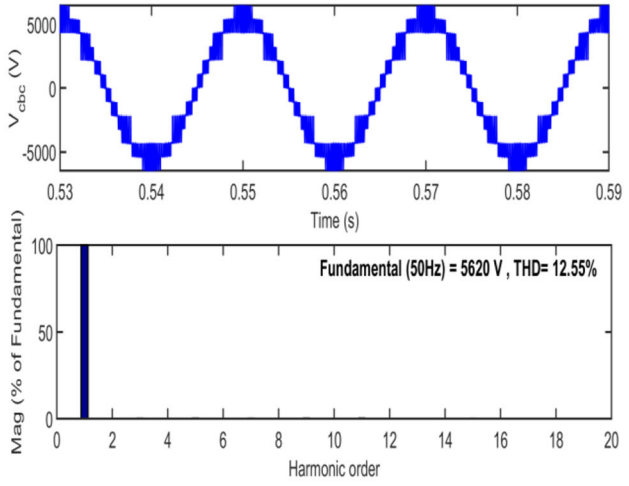
Fig. 6 **a** Steady-state performance at insolation level 1000 W/m^2, **b** dynamic performance when insolation level is changed from 1000 to 400 W/m^2 at 0.6 s for seven-level CHB grid-interfaced system



(a)



(b)



(c)

Fig. 7 Waveform and harmonic spectra of **a** line voltage, **b** line current and **c** converter line voltage at 1000 W/m^2

Improvement in Power Quality

Figure 7 shows the waveforms for the line voltage, line current (I_a) and converter line voltage (V_{cbc}) and harmonic spectra along with THD levels at 1000 W/m^2 levels. The THD values of 0.62%, 1.15% and 12.55% at insolation of 1000 W/m^2 and temperature of 25°, respectively, are observed for line voltage, line current and converter line voltage. The THD of the supply current is observed to be within the limits set by the IEEE-519 standard [26]. A minor variance is seen, though the power factor is also close to unity.

Conclusion

The design and control of single-stage transformerless seven-level H-bridge cascaded multilevel converters have been carried out with the decoupled current controller based on SRF-PLL and phase-shifted PWM at lower (500 Hz) switching frequency for large-scale photovoltaic grid-integrated plant. The low switching frequency reduces the switching losses, and hence, an improvement in the efficiency is achieved. An enhanced power quality is observed because of increased steps in output multilevel converter voltage waveform as compared to a two-level converter. It is attributed to the reduction in the supply current THD and reduction in size of filter components. Only an inductive filter is used for filtrating, which further results in reduction in the cost. The performance of the system is observed satisfactory at 1000 W/m^2 insolation level and 25° temperature and has been validated in accordance with the IEEE-519 standard [26].

References

1. F. Blaabjerg, R. Teodorescu, M. Liserre, A.V. Timbus, Overview of control and grid synchronization for distributed power generation systems. IEEE Trans. Ind. Electron. **53**(5), 1398–1409 (2006)
2. S. Kouro, M. Malinowski, K. Gopakumar, J. Pou, L. Franquelo, B. Wu, J. Rodriguez, M. Pérez, J. Leon, Recent advances and industrial applications of multilevel converters. IEEE Trans. Ind. Electron. **57**(8), 2553–2580 (2010)
3. J. Rodriguez, J.S. Lai, F.Z. Peng, Multilevel inverters: a survey of topologies, controls, and applications. IEEE Trans. Ind. Electron. **49**(4), 724–738 (2002)
4. R. Doga, M. Pathak, Recent trends in solar PV inverter topologies. Sol. Energy **183**, 57–73 (2019)
5. B. Xiao, L. Hang, J. Mei, C. Riley, L.M. Tolbert, B. Ozpineci, Modular cascaded H-bridge multilevel PV inverter with distributed MPPT for grid-connected applications. IEEE Trans. Ind. Appl. **51**(2), 1722–1731 (2015)
6. T. Kerekes, R. Teodorescu, M. Liserre, C. Klumpner, M. Sumner, Evaluation of three-phase transformerless photovoltaic inverter

topologies. IEEE Trans. Power Electron. **24**(9), 2202–2211 (2009)
7. M.R. Islam, Y. Guo, J. Zhu, A multilevel medium-voltage inverter for step-up-transformer-less grid connection of photovoltaic power plants. IEEE J. Photovolt. **4**(3), 881–889 (2014)
8. B. Singh, I. Hussain, Grid integration of single stage solar PV power generating system using 12-pulse VSC, in *2014 IEEE 6th India International Conference on Power Electronics (IICPE)*, Kurukshetra (2014), pp. 1–6
9. I. Hussain, B. Singh, Investigations on solar PV grid interfaced power generating system using two-level twelve-pulse double bridge converter, in *2014 9th International Conference on Industrial and Information Systems (ICIIS)*, Gwalior (2014), pp. 1-6
10. C. Zhang, S. Du, Q. Chen, A novel scheme suitable for high-voltage and large-capacity photovoltaic power stations. IEEE Trans. Ind. Electron. **60**(9), 3775–3783 (2013)
11. I. Hussain, B. Singh, Grid integration of large capacity solar PV plant using multipulse VSC with robust PLL based control, in *6th IEEE Power India International Conference (PIICON)*, Delhi (2014), pp. 1–6
12. N. Mishra, P. Kant, B. Singh, An asymmetric seven levelmultilevel converter for grid Integrated systems, in *2018 8th IEEE India International Conference on Power Electronics (IICPE)*, Jaipur, India (2018), pp. 1–6
13. S. Kuncham, K. Annamalai, S. Nallamothu, Single phase two stage seven-level power conditioner for photovoltaic power generation system. IEEE J. Emerg. Sel. Top. Power Electron. **8**(1), 794–804 (2020)
14. B. Singh, N. Mishra, P. Kant, Power quality improvement in single phase five level cascaded grid interfaced systems, in *2018 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES)*, Chennai, India (2018), pp. 1–6
15. R. Gonzalez, E. Gubia, J. Lopez, L. Marroyo, Transformerless single-phase multilevel-based photovoltaic inverter. IEEE Trans. Ind. Electron. **55**(7), 2694–2702 (2008)
16. B. Singh, M. Kandpal, I. Hussain, Grid integration of single stage SPV system using asymmetric cascaded 7-level VSC, in *39th National Systems Conference (NSC)*, Noida (2015), pp. 1–6
17. A. Acharya, M. Rico, D. Sera, R. Teoderscu, L. Norum, Performance analysis of medium-voltage Grid Integration of PV plant using modular multilevel converter. IEEE Trans. Energy Conv. **34**(4), 1731–1740 (2019)
18. A. Elmelegi, M. Aly, E. Ahmed, A. Alharbi, A simplified phase shift PWM-based feedforward distributed MPPT method for grid-connected cascaded PV inverters. Sol. Energy **187**, 1–12 (2019)
19. B. Subudhi, R. Pradhan, A Comparative Study on maximum power point tracking techniques for photovoltaic power systems. IEEE Trans. Sustain. Energy **4**(1), 89–98 (2013)
20. A. K. Singh, I. Hussain, B. Singh, An improved P&O MPPT algorithm for single stage three-phase grid integrated solar PV system, in *IEEE 7th Power India International Conference (PIICON)*, Bikaner (2016), pp. 1–6
21. M. Killi, S. Samanta, Modified Perturb and observe MPPT algorithm for drift avoidance in photovoltaic systems. IEEE Trans. Ind. Electron. **62**(9), 5549–5559 (2015)
22. B.P. McGrath, D.G. Holmes, Multicarrier PWM strategies for multilevel inverters. IEEE Trans. Ind. Electron. **49**(4), 858–867 (2002)
23. B. Wu, *High-Power Converters and AC Drives* (IEEE, 2006)
24. B. Singh, A. Chandra, K. Al-Haddad, *Power Quality: Problems and Mitigation Techniques* (Wiley, Hoboken, 2015)
25. R. Teodorescu, M. Liserre, P. Rodriguez, *Grid Converters for Photovoltaic and Wind Power Systems* (Wiley, Hoboken, 2011)
26. IEEE Recommended Practices and Requirements for Harmonic Control in Electrical Power Systems, IEEE Std 519-1992 (1993), pp. 1–112, 9

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/342523500

COVID-19 Global Pandemic: Impact on Management of Supply Chain

Article · April 2020

DOI: 10.46338/IJETAE0416

CITATIONS
0

READS
55

2 authors, including:



Rakesh Kumar
Delhi Technological University
7 PUBLICATIONS **0** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Performance Measurement of TQM in Indian FMCG Industry View project

Project Population Statistics View project

All content following this page was uploaded by Rakesh Kumar on 12 July 2020.

The user has requested enhancement of the downloaded file.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 10, Issue 04, April 2020)

COVID-19 Global Pandemic: Impact on Management of Supply Chain

Rakesh Kumar[1], R.S. Mishra[2]
[1,2]*Department of Mechanical Engineering, Delhi Technological University, Delhi, India-110042*

***Abstract*— To understand more about the global impact of (corona virus disease, detected in December, 2019, hereafter named) COVID-19 on supply chain management, a set of questionnaire is designed which addressed four pertinent themes focusing on product classification including export product, import product, elective product and emergency product. The questionnaire consisted of ten multiple choice questions including three open-ended questions (appendix A). The questionnaire was targeted at those organizations which were practicing supply chain management function as a core or a department of organization and was distributed online in interest groups using various social media platforms. The questionnaire was open from 26th March 2020 to 20th April 2020 and did not collect any identifiable data. Valid responses were represented as categorical data and presented in percentages.**

***Keywords*— COVID-19; supply chain management, product classification**

I. INTRODUCTION

The global downturn's speed and harshness due to COVID- 19 have significant implications for the supply chains of global manufacturers. That kind of volatility wreaks havoc on traditional supply chain planning: the process for determining production levels, raw-material purchases, transport capacity, and other vital factors, largely by examining historical patterns of demand. Indeed, the forecasting challenge is particularly acute because in many upstream industrial settings, as supply partners along the chain anticipate that demand will fall, the supply chain appears to be decoupling from downstream consumption—the focus of most forecasting models. In fact, risk managers and other executives fail to anticipate the effects, both negative and positive, of events that occurred. At first glance, for instance, a thunderstorm in a distant place wouldn't seem like cause for alarm. It was initially reported to the World Health Organization (WHO) on December 31, 2019 about Coronavirus disease 2019 (COVID-19) identified amid an outbreak of respiratory illness cases in Wuhan City, Hubei Province, China. On January 30, 2020, the WHO declared the COVID-19 outbreak a global health emergency. On March 11, 2020, the WHO declared COVID-19 a global pandemic.

The appeal of Janata Curfew on March 22,2020 by Hon'ble Prime Minister of India that the nation is set to observe an unprecedented shutdown and on March 24,2020 India imposed a 21-day lockdown, which was announced, comes after a call from the UN health agency, WHO, for the country to take "aggressive action".

What can companies do to prepare themselves for this sort of situation? True, there's no easy formula for anticipating the way risk cascades through a company or an economy. But we've found that executives who systematically examine the way risks propagate across the whole value chain—including competitors, suppliers, distribution channels, and customers—can foresee and prepare for second-order effects more successfully. Especially China, Korea, Japan, Germany and the US are also at of global value chains, so their woes will produce 'supply-chain contagion' in virtually all nations, Richard Baldwin and Beatrice Weder di Mauro apprehended this fact in their e-book on Economics in the Time of COVID-19. These nations are central to the global supply chains in many manufactured goods. The near term impacts—missed shipments, border-crossing delays, and communications problems—are easily understood. Less clear is the longer-term effect of increased uncertainty on lead times and demand and what that means for supply chain strategy. "Risk is something you can predict. If you flip a coin, over time you know that half will be heads and half will be tails. The nature of choice between heads and tails involves a risk because you can bet with these probabilities. So, in conditions of risk, you can distribute your investment into certain proportions against different outcomes that you may expect," Kim said (Minyoung Kim, associate professor of strategy and international business at the University of Kansas)."But uncertainty means there's nothing you can do about it beforehand. You know neither the outcomes nor the probability. I would call that the 'unknown unknown.' This could be a natural disaster or Brexit. If you knew, you might prepare against it. But you don't know that you don't know." The new uncertainty COVID-19 will require companies to rethink their supply chain strategy. For many, that effort begins with a reassessment of their demand-based strategy, which in recent years has been touted as the preferred approach to supply chain management.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 10, Issue 04, April 2020)

Observing the scenario of Indian market S. Srinivasamurthy, Research Director, Enterprise Solutions and ICT Practices, IDC India says, "While the actual impact of COVID-19 on India market will be evident by middle of 2020, we expect a slowdown in terms of discretionary IT spending, contract renewals and new deals getting signed as enterprises recalibrate by cost structure in coming months. Existing project executions have also taken a hit due to travel restrictions in place. Corporates across the country implementing alternative ways of working, i.e. work from home, it is generating a parallel corporate line that demands to be connected from where they want, when they want and to who they want. But the different sets of challenges that the IT teams within the organization are grappling with - how to secure data and assets from cyber threats. Companies can meet most challenges noted above with a sound supply chain strategy, with IT as the key enabler. Though the complexities of India's supply chain may appear overwhelming, understanding and mastering them is a critical success determinant for an organization attempting to serve customers in India. An efficient review of supply chain design will help better position companies in what is becoming an increasingly competitive marketplace.

Supply chain in Indian Economy

Supply chain costs in India represent as much as 13 percent of the Gross Domestic Product (GDP). This is almost twice the percentage in developed countries: in the U.S., supply chain costs amount to 8.5 percent of the GDP. ARC Advisory Group categorizes supply chain challenges in India into two groups: demand side challenges and supply side challenges. Demand side supply chain challenges in India relate to price and variety. With 28 States and 7 Union Territories, differing considerably in dress, food tastes, customs, traditions and purchasing habits, it's almost impossible for a single manufacturer to address the wide variety of consumer requirements. Thus, it is important for manufacturers to partner with local distributors and retailers who can help develop products that will appeal to local consumers, at appropriate price points. When it comes to COVID-19's immediate aggregate demand shock (i.e purchases drop) is observed, for which, two aspects are to distinguish: real-world and emotional. Real-world since some consumers are or will be prevented from getting to stores, so their demand disappears from the market. Similarly, some home delivery services are suspended, so goods and consumers are approaching together less frequently.

Emotional since – as happened in the wake of the Global Crisis – consumers and firms tend to embrace a 'wait-and-see' attitude when faced with massive Knightian uncertainty (the unknown-unknowns) of the type that COVID-19 is now presenting to the world. This, quite naturally, makes trade flows susceptible to demand shocks (purchases drop) and supply shocks (production drop). The supply shock aspects of COVID-19 – factory closures, travel bans, border closings and the like – will reduce exports across the border. Many vendors have been successful in selling "no-frills," standardized products to rural segments in India only. Big Bazaar, Hindustan Unilever, Maruti, Tata Motors and Hero have offerings relevant to this price-sensitive segment of the population. Other successful examples of localized low-price products include Siemens' pacemakers, Godrej's Chotukool refrigerator, and General Electric's ECG. Supply side supply chain challenges in India mainly relate to poor infrastructure, complex tax infrastructure, weak distribution system, fragmented market, and lack of technology adoption. Richard Baldwin and Beatrice Weder di Mauro as responsible editor mentioned in their recent e-book on Economics in the Time of COVID-19, the disruptions that companies, individuals and governments are experiencing imply that globalisation and integration may be at risk from such health shocks i.e. COVID 19. Firms will probably take into account the lesson they are learning that global supply chains can be abruptly broken by this health shock.

II. Literature Review

Businesses dependent on global sourcing are facing hard choices in crisis management amid the supply chain disruptions due to COVID-19. The impact of the coronavirus pandemic on global supply chains is a major disruption, along the lines of having an earthquake or a tsunami. This is an unprecedented type of disruption. The uncertainties ahead swing between extremes. The main target of managing the supply chain is to realize and neutralize the uncertainties in the supply chain. Uncertainties in the supply, process and demand are recognized to have a major impact on the manufacturing function. As this paper focuses on the impact of COVID-19 on the supply chain management, the literatures were reviewed nearby similar context of reputed journals related to the flexibility, agility and pro-activeness of the supply chain.

Angkiriwang, R. et al. (2014) in his study obtain insights into the typology of uncertainty and relevant strategies adopted by manufacturing companies to achieve better supply chain flexibility.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 10, Issue 04, April 2020)

Strategies are classified into reactive (buffering) and proactive (redesigning). A framework is developed that links supply chain uncertainty, as result the two types of strategies for achieving supply chain flexibility and the relevant objectives to be achieved. The research is endorsed by four case studies and are compared in terms of uncertainty typology and strategies being adopted to improve supply chain flexibility. supply chain flexibility (SCF) entails the implicit requirement of flexibility within and between all partners in the chain (Duclos *et al.*, 2003). Vickery, S.(1999) to establish the linkage, examined dimensions of supply chain flexibility and their relationships with environmental uncertainty, business performance, and functional interfaces. Prater et al. (2001) used five case studies to show the techniques of better trade-off between vulnerability (uncertainty) and supply chain agility by creating a link between the two called supply chain exposures. They considered and found two main problems of European industries, that is, in accurate forecasting and in developing an efficient and effective supply chain (SC) structure. The work described exposure as the degree to which an agile supply chain must be overextended (i.e. vulnerable) and, therefore, should be restructured or adjusted to address the issues in international SCs. To obtain high service levels at a low cost for emergency supply options in terms of lateral transshipments and direct deliveries, Alfredsson, P. et al. (1999) consider a two-echelon inventory system. An analytical model is used to calculate relevant performance measures. Comparison of the results of the model with the results of other models indicates that the combined use of lateral transshipments and direct deliveries can lead to significant cost savings. Agrawal, S. et al. (2019) emphasized about the segregation of products into different classified categories is essential for reverse supply chain also.

Like COVID-19, the earthquake and tsunami that struck Japan in March 2011 directly impacted over 27000 businesses whose production, warehousing, and retail facilities were destroyed or disabled by the natural disaster. Supply chain disruption is defined MacKenzie, C. A (2014) as a disruptive event that causes production difficulties for multiple suppliers and in which at least two of these suppliers deliver non-identical goods or services to at least two competing firms. The severity of a supply chain disruption can be measured by the number of entities that encounter difficulties in receiving or delivering materials or goods due to an unplanned event (Craighead et al., 2007).

A number of authors have tried to develop a framework for relationship between the typology of uncertainty and the type of strategies chosen for better supply chain. The study proposed in this article follows real time approach of the disruption COVID-19. The survey based work analyzes what supply chain firms should do during a severe supply chain disruption under uncertain conditions.

III. Motivation For The Research

The current COVID-19 global effect greatly motivated for this work. As India goes into a 21-day lockdown, there is some confusion arises about the e-commerce delivery of essential goods. According to lockdown rules, delivery of essential goods and groceries from e-commerce sites is allowed. Services like Grofers, BigBasket and Milkbasket allege their delivery people have been beaten up on the road. Facing difficulties, Flipkart has temporarily shut down service. Others too vary(www.indiatoday.in). The three East Asian manufacturing giants – China, South Korea, and Japan – account for over 25% of US imports, and over 50% of US imports of computer and electronics products. Apparel and footwear companies are particularly vulnerable to East Asian supply disruptions. The Economist magazine pointed to the vulnerability of the electronics industry with its practice of keeping very lean inventory levels and the lack of alternative sources for many electronic components. The optics sector is likewise highly exposed. The heart of the outbreak, Hubei province, is known as China's 'optics valley' since so many firms manufacturing fiber optic components are located there (these are essential inputs for telecoms networks). The automobile sector, especially in East Asia, is already significantly disrupted by ruptures in international supply chains. For example, a shortage of parts coming from China has forced Korean carmaker Hyundai to shut all its car plants in Korea. The Japanese firm Nissan closed a factory in Japan temporarily. The shock has even reached Europe. Fiat-Chrysler has recently warned that it could soon halt production at one of its European factories. Jaguar Land Rover, a UK-based auto company, announced it might to run out of parts from the end of February,2020. To stave this off, it had flown in emergency supplies from China in suitcases [7].

IV. Problem Description

The basis used by marketers to classify the products are: (i) On the basis of durability and tangibility of the product (ii) On the basis of Consumer Shopping Habits and (iii) On the basis of use in industry.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 10, Issue 04, April 2020)

Traditionally products have various classifications, consumer products and industrial/business products is based on use is classified. Consumer product is further classified as: convenience products, shopping products, specialty products and unsought products and industrial product is classified as: material and parts (these goods enter the product directly), installation, accessory equipment, operating supplies. In general, the consumer goods directed to ultimate end-users, but the buyers seek product and marketing is important for it, but the industrial products used to produce other products and services, and usually vendors seek buyers. Based on durability it's classified as durable and non-durable and few classified products on the basis of tangibility, tangible and non-tangible also. But It is vital to design a product classification method in order to support global supply planning and improve the supply chain performance.

A. Product classifications

With respect to the present study and to measure the impact of COVID-19 on SCMs, the products are classified on basis of exigency as export products, import products, elective products and emergency products. Categorization of the product is important one for those organizations which were dealing with supply chain, the basis they can maintain it in their data base and exploit it at the time of need. This categorization helps to do planning accordingly by evaluating its gravity to the present situation of the demand.

Export products: those products which is manufactured by manufacturer for the specifically export purposes.

Import products: The product which will arrive from other country.

Elective products: Also called optional product which have substitute for the customer.

Emergency products: In this category those products will lie which are critically necessary for providing care for people. Viz. medical devices, medical products, productive equipment, masks, screeners, disinfectants and so on.

Companies use to handle complicated supply chain operation but lacks in product classification system. It is vital to design a product classification method in order to support global supply planning and improve the supply chain performance. That part of this research for future work. Also, Product Classification System is supposed to be applied in supply planning prioritization, inventory management and workforce balancing, etc.

The things which are considered as mechanism for the propagation of economic shocks, or economic contagion are: Products (tangible), people, services, know-how(skill), financial capital, foreign direct investment (FDI), international banking, and exchange rates. The flow of global economies dependencies is connected by cross-border by these mechanisms, and product is considered at top priority among all the above. Companies are learning the lesson that global supply chains can be abruptly broken by a health shock and will adjust accordingly. Financial intermediaries and regulators are likely to incorporate pandemic shocks into their future risk assessments and stress tests.

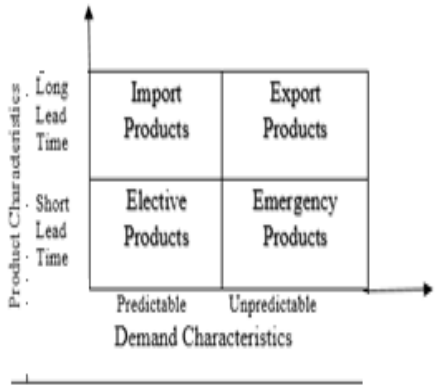


Figure 2. Characteristics of Export, Import, Elective and Emergency Product

V. Data Collection And Analysis

A. Questionnaire development

A questionnaire was developed based on the practical practice of supply chain in various organizations proposed for the study and aimed at gathering information related to dealing with the basis of classification of the product and its impact in context to Indian supply chain industry. With the help of available literature and telephonically talk with the experts the questionnaire is developed and suggestions were incorporated and questionnaire was improved before distribution for data collection. Survey form was developed with the help of google forms. The survey questionnaire is available in Appendix A. The survey was sent to the top management person who was handling these organizations via mail as well as in person asking people to participate in the survey. The given deadline for data collection for the responses was of 26th March to 20th April 2020.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 10, Issue 04, April 2020)

B. Questionnaire administration

The questionnaire was self-administered, and the survey for the study with target population of supply chain industries were conducted through google form, online. The survey link was sent on email addresses of the respondents with a cover letter. Nowadays, web-based survey is being widely used by researchers because of less time required, ease of availability, real time data compilation, and lower cost than other surveys. With the help of google search engine, the dedicated organization were searched with their e-mail address to consult for the survey. Justdial.com helps more to lean the time of searching to those companies. To cover the entire India, the companies of Delhi, Kolkata, Mumbai, Chennai, Bangalore, Hyderabad and Ahmedabad were selected, besides these cities Pune, Goa, Nagpur, Coimbatore, Indore, Nashik, Vadodara, Jaipur, Mysore, Vizag, Chandigarh and Surat, many companies are there which associated with supply chain operations. The questionnaires were sent to the respondents in single phases. 300 forms were sent to different organizations through google forms and responses were analyzed. Total 130 responses were received 35 were incomplete so discarded. An effective response rate of 43.3% was achieved which is found adequate for the present study.

C. Respondent's profile and non-response bias

Due to time constraint, in single phases, total 300 e-mails were sent to the selected supply chain dedicated organizations for the survey. Overall, 130 responses were found suitable for the study. Participating organizations in the survey belong to (Original Equipment Manufacturer) OEM, Fast Moving Consumer Goods (FMCGs) and SCM consultant. Respondents who have participated in the survey represent different levels of management including top position to middle level and executive level employees and mostly belong to logistics department, marketing & sales, corporate strategy, and supply chain departments. Non-response bias with web-based survey was accessed by splitting the sample into two groups based on receiving date of the responses. Early response group consisted of 67 responses and late response group consisted of 63 responses. The t-test was performed on the difference in early response group and late response group. It yielded no statistically significant difference between the mean of two groups ($p > 0.05$). In addition, statistical tests show that there is no significant difference among samples.

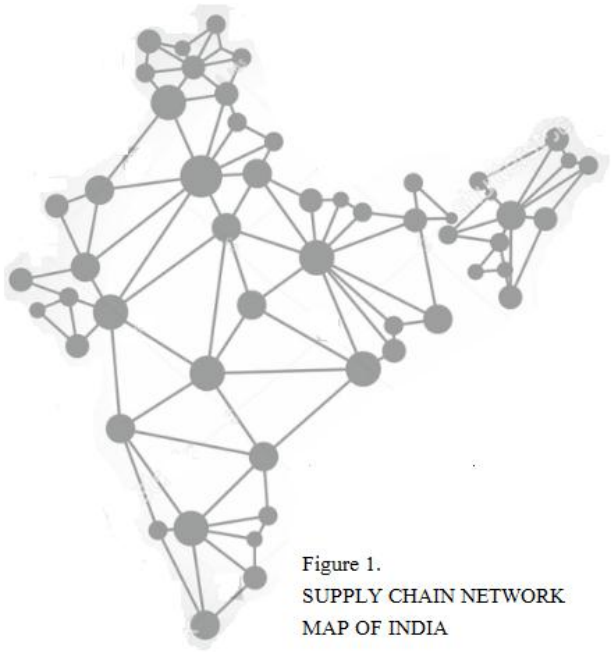


Figure 1.
SUPPLY CHAIN NETWORK MAP OF INDIA

VI. Result And Discussion

To understand more about the global impact of COVID-19 on supply chain management, we designed a questionnaire that addressed four pertinent themes including export product, import product, elective product and emergency product. The questionnaire consisted of 7 multiple choice questions and three open-ended questions (appendix A). The questionnaire was targeted at supply chain organizations and was distributed online in interest groups using various social media platforms. The questionnaire was open from 26th March to 20th April 2020 and did not collect any identifiable data. Valid responses were represented as categorical data and presented in percentages.

Impact studies with individual SCM organization data

This section describes a new wave of research that is based on samples of organization collected precisely to study strength of chain of supply with the help of product classification. The sampling was designed to examine organizations in different product category zones and to select organizations within country across a wide range of product. The survey instrument was designed to measure accessibility of products in places without barriers and hindering when situation is unfavorable.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 10, Issue 04, April 2020)

The instrument collected information about the choices that organizations made: which product to ship, which product to hold and which inputs to purchase. Precise Products classification give solutions and support decision-making for Global Supply Planning. For example, higher priority of products requires shorter lead time of supplying. The crisis never took place for time with the organization, but the plan was put in place and organization feel secure knowing that should it happen, they would be safe. For all the effort, resources, and consideration put into designing the most efficient manufacturing facilities and systems, it should be a natural next step to protect them by developing crisis response plans. Fisher is an acknowledged pioneer in the classification of supply chain strategies, claiming that there are basically two types of supply chain strategies, physically efficient and market responsive (Harris, 2007).

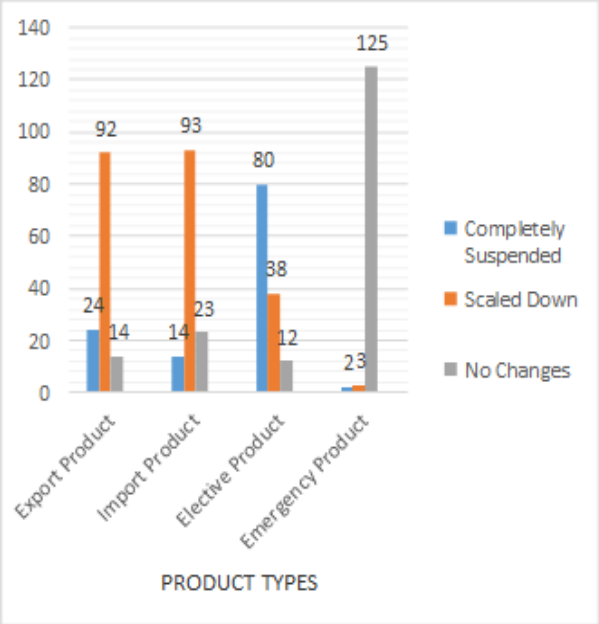
A total of 130 individual responses were obtained from the supply chain firms. Of these, 32(24.6%) were from the Delhi, 21 (16.1%) were from Kolkata, 10 (07.6%) were from Mumbai, 17 (13.0%) were from the Chennai, 17 (13.0%) were from Bangalore,19 (14.6%) were from Hyderabad and 14 (10.7%) were from Ahmedabad. Of the 130 respondents, 85(65.3%) were working Supply Chain as a part of OEM. When asked about export product, 102 (78.4%) replied that export product orders were either suspended or scaled down. When asked about import product, 118 (90.7%) replied that import product were either suspended or scaled down. With regards to elective products, 120 (92.3%) replied that at least some, if not all elective products orders were cancelled. The majority of orders cancelled were due to its availability of substitute product or not felt essential and emergency by the customer. On the contrary, 125 out of 130 respondents (96.2%) are still delivering all emergency products (Figure 1). For a large majority of respondents (113 out of 130, 87.0%), face-to-face multidisciplinary or team meetings have been reduced or cancelled entirely, and have been replaced by online messaging platforms, video conferencing, teleconferencing or email discussions. The rate of response also reflects the effect of lockdown.

As only 13% response from Mumbai and also we got news that Maharashtra counts highest death due to COVID-19. The impact of COVID-19 on supply chain services has been far-reaching. The majority of respondents have reported a reduction or cessation of export products orders and elective product orders, whilst continuing with emergency products.

Numbers from group purchasing organization Premier and data software companies show the supply chain gaps in getting needed personal protective equipment (PPE) to those on the front lines. Data on increasing orders of PPE in the hospital setting and the arc of COVID-19 calls compared to a typical flu season show the pandemic in a new light. Healthcare systems, distributors and purchasing organizations, along with state and federal governments, rely on real-time data to manage supplies and model their equipment and staffing needs. so manufacturers and their complex webs of suppliers, who in many cases lack that same experience or large-scale backup plans. The result: a chaotic scramble resulting in production delays, product shortages, and higher prices.

E-commerce giant Amazon has announced that it will only deliver high-priority products. In view of the challenges faced by the delivery executives, the Karnataka government on Monday issued a notice stating that e-commerce and home delivery fall under essential services and hence should be exempted from the restrictions imposed on citizens to curb the spread of the coronavirus. Following the suit, Maharashtra too issued an order stating that the e-commerce services such as food, medicine, etc should not be included in the restrictions that are imposed to contain the virus. Grofers founder had tweeted his apology to the customers for not being able to deliver their products on time. He stated “We @grofers are having a lot of trouble getting essentials to people who need them as our warehouses are asked to shut, and trucks and delivery partners are being stopped by the police. We apologize to our customers and are working hard to find a solution.” Flipkart also suspended its services temporarily.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 10, Issue 04, April 2020)



REFERENCES

[1] Reina Angkiriwang, Nyoman Pujawan & Budi Santosa. 2014. Managing uncertainty through supply chain flexibility: reactive vs. proactive approaches. Production & Manufacturing Research, 2:1, 50-70, DOI: 10.1080/21693277.2014.882804

[2] Buyurgan, N., Zhang, S., & Okyay, H. K. (2019). Methodical analysis of inventory discrepancy under conditions of uncertainty in supply chain management. International Journal of Logistics Systems and Management, 32(2),

[3] Patrik Alfredsson, Jos Verrijdt. (1999). Modeling Emergency Supply Flexibility in a Two-Echelon Inventory System. Management Science45(10):1416-1431.

[4] Richard Baldwin and Beatrice Weder di Mauro, "Economics in the Time of COVID-19", A CEPR Press VoxEU.org eBook, ISBN: 978-1-912179-28-2.

[5] Prater, E., Biehl, M. and Smith, M.A. (2001). International supply chain agility: tradeoffs between flexibility and uncertainty. International Journal of Operations and Production Management, Vol. 21 Nos 5/6, pp. 823-839.

[6] Vickery, S. nee, Calantone, R., & Droge, C. (1999). Supply Chain Flexibility: An Empirical Study. The Journal of Supply Chain Management, 35(3), 16–24.

[7] Duclos, L.K., Vokurka, R.J. and Lummus, R.R. (2003). A conceptual model of supply chain flexibility. Industrial Management & Data Systems, Vol. 103 No. 6, pp. 446-456.

[8] McKibbin, Warwick J. and Fernando, Roshen. (2020). The Global Macroeconomic Impacts of COVID-19: Seven Scenarios (March 2, 2020). CAMA Working Paper No. 19/2020.

[9] MacKenzie, C. A., Barker, K., & Santos, J. R. (2014). Modeling a severe supply chain disruption and post-disaster decision making with application to the Japanese earthquake and tsunami. IIE Transactions, 46(12), 1243–1260.

[10] Craighead, C.W., Blackhurst, J., Rungtusanatham, M.J. and Handfield, R.B. (2007) The severity of supply chain disruptions: Design characteristics and mitigation capabilities. Decision Sciences, 38(1), 131–156.

[11] Harris, G.A. (2007). Alignment of Supply Chain Strategy with Product Characteristics. Ph.D. diss. The University of Alabama in Huntsville.

[12] Agrawal, S., Singh, R. K. (2019). Analyzing disposition decisions for sustainable reverse logistics: Triple Bottom Line approach. Resources, Conservation & Recycling, 150 (2019),104448

[13] https://www.idc.com/url.do?url=/getdoc.jsp?containerId=prAP46181620&position=5&transactionId=74540281&term=&page=1&perPage=25

[14] https://www.supplychainbrain.com/articles/15619-indias-supply-chain-is-improving-but-still-has-a-long-way-to-go

[15] https://phys.org/news/2020-04-global-uncertainty-businesses-flexibility-efficiency.html?utm_source=TrendMD&utm_medium=cpc&utm_campaign=Phys.org_TrendMD_1

[16] https://www.economist.com/finance-and-economics/2020/02/22/covid-19-presents-economic-policymakers-with-anew-sort-of- threat?cid1=cust/ednew/n/bl/n/2020/02/

[17] https://www.justdial.com/Delhi/Supply-Chain-Management-Companies/nct-10808773

Appendix A

Questionnaire

Name of the organization:
_____
Year of establishment:
_____
Annual Turnover (in Rupees):
_____
Number of Employees:
_____

1. What is your country of practice of supply chain

2. Is your company is dealing with Supply Chain as a part of OEM or as a consultant company?

a. Yes b. No c. Other (please specify)

3. How has COVID-19 affected your export product order demand?

a. export product order completely cancelled

b. export product order scaled down i.e. non-urgent products postponed/cancelled

c. export product order demands operate as per usual

d. Other (please specify)

4. How has COVID-19 affected your import product demand?

a. import product completely cancelled

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 10, Issue 04, April 2020)

b. import product scaled down i.e. reduction in number of orders

c. import product demand as per usual

d. Other (please specify)

5. How has COVID-19 affected your elective product demand?

a. All elective product demands are cancelled

b. Some elective product demands are cancelled

c. No elective product demands are cancelled

d. Other (please specify)

6. If elective product demands have been cancelled, what type of products are they? Leave blank if no elective product demands have been cancelled.

7. How has COVID-19 affected your emergency product demand?

a. All emergency product orders are rejected

b. Only some emergency product orders are rejected

c. No emergency product orders are rejected

d. Other (please specify)

8. If emergency product orders have been rejected, what type of products are they? Leave blank if no emergency product orders have been rejected.

9. How has COVID-19 affected face-to-face multidisciplinary meetings?

a. All face-to-face multidisciplinary meetings are cancelled

b. Some face-to-face multidisciplinary meetings are cancelled

c. No face-to-face multidisciplinary meetings are cancelled

d. Other (please specify)

10. What modality has replaced face-to-face multidisciplinary meetings? Leave blank if your face-to-face multidisciplinary meetings have not been cancelled. (You can choose more than 1 option)

a. Teleconference

b. Video conference

c. Online messaging platforms i.e. WhatsApp, Telegram, etc

d. Email e. Other (please specify)

About Author(s)

Mr. Rakesh Kumar received Bachelors in Industrial Engineering and Master's in Mechanical Engineering with specialization in Computer Integrated Manufacturing from India. He is currently a assistant professor at Delhi Technological University in Department of Mechanical Engineering. His research interest includes *Total Quality Management, Reliability Engineering, Operations and Supply chain, Fuzzy Logic, AI and Decision Science*.

R.S. Mishra (Professor) received the Ph.D. degree in Mechanical engineering from IIT Delhi, India and currently at Delhi Technological University serving since 1997. He having 150 research papers in peer reviewed in International journals and 75 research papers in proceedings of International and National conferences. Prior to joining DTU, he served 09 years at different engineering colleges of India, viz, REC Hamirpur, PAU Ludhiana (Punjab)and HAU, Hisar (Haryana.).His research area includes *Green Technology, Refrigeration & Air Conditioning, Metal Cutting Technology, Total Quality Management, Solar Energy Technology, Bio Fuels, power Plant engineering*.

View publication stats

Proceedings of the Fourth International Conference on Trends in Electronics and Informatics (ICOEI 2020)
IEEE Xplore Part Number: CFP20J32-ART; ISBN: 978-1-7281-5518-0

Deep Learning Approach for Arbitrary Image Style Fusion and Transformation using SANET model

[1]Pratibha Rathi
Department of Computer Science & Engineering
Delhi Technological University
Delhi, India
pratibharathi96@gmail.com

[2]Pranav Adarsh
Department of Computer Science & Engineering
Delhi Technological University
Delhi, India
pranavadarsh95@gmail.com

[3]Dr. Manoj Kumar
Department of Computer Science & Engineering
Delhi Technological University
Delhi, India
mkumarg@dce.ac.in

***Abstract*—For real-time applications of arbitrary style transformation, there is a trade-off between the quality of results and the running time of existing algorithms. Hence, it is required to maintain the equilibrium of the quality of generated artwork with the speed of execution. It's complicated for the present arbitrary style-transformation procedures to preserve the structure of content-image while blending with the design and pattern of style-image. This paper presents the implementation of a network using SANET models for generating impressive artworks. It is flexible in the fusion of new style characteristics while sustaining the semantic-structure of the content-image. The identity-loss function helps to minimize the overall loss and conserves the spatial-arrangement of content. The results demonstrate that this method is practically efficient, and therefore it can be employed for real-time fusion and transformation using arbitrary styles.**

Keywords—image processing; deep learning; neural style transfer; computer vision; arbitrary image stylization; SANET; ASPM; convolutional neural networks.

I. INTRODUCTION

In the recent few years, various research works contributed to improve and accelerate the development of deep learning methods for continuous reformations of style transformation procedures. Arbitrary style transformation is the process of fusing new artistic styles with the content of the given image to create impressive artworks. NST was first introduced and explained by Leon A. gatys in 2015. Based on this idea presented by Gatys, many NST applications like a deep dream, prisma, and ostagram, etc. have been released to generate amazing artworks. The method proposed by Gatys [1] [2] follows an iterative procedure for the optimization of the generated image, which makes it computationally expensive and slow. However, the quality of the generated output is quite impressive. After that, numerous models have been proposed based on the optimization of the employed model. Johnson et al. [3] and Ulyanov et al. [4] [5] proposed fast offline approaches for generating output in a single feed-forward-pass, but their model is limited in the number of styles, and the quality of the result is also reduced. For fusing arbitrary styles using a single model, some ASPM approaches have been introduced. The swap model [6] was stated that employs the patch match procedure to substitute the features of content by the nearest match of characteristics of the style. Li et al. [7] specified multilevel stylization by recursively employing whitening-color-transformation to enhance the quality of output and to protect the content-structure. AdaIN [8] model changes the mean-variance of the content-image to match with style-image for transforming features. Avatar-net [9] model contains a style-decoder model based on patch-match and an hourglass network for adapting styles on multiscale. To solve the trade-off between content and style losses, this paper presents a new loss function identity-loss and a network consist of an integration of two SANET models for preserving both content-structures and style-characteristics. The arrangement of the paper is as follows. Below in section 2, background information of style transformation methods is given, which covers the explanation of some IBAR methods along with their drawbacks. Section 3 provides information about the existing neural methods of style transformation and their categorization into online and offline divisions. Section 4 illustrates the analysis of optimization approaches and expansion areas of NST by providing some detailed information about them. Section 5 demonstrates the implementation results and its comparison with other existing models. Finally, section 6 summarizes the conclusion.

II. BACKGROUND

In this section, the background information of the approaches used in style transformation before the evolution of deep learning is explained. The techniques in the taxonomy of image-based artistic rendering (IBAR) were employed for the stylization of two-dimensional images. First, the procedure of

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:25:38 UTC from IEEE Xplore. Restrictions apply.

some IBAR [10] methods are described and then their drawbacks are reviewed.

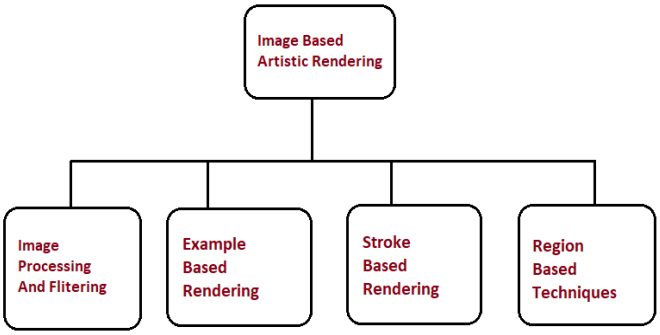


Fig.1: Categorization of IBAR techniques

A. *Stroke-based rendering*

It is the process of integrating portraits, paintings, and artworks by providing the strokes or marks on a digital canvas using brushes, lines, and tiles [11]. For converting a photograph into a particular style, virtually some brush strokes are superimposed at some positions on a digital canvas. It can create conventional arts based on the brush, for instance, an oil painting. This procedure generally starts from a source picture by sequentially adding strokes to modify that picture and ultimately composing a nonphotorealistic image that looks like the original photograph carrying an exquisite style. The limitation of SBR methods is that all SBR approaches are intended to create a single distinct style and therefore is not proficient for composing arbitrary styles. It illustrates the non-flexibility of this approach.

B. *Region-based techniques*

RBT is used to accommodate the effects based on the content in regions. Previous RBT algorithms [12] deteriorate the shape of structures in managing the arrangement of strokes. The latter method introduces an IBAR procedure [13] based on regions to alter the semantic geometry of photographs for generating creative effects. This algorithm replaces parts of the picture by many canonical patterns for constructing uncomplicated shape rendering results. In this approach, a source image and target photograph are distributed into many areas using the method based on graphs. The semantic conformity is estimated between them based on colors, patterns, and texture. Ultimately, the colors and textures are transferred for each area by employing a patch match method. Drawbacks of RBT methods are similar to SBR, all RBT approaches are intended to create a single distinct style and therefore is not proficient for composing arbitrary styles. It represents the non-flexibility of this approach.

C. *Image-processing and filtering*

Usually, the IPF approach for rendering is simple to implement and effective in application. IPF techniques [14] apply to real-time rendering and accommodate multicore CPU and GPU. Bilateral and Gaussian filters are employed to compose the effects of cartoons. Kuwahara filter is an exciting range of filters for preserving edges. It works relatively well on pictures of high contrast. These filters eliminate features in sharp regions and preserve shape outlines in areas of low contrast. Some other filters are morphological, diffusion, and shock filter. Gradient-domain techniques are also utilized in which the result is interpreted by creating a gradient field. The drawback of IPF is that style variety, and diversity is inadequate.

D. *Example based techniques*

EBR [15] understands the mapping within the prototype pairs containing the source photograph and the corresponding styled artwork. The learned mapping is employed to style arbitrary pictures. EBR can work in two categories by performing texture transformation and color transformation. The proposed framework named image analogy is a supervised method containing two stages, namely the design stage and application stage. In the design stage, it intends to study the analogous mapping between the pairs of source pictures and target outputs. The data used for training in image-analogy contains combinations of input-pictures along with the corresponding stylized output-pictures. In the application phase, learned mapping is applied to new target photographs to produce relevant analogous results. The main drawback of this method is the unavailability of training data. This method fails to capture the structure of images accurately and uses only low-level features of the picture.

III. NEURAL STYLE TRANSFER

NST is a machine learning technique that takes two images in input, one of which is the content-input image and another one is called a style-input image, for example, the artwork in a famous painting and blend them to transform the input images into an output image that looks like the picture called content-image but it will be painted similarly to the picture called style-image. Neural style transfer uses deep learning and its convolution neural networks to perform this amazing task. It is much propitious for artists and widely used by them to impose their artwork on digital photographs. It was first introduced and explained by Leon A. Gatys in 2015. After the Gayts, several endeavors are in progress related to the style transfer and its algorithms to perform training and learning faster as much as possible and to extend the procedure of style transfer techniques from static images to videos, audios, and other mediums. Neural methods of style transfer are broadly divided into two categories. One of which is the online method and the other one is the offline method.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:25:38 UTC from IEEE Xplore. Restrictions apply.

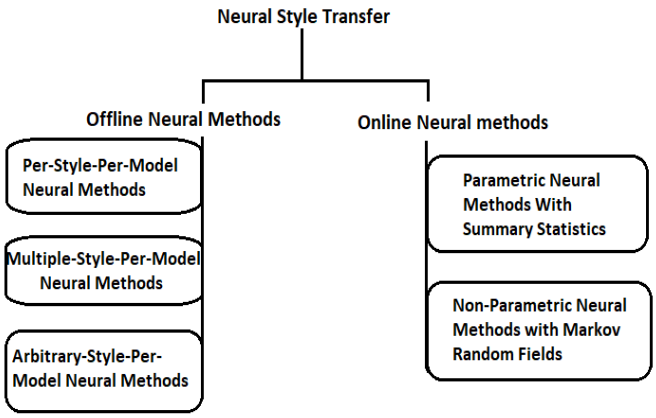


Fig.2: Categorization of NST techniques

A. *Online neural methods*

One of the famous applications in the field of style transfer is Deep-dream, which is based on the online approach of NST. The online method focuses on the optimization of the generated output image but not on the improvement of the applied model. This approach employed several iterations for enhancing the quality of the output image. Due to the multiple iterations, this procedure is much time consuming and computationally expansive. Therefore it is also known as the slow neural method for a style transformation. The process is to extract the content information and style representation from input images and then apply this extracted information to the CNN model for recombining both components to generate the artistic output. Gradient descent algorithm is used for the optimization of the result by minimizing the value of loss function. These online methods differ in the approach of applying the content and style representation to the CNN model so categorized into two classes.

1) Parametric online neural methods: These methods employed spatial representation summary and statistics to generate artworks. Gatys et al. [1] [2] in 2015 proposed an algorithm that applies deep CNN model VGG 19 for extracting structure, shape, and texture, etc. related information of content and style features from preprocessed input pictures. This information is spatially modeled in different layers of VGG 19 and used for the reconstruction of target artwork. VGG nineteen network is composed of sixteen layers of convolution and five layers for pooling, where higher layers are handled to represent the content characteristics. Style-image input is passed through the VGG network, and the gram matrix is employed to store style characteristics. The overall loss is a linear combination of both content and style losses balanced by parameters calculated through the one thousand iterations of gradient descent. Content loss is computed in terms of mean squared difference. The α and β hyperparameters provide the appropriate weights to both components of loss for balancing the equation of total loss function.

$$L_{total} = \alpha L_{content} + \beta L_{style} \quad (1)$$

$$L_{content}(C, G, L) = 0.5 \sum_{ij} \left(a[L](C)_{ij} - a[L](G)_{ij}\right)^2 \quad (2)$$

$$L_{GM}(S, G, l) = \frac{1}{4N_l^2 M_l^2} \sum_{ij} \left(GM[l](S)_{ij} - GM[l](G)_{ij}\right)^2 \quad (3)$$

$$L_{style}(S, G) = \sum_{l=0}^{L} w_l \times L_{GM}(S, G, l) \quad (4)$$

Here S, C, G, and GM are used to represent style-image, content-image, output generated-image, and gram-matrix, respectively. Activation of layer L is used for the calculation of content-loss. N_l is the number of channels in the feature map, and M_l is the product of the height and width of the feature map. After that demystifying approach [16] of NST was introduced that employs a different method for the statistical representation of style features. In this procedure, the gram matrix was replaced by MMD. It states that the minimization of maximum-mean-dispersion is comparable to the matching of gram-matrices. Several different methods for the representation of style were used, which include Gaussian, linear, polynomial and batch-normalization, etc. In 2017, Lapstyle NST was presented [17], which added a laplacian-loss function to calculate the overall loss. This laplacian-loss preserves the low-level fine details of the content characteristics and is aimed to reduce the distortion of edges, shapes, structure, contour, and colors of original input when converted to artistic output.

2) Non-parametric online neural methods: In 2016, Li and wand [18] proposed a method of style transformation that employed a combination of Markov-random-field models and deep CNN for synthesizing two-dimensional artworks. This approach works well with both photo-realistic and non-photo-realistic transformation of style. For minimizing loss, it uses limited memory BFGS algorithm. This algorithm matches style on the level of neural-patches. Hence, it is suitable for preserving the fine details of the structure. They formed a new function for calculating the style-loss and content-loss by calculating Euclidean distance between patches using the energy function of MRF-neural-patch-matching. The characteristic of this algorithm is that it produces considerably well-resulting artwork for photo-realistic styles, especially if the shape, structure, and panorama of both content-image and style-image are alike. Though, it usually doesn't work well if there is a great difference in view and arrangement of structures of content-image and style-image because the patches in both pictures could not be paired accurately.

B. *Offline neural methods*

Earlier explained the online method uses the iterative approach of gradient descent for improving the quality of generated output. Therefore it consumes a large amount of computation time when image size is considerably large. To overcome this limitation of speed, offline neural methods were introduced. This approach is much speedy compared to the previous procedure because it generates the output artwork in a single feed-forward-pass by optimizing the adopted model. As it mainly focuses on the optimization of the applied model, hence it is also known as the model-optimization-based offline method of NST. This offline approach is further divided into three categories based on the number of styles generated by the model in its output artwork.

1) The offline neural model generating single style-PSPM: For adopting an individual style, this approach optimizes the employed model. Johnson et al. [3] proposed a parametric method of training a feed-forward-network for the task of style transformation. This method replaces the per pixel losses by perceptual losses computed using high-level features of

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:25:38 UTC from IEEE Xplore. Restrictions apply.

images. The system contains two components of the network, which are image-style transformation network and loss-network. Image-style transformation network is a deep residual CNN that can convert input images to an artistic output by understanding the mapping. It is trained using stochastic-gradient-descent for minimizing the weighted combination of loss-functions, which is computed by employing loss-network. Perceptual losses measure the similarities and dissimilarities of images more precisely compared to per pixel losses. Another similar approach proposed by Ulyanov et al. [4] uses multi-scale architectures that can be trained faster and result in a smaller loss in content and texture characteristics and better visual quality while working with comparatively few parameters. These are very lightweight models that are capable of generating artworks comparable to the quality of output produced by Gatys et al., and their speed is almost 100 times faster. However, these are also limited in considering the small details and depth information. After that, a new advance concept was introduced, which is known as instance-normalization [5]. It means normalization is applied to every single image in place of the batch of images. It converges faster, learns quickly, and significantly improves the visual quality of generated stylized artwork. This model can be trained rapidly, reduces the overall loss, and enhances the diversity of network output.

The non-parametric approach of single-style per model offline neural methods uses MGAN [19]. It is a markovian model applied to generative adversarial networks for learning the mapping between different representations of the same content-image. MGAN utilized the imagenet dataset for adversarial-training. It successfully preserved the coherent texture of compact content-images, and the speed of synthesis is remarkably fast. MGAN network consists of two subnetworks, namely discriminator and generator, for improving the model in iterations. This method lacks in transferring non-texture style, such as facial-features of two different face-pictures. Facial-features cannot be interpreted as texture because these require the understanding of expression, poses, and gender, etc. semantic characteristics.

2) The offline neural model generating multiple styles-MSPM: Above mentioned PSPM procedures can generate artistic output hundred times speedier than previous offline neural methods based on image-optimization. But it needs to train several network models individually to generate output for every distinct style-image. It exhibits inflexibility and excessive time consumption in training for each style. For reducing the irrelevant time consumption during the separate training of various models in PSPM, a new approach was introduced for consolidating multiple-styles into a single model, which is known as MSPM. In 2017, Dumoulin et al. [20] proposed that it is possible to model different styles at the same time by using the same parameters in the convolutional network. For incorporating different styles, shifting and scaling of parameters is required in the layers of instance normalization, which is called conditional-instance-normalization. CIN is an alteration of style-transformation networks that is uncomplicated, scalable, flexible, and efficient for incorporating diversity in output artwork. By consolidating affine parameters of different styles, it can be applied to incorporate multiple-styles in one output artwork.

After this style-bank method [21] was introduced, that is an aggregate of various convolution filter-banks. One style is expressed by every filter-bank explicitly in that combination. For converting a picture to a specific style, the corresponding filter-bank is turned over the intermediate feature-embedding generated by a single autoencoder. While keeping a fixed autoencoder, it is possible to carry out incremental-learning for appending a new style by training a new filter-bank. This style-bank fusion approach works in two ways, linear and region-based. In linear fusion, the style-bank layer is fed with a linear combination of multiple styles. While in region-based fusion, content-image is disintegrated into some disjoint regions using automatic k-means clustering, and then combined style transformation can be conducted using more than one style simultaneously. But the size of the model expands as the number of styles learned by it increases, which is a drawback. To eliminate this limitation, Zhang and Dana [22] explored the capabilities of a single CNN model VGG 16 for consolidating both content and style in one network. The Multiple-style-generative network was introduced, which can control the size of brush strokes in the real-time transfer of style. In MSG-Net, the comatch layer was used to represent styles in two-dimensional ways. For the comatch layer, it is possible to differentiate and train end to end, but tough to conserve the fine details of structures.

3) The offline neural model generating arbitrary styles-ASPM: The third type of offline neural method is ASPM that intends to build one model for all styles. It states that a single model can be trained to transfer any arbitrary style. ASPM is also divided into two types. Parametric-ASPM employed spatial representation summary and statistics to generate artworks while nonparametric-ASPM applies MRF for texture-modeling.

Style-Swap: Chen and Schmidt [6] proposed the first algorithm for nonparametric-ASPM in which they introduced the concept of style-swap for training an inverse network to generate the stylized output for any arbitrary style. At first, a set of patches having sufficient overlap for both content and style activations are extracted for computing in a previously trained VGG network. Each content patch is matched with style patches to find the closest and similar match for swapping. Then content activation is reconstructed by averaging areas of overlapping. This approach of one-model-for-any-style is more flexible than the previous methods but generates less appealing results in which the content is preserved thoroughly, whereas the style is not reflected adequately.

AdaIN: In parametric-ASPM, Huang, and Belongie [8] modified conditional-instance-normalization (CIN) to adaptive-instance-normalization (AdaIN) instead of training a network for parameter prediction. The above-mentioned style swap method takes a considerable amount of time and occupies much memory, but AdaIN is a simplistic approach with the low cost of computation. For the given input pictures, AdaIN modifies the mean value and variance value of both content-image and style-image for matching. AdaIN layer efficiently

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:25:38 UTC from IEEE Xplore. Restrictions apply.

transfers the feature-characteristic statistics for performing style transformation. After that, a decoder is used for inverting the output of the AdaIN layer, and the VGG encoder is utilized for measuring loss. It is a good parametric-ASPM approach for performing stylization in real-time, but it is insufficient in the generalization of new styles. Therefore, it is challenging to synthesize intricate style patterns consist of fine details and complex structures.

WCT: Li et al. [7] changed the AdaIN-layer of the above-described model by a combination of the whitening-color-transformation network. The objective is to protect the arrangement and composition of content. Whitening-transformation receives content-activations from the encoder and then produces a filtered representation of content-image. After that, coloring-transformation is applied for consolidating the style-patterns into the filtered content design, and the decoder provides stylized output artwork. This method is good in generalizing distinct styles as learning is not required of every style, but not useful in delivering sharp-features and small strokes-details.

Avatar-net: It blends the images based on the semantic-arrangement of content and employs multi-scale stylization in a single feed-forward pass [9]. It is better than WCT that requires several recursive feed-forward passes. Avatar-net uses a style-decorator and an hourglass-network. Patch-based style decorator module is used to decorate the content features with style patterns while retaining the semantic properties of content. The hourglass network is employed for a multi-scale adaptation of style. But its output depends on the patch size, and therefore usually, it is unable to express the local and global styles simultaneously.

SANET Model: It stands for the style attentional network that can blend style-patterns in the content-image efficiently and flexibly [23]. It uses a kernel that can learn similarities rather than fixed kernels. This neural-network model modifies the self-attention mechanism to understand the mapping between content and style features. For synthesizing artworks, this model takes two images in input, which are content-image I_c and style-image I_s. By applying the encoder module, which is a previously trained VGG nineteen network, both the input-images are encoded into an intermediate representation. After encoding both input-images, both the obtained feature-maps F_c and F_s, are applied to two distinct style attentional networks to produce the combined feature-map F_{cs}.

$$F_c = E(I_c) \tag{5}$$

$$F_s = E(I_s) \tag{6}$$

$$F_{cs}^{r41} = SANET_{r41}(F_c, F_s) \tag{7}$$

$$F_{cs}^{r51} = SANET_{r51}(F_c, F_s) \tag{8}$$

$$F_{csc}^{r41} = F_c + W_{cs} F_{cs}^{r41} \tag{9}$$

$$F_{csc}^{r51} = F_c' + W_{cs}' F_{cs}^{r51} \tag{10}$$

$$F_{csc} = conv_{3\times3}\left(F_{csc}^{r41} + upsamle\left(F_{csc}^{r51}\right)\right) \tag{11}$$

$$I_o = D(F_{csc}) \tag{12}$$

Here F_{csc}^{r41} and F_{csc}^{r51} are the synthesized feature-map values generated by applying 1xl conv-net to the output feature-maps of SANETs. F_{csc} is the integrated output feature-map achieved by applying 3x3 conv-net to the resulting feature-maps obtained by relu_4_1 SANET and upsampled relu_5_1 SANET. The combined output of these SANETs is concatenated by passing through a symmetric-decoder, which produces the final output-image I_o. This method is both effective and practically efficient. It flexibly performs style transformation using the loss function that unites traditional style-reconstruction-losses and the newly introduced identity-loss.

$$L = w_c L_c + w_s L_s + L_{id} \tag{13}$$

$$L_c = \left\| \overline{E(I_o)^{r41}} - \overline{F_c^{r41}} \right\|^2 + \left\| \overline{E(I_o)^{r51}} - \overline{F_c^{r51}} \right\|^2 \tag{14}$$

$$L_s = \sum_{i=1}^{L} \left\| \mu(\varphi_i(I_o)) - \mu(\varphi_i(I_s)) \right\|^2 + \left\| \sigma(\varphi_i(I_o)) - \sigma(\varphi_i(I_s)) \right\|^2 \tag{15}$$

$$L_{id} = w_{id1}(\|I_{cc} - I_c\|^2 + \|I_{ss} - I_s\|^2) + w_{id2} \sum_{i=1}^{L} \begin{pmatrix} \|\varphi_i(I_{cc}) - \varphi_i(I_c)\|^2 \\ + \|\varphi_i(I_{ss}) - \varphi_i(I_s)\|^2 \end{pmatrix} \tag{16}$$

Here L_c is content-loss, L_s is style-loss and w_c, w_s are the weights assigned respectively. $\overline{F_c^{r41}}$ and $\overline{F_c^{r51}}$ are channel-wise-normalized mean, variance feature-map values of content-image. $\overline{E(I_o)^{r41}}$ and $\overline{E(I_o)^{r51}}$ are channel-wise-normalized mean, variance feature-map values of output-image after passing through the encoder. φ is the feature-map value of the layer of encoder denoted by i. L_{id} is identity-loss and w_{id1}, w_{id2} are the weights assigned to identity-loss. This loss function helps to preserve the structure of content-image while simultaneously maintains the style-patterns.

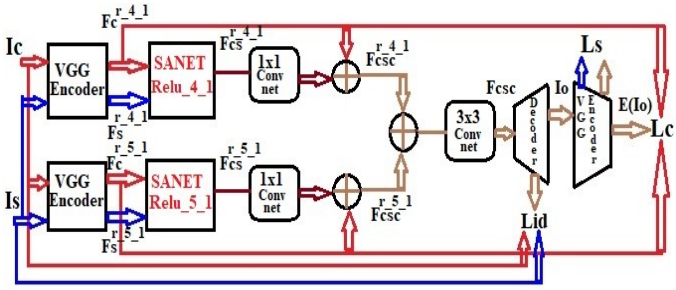


Fig.3: Network Architecture of SANET model

IV. OPTIMIZATION AND EXPANSION OF NST

Optimization algorithms are applied to optimize the output of NST by minimizing the value of loss-function. They work during training and learning of the model to update the weights and finding the relevant values for parameters to get an optimal solution. In Table 1, some popular optimization methods [24] are discussed with their merits and drawbacks.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:25:38 UTC from IEEE Xplore. Restrictions apply.

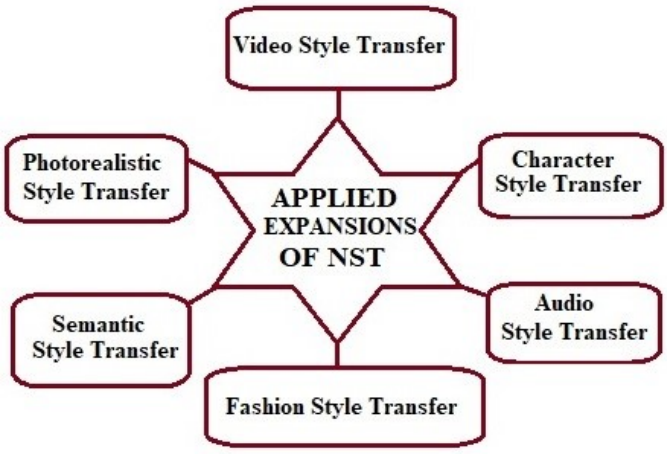


Fig.4: Expansion of NST in various applications

approach [26] was suggested that applies a flow-subnetwork to compose feature-flow and to consolidate the knowledge of optical-flow. This model is based on a network that contains a previously trained combination of encoder and decoder. The semantic transformation [27] of style is an expansion of NST, which is applied to a pair of input images holding analogous content. The method is to obtain a semantic correlation between content-image and style-image. Hence, the style of every region of style-image is applied to the corresponding semantically analogous part of the content-image. Style transformation is also operated on characters [28] for producing unique fonts and unusual text-effects. By collectively training the conditional-generative model and an ornamentation-network model, the style transformation of characters can be accomplished. Another expansion is to generate distinct sound effects by allotting the desired style from the target-audios [29]. It uses two approaches based on iterative optimization of audio and optimization of the employed model. Photorealistic transformation [30] refers to

Table 1: Analysis and comparison of optimization algorithms

Type	Description	Merits	Drawbacks
Gradient Descent-optimizer	It is a traditional algorithm for the optimization of neural network models. It updates weights and tunes the parameters of the model iteratively to minimize the value of loss function.	Updates weights in neural network models iteratively.	It is very slow and takes ample computation time for huge datasets. It is difficult to regulate the learning rate as the algorithm sometimes stuck in local minima.
AdaGrad-optimizer	It applies a separate learning-rate for each parameter. It modifies the learning rate of the parameter at each step based on the computation of the previous gradients of that parameter.	Very helpful for handling scattered data and it is not required to set the learning rate manually.	Its main limitation is the decrease and decay in its learning rate value.
AdaDelta-optimizer	It is an improvement of AdaGrad-optimizer. It entirely replaces the learning-rate parameter by the exponential moving-average of squared-deltas.	No need for the learning-rate parameter.	There is a necessity of two-state variables for saving the second-moments of gradients and the variation in parameters.
Adam-optimizer	It is an adaptive-method for defining separate learning-rates for every parameter using momentum and therefore termed as adaptive-moment-estimation.	It separately defines the learning-rate for every parameter.	Its gives better result in the starting of training, but performance diminishes as the time increases. Therefore, for some tasks, it does not converge to an optimum value.
L-BFGS-optimizer	It is a limited-memory-BFGS algorithm in the family of QUASI-Newton-methods because it takes a small amount of memory. It uses a line-search approach, which makes it much steady in training and more accessible to examine for convergence.	It is the most efficient, speedy, and popular approach of optimization compared to previous methods.	It does not scale well while updating and calculating gradients. Hence, there is a necessity of mini-batch-training that needs to estimate gradients on small subsets of data.

Several approaches of NST are expanded to style transformation of audio, video, and characters, etc. For altering the style of videos, the requirement is to achieve a smooth variation among contiguous frames of video. A method [25] based on iterative optimization of images uses temporal-consistency loss obtained by calculating optical-flow. It produces stylized videos by excluding temporal-artifacts but requires significant time for computing results. Another

the transfer of style of arrangements of colors. Its primary intent is to conserve the fundamental structure of the content-image by utilizing a two-stage optimization method, namely stylization, and smoothing. Hence, for decreasing the deformation of content-image, it employs photorealism-regularization. Another expansion of NST approaches is in the field of fashion [31] to create clothes containing the desired styles of fashion. In the synthesized output, the applied GAN-

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:25:38 UTC from IEEE Xplore. Restrictions apply.

model should protect the basic design of cloth while blending with the given target style.

V. IMPLEMENTATION RESULTS AND ANALYSIS

The characteristics of various distinct procedures for neural style fusion and transformation are thoroughly analyzed [32]. The charts given below in Fig.5 and Fig.6 represent the comparison of running time performance on input-images of distinct dimensions. It has been observed that the algorithm based on optimization of image proposed by Gatys et al. [1] takes a considerable value of computation time when compared with offline methods based on optimization of employed models [3] [4] [5]. However, the quality of generated results is better when compared to others. For real-time applications of arbitrary style transformation, there is a trade-off between the quality of results and the running time of existing algorithms. Therefore, a balance should be maintained between quality and speed.

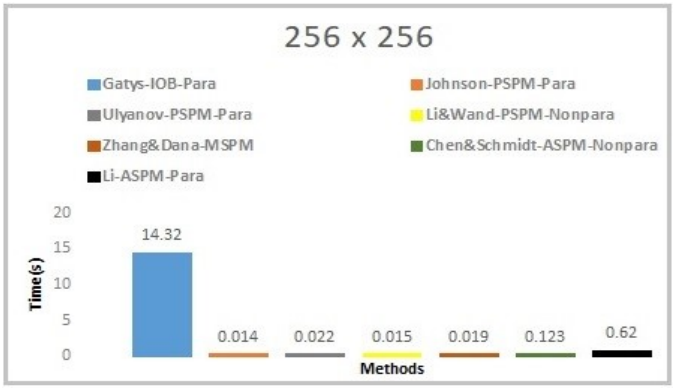


Fig.5: Comparison of the run-time of some NST approaches on 256×256 dimensional images

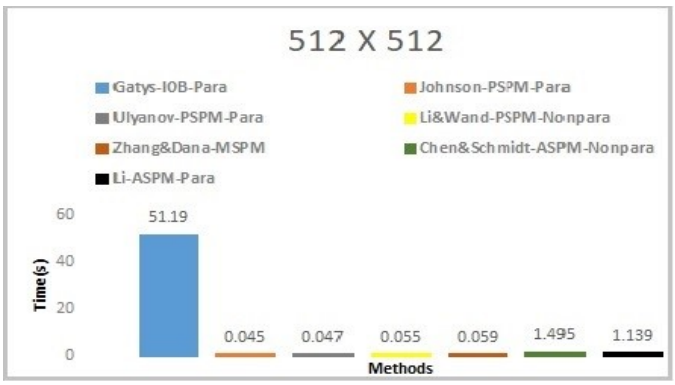


Fig.6: Comparison of the run-time of some NST approaches on 512×512 dimensional images

A network of the SANET model is deployed using Anaconda, PyTorch 0.4.1, CUDA 9.2, and open CV 3.4.2 python. During the training of the model for content-images VOC 2012 dataset is used, and for the style-images WikiArt dataset is used. A PC comprising Windows 10, Intel Core i5 eighth-generation quad-core CPU, NVIDIA Geforce GTX 1050 GPU, and a RAM of size 8 gigabytes is used. It also demands an internet of extremely high-speed. Programming Language used is python. The input is taken as two pre-processed images and resized them to 512×512. During training, the size of each batch is six, and 180000 default iterations are applied. The initial values assigned to weights while computing the loss are as follows, w_c=1.0, w_s=4.0, w_{id1}=50, and w_{id2}=1. It has been perceived that, if the weight w_c is increased in the absence of identity-loss function, then spatial-distribution and structure of content-image can be protected but with a lack of style-features. However, if the weights for identity-loss function are increased, then it's possible to preserve the semantic-arrangement of content along with sustaining the style-features. Consequently, the total loss is estimated by incorporating identity-loss along with content and style losses. Adam optimizer is applied to minimize the value of loss function. The encoder uses a previously trained VGG nineteen network. Two SANET models are jointly trained and integrated by taking the VGG feature-maps as inputs and encodes output feature-maps using distinct layers Relu_4_1 and Relu_5_1. Outputs of these SANETs are combined by using the symmetric decoder. After that, the trained network of SANET models is applied to the pictures taken by us on our college campus to examine the execution time and quality of the obtained results.

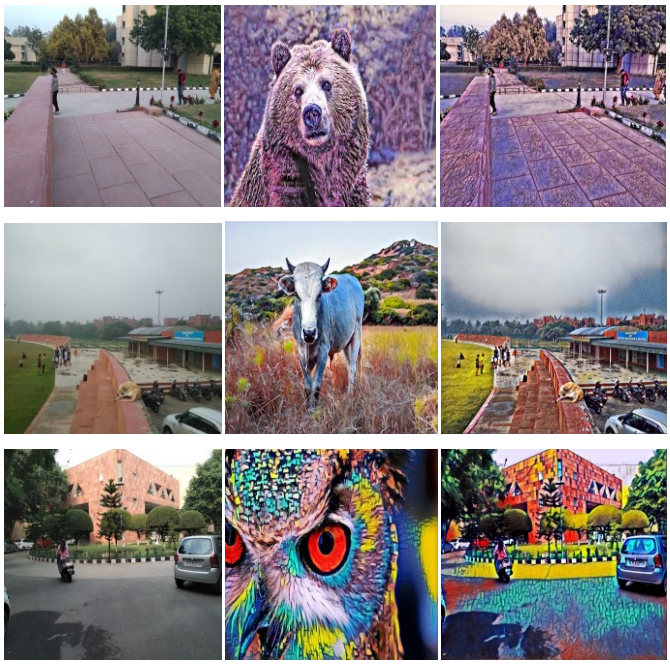


Fig.7: Experimental results of style transformation using SANET model

For determining the quality of the SANET network, its comparison is analyzed with four types of models for fusing and transforming arbitrary-styles. It is observed that AdaIN [8] is the fastest ASPM approach, but it renders a suboptimal result as it copies the channel wise mean-variance value, and the trained decoder usually attach frequently seen similar patterns and style-textures to all stylized outputs. Hence, its resulting output is less appealing and sometimes fails to retain the distribution of colors. Despite matching the second-order-statistics optimally in WCT [7], sometimes, it creates unseen and collapsed-patterns. Avatar-Net [9] blends the images based on the semantic-arrangement of content and employs multi-scale stylization. Still, because of its dependence on the patch size, it fails to realize the local and global style-patterns

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:25:38 UTC from IEEE Xplore. Restrictions apply.

simultaneously. Seldom Avatar-Net and WCT render the blurred hair-texture and distorted vision of content due to the collapsed brush strokes. In contradiction, the network containing two integrated SANET models can generate various exquisite styles efficiently.

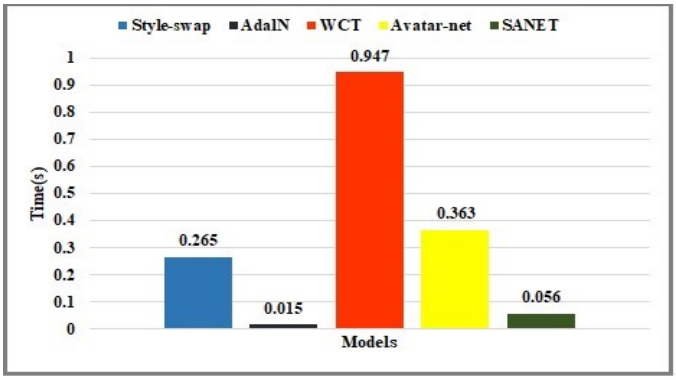


Fig.8: Comparison of execution time of SANET with other ASPM models

The graph in Fig.8 displays a comparison of the running time of SANET with other existing methods of arbitrary style transformation for images of 512×512 dimensions. SANET is much faster than style-swap, Avatar-net, and WCT models but a bit slower in comparison to AdaIN. However, the quality of the result surpasses AdaIN and others. Therefore, this method is both effective and practically efficient.

VI. CONCLUSION

This paper presents an advanced network containing a combination of two SANET Models for style fusion and transformation. This network ensures the semantic-arrangement and spatial-distribution of content-image while fusing with the characteristics of new arbitrary style-patterns by computing a new identity-loss component. The execution time of different NST algorithms are compared on input-images of distinct dimensions. It is observed that online neural models are slow but deliver better results in comparison to fast offline neural models. In all existing approaches, there is a trade-off between the quality of results and the running time of existing algorithms. Our results demonstrate that this model is much faster than other ASPM models and produces results with better quality and flexibility, which makes it much effective and practically efficient. They are planning to expand this model for the style transformation of videos in the future.

REFERENCES

[1] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2414–2423, 2016.

[2] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," in ArXiv e-prints, Aug. 2015.

[3] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for realtime style transfer and super-resolution," in European Conference on Computer Vision, pp. 694–711, 2016.

[4] D. Ulyanov, V. Lebedev, A. Vedaldi, and V. Lempitsky, "Texture networks: Feed-forward synthesis of textures and stylized images," in International Conference on Machine Learning, pp. 1349–1357, 2016.

[5] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6924–6932, 2017.

[6] T. Q. Chen and M. Schmidt, "Fast patch-based style transfer of arbitrary style," in Proceedings of the NIPS Workshop on Constructive Machine Learning, 2016.

[7] Y. Li, C. Fang, J. Yang, Z. Wang, X. Lu, and M.-H. Yang, "Universal style transfer via feature transforms," in Advances in Neural Information Processing Systems, pp. 385–395, 2017 .

[8] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in Proceedings of the IEEE International Conference on Computer Vision, pp. 1501–1510, 2017.

[9] L. Sheng, Z. Lin, J. Shao, and X. Wang. Avatar-Net: Multiscale zero-shot style transfer by feature decoration. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8242–8250, 2018.

[10] J. E. Kyprianidis, J. Collomosse, T. Wang, and T. Isenberg, "State of the 'art': A taxonomy of artistic stylization techniques for images and video," IEEE transactions on visualization and computer graphics, vol. 19, no. 5, pp. 866–885, 2013.

[11] A. Hertzmann, "Painterly rendering with curved brush strokes of multiple sizes," in Proceedings of the 25th annual conference on Computer graphics and interactive techniques. ACM, pp. 453–460, 1998.

[12] B. Gooch, G. Coombe, and P. Shirley, "Artistic vision: painterly rendering using computer vision techniques," in Proceedings of the 2nd international symposium on Non-photorealistic animation and rendering. ACM pp. 83–ff, 2002.

[13] Y.-Z. Song, P. L. Rosin, P. M. Hall, and J. Collomosse, "Arty shapes," in Proceedings of the Fourth Eurographics conference on Computational Aesthetics in Graphics, Visualization and Imaging, Eurographics Association, pp. 65–72, 2008.

[14] H. Winnemoller, S. C. Olsen, and B. Gooch, "Real-time video abstraction," in ACM Transactions On Graphics (TOG), vol. 25, no. 3. ACM, pp. 1221–1226, 2006.

[15] M. Zhao and S.-C. Zhu, "Portrait painting using active templates," in Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering. ACM, pp. 117–124, 2011.

[16] Y. Li, N. Wang, J. Liu, and X. Hou, "Demystifying neural style transfer," in Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, pp. 2230–2236, 2017.

[17] S. Li, X. Xu, L. Nie, and T.-S. Chua, "Laplacian-steered neural style transfer," in Proceedings of the 2017 ACM on Multimedia Conference. ACM, pp. 1716–1724, 2017.

[18] C. Li and M. Wand, "Combining markov random fields and convolutional neural networks for image synthesis," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2479–2486, 2016.

[19] C. Li and M. Wand, "Precomputed real-time texture synthesis with markovian generative adversarial networks," in European Conference on Computer Vision, pp. 702–716, 2016.

[20] V. Dumoulin, J. Shlens, and M. Kudlur, "A learned representation for artistic style," in International Conference on Learning Representations, 2017.

[21] D. Chen, L. Yuan, J. Liao, N. Yu, and G. Hua, "Stylebank: An explicit representation for neural image style transfer," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1897–1906, 2017.

[22] H. Zhang and K. Dana, "Multi-style generative network for real-time transfer," arXiv preprint arXiv:1703.06953, 2017.

[23] D. Y. Park and K. H. Lee., "Arbitrary style transfer with style-attentional networks," arXiv preprint arXiv:1812.02342, 2018.

[24] S. Ruder, "An overview of gradient descent optimization algorithms" arXiv preprint arXiv:1609:04747, 2017.

[25] M. Ruder, A. Dosovitskiy, and T. Brox, "Artistic style transfer for videos," in German Conference on Pattern Recognition, pp. 26–36, 2016.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:25:38 UTC from IEEE Xplore. Restrictions apply.

[26] D. Chen, J. Liao, L. Yuan, N. Yu, and G. Hua, "Coherent online video style transfer," in Proceedings of the IEEE International Conference on Computer Vision, pp. 1105–1114, 2017.

[27] A. J. Champandard, "Semantic style transfer and turning two-bit doodles into fine artworks," ArXiv e-prints, Mar. 2016.

[28] S. Azadi, M. Fisher, V. Kim, Z. Wang, E. Shechtman, and T. Darrell, "Multi-content gan for few-shot font style transfer," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[29] P. Verma and J. O. Smith, "Neural style transfer for audio spectograms," in Proceedings of the NIPS Workshop on Machine Learning for Creativity and Design, 2017.

[30] Y. Li, M.-Y. Liu, X. Li, M.-H. Yang, and J. Kautz, "A closed-form solution to photorealistic image stylization," in European Conference on Computer Vision, 2018.

[31] S. Jiang and Y. Fu, "Fashion style generator," in Proceedings of the 26th International Joint Conference on Artificial Intelligence. AAAI Press, pp. 3721–3727, 2017.

[32] Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu, M. Song. "Neural Style Transfer: A Review", IEEE Transactions on Visualization and Computer Graphics,2019.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:25:38 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Dementia Detection Using Machine Learning by Stacking Models

Ram Murti Rawat
Assistant Professor, Department of Computer Science
Delhi Technological University, Delhi, India
rammurtirawat@dtu.ac.in

Mohammad Akram
Department of Computer Science
Delhi Technological University, Delhi, India
mohammadakram077@gmail.com

Mithil
Department of Computer Science
Delhi Technological University, Delhi, India
mithilmk7@gmail.com

Setti Suneel Pradeep
Department of Computer Science
Delhi Technological University, Delhi, India
ernestsolomoninzion@gmail.com

***Abstract*— Dementia is a syndrome which weakens the memory, thinking, behavior, and the ability to perform normal day-to-day activities. Although dementia mainly affects older people, it is not a part of aging. Machine Learning (ML) is considered as one of the contemporary approaches in predicting, identifying, and making decisions without having human involvement. ML is quickly evolving in the medical industry ranging from diagnosis to visualization of diseases and the study of disease transmission. The present study aims to explore the usage of the stacking of two models, namely gradient boosting machine and artificial neural network is used in the prediction of dementia and validate its performance through statistical analysis.**

Keywords*— *Machine learning, OASIS, Artificial Neural Network, Gradient Boosting Machine, Stacking, Ensemble Learning

I. INTRODUCTION

Machine learning (ML) is considered as an integral part of Artificial Intelligence (AI), also a data analysis technique that computerizes the explanatory model structure. In most scenarios, based on the learning method, two types of ML algorithms (supervised & unsupervised) were used [1]. At present, these algorithms are engaging in all the major industries like healthcare, banking, transport, social media, etc. [1] [2]. Above all, the medical industry is advancing quickly with high volumes of information and increasing difficulties in inventory and patient outcomes. Economically developed nations such as the USA, Japan, and European countries are even facing problems with the enormous collection of medical data [3]. However, by using conventional techniques, it is not possible to analyze this significant volume of information because of time consumption and efforts. Therefore, ML techniques are coming up with various algorithms and programs to avoid these issues. Besides that, the selection of a proper algorithm is not an easy task since it depends on multiple factors such as data volume, information type, and outcomes related to industry requirements [1][4].

Nowadays, Machine Learning algorithms are progressively utilized in neuro-imaging studies like a prediction of Alzheimer's disease (AD) from auxiliary MRI. Also, many studies attempted different ML strategies in predicting AD and their causes [3] [5]. In the study of AD prediction and retrieval, a multistage classifier utilizing ML, including Naive Bayes classifier, support vector machine (SVM), and K-nearest neighbor (KNN) is used to group Alzheimer's illness in the more acceptable and effective way [6]. Similarly, a study from Ref. [7], concluding that the utilization of locally linear embedding (LLE) kind of unsupervised learning is utilized to categorize AD based on fundamental MRI data. Besides, some preliminary studies with ML techniques concluded that these methods are valid and accomplish with high precision (up to 98%) in diagnosing clinical events with analysis of patient medical records [8][4].

Despite it, AD is one of the standard types of dementia and associated mostly with older people [9]. In this paper, a brief explanation is given on the prediction of dementia and calculate performance by using a Stack of two models, namely Gradient Boosting Machine and Artificial Neural Network, i.e., Ensemble Learning. Ensemble learning is a ground-breaking ML method that has shown significant advantages in numerous applications. An ensemble with regards to ML can be extensively characterized as an ML framework that is built with a lot of individual base learners as ensemble members working in parallel and whose predicted results are joined to deliver a single output for a given issue. This single output should have better performance on average than any other ensemble member with uncorrelated error on the target datasets [10].

II. LITERATURE REVIEW

The use of AI in the domain of healthcare has seen vast popularity in recent years, albeit it has a long history. In the 1970s, an expert system was designed for aiding progress in healthcare. Continuous research and development are features of this domain. Many researchers have attempted to address

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:52:38 UTC from IEEE Xplore. Restrictions apply.

this topic and a brief overview of which is presented in the forthcoming section [11].

A short survey of the related work is included in this section. Tohka et al. [12] used Alzheimer's disease Neuroimaging Initiative dataset for comparison of various feature selection techniques by using machine learning for dementia for anatomical brain MRI. SVM and Logistic Regression are used as classifiers, and then it is applied post feature selection techniques. It shows that removing age feature improves the average accuracy with all the classifiers, although the author isn't clear about the discoveries of this investigation for other brain diseases [13].

Zhang Y et al. [14] have proposed a computer-aided diagnosis (CAD) system for MRI images of brain-based on Eigenbrains. This approach generates eigenbrain images based on EEG using key slices from the 3D volumetric data from the MRI. Subsequently, Kernel SVM is included and trained by particle swarm optimization algorithm. The proposed approach was effective in AD subject prediction and discriminant brain-region detection in MRI scanning; still, some advantages and disadvantages of eigenbrain also exist, such as it achieves very high classification accuracy, which can be improved further by combining other features. However, it is essentially two-dimensional, which does not reduce the redundancy along the slice direction; also, it needs preprocessing of spatial registration, which costs a large amount of computation power.

Moradi E et al. [15] has proposed a technique for identifying mild cognitive impairment (MCI) as a transitional stage between age-related cognitive decline and AD. A novel MRI – based biomarker that was developed using machine learning strategies. Noteworthy techniques used in this model included using semi-supervised learning on data available from AD patients and performing feature selection using regularized logistic regression, removing aging effects from MRI data before classifier training to prevent possible confounding between changes due to AD and those due to standard aging, and construction of an aggregate biomarker by learning separate MRI biomarker and subsequently combining age and cognitive measures about MCI subjects by applying random forest classifier. Their aggregate biomarker achieved a very high predictive performance. However, it needs preprocessing of each MRI image, which requires 8 min per image on a high-performance system, thus costs a large amount of computation power.

III. EXPLORATORY DATA ANALYSIS

Analysis of the relationship between the various features of OASIS data explicitly through the graph is a crucial step to understand the nature of the data and eventually to develop a machine learning model. The following figures are implemented to achieve the same.

The graph illustrated in Fig.1 suggests that gender and dementia are correlated and that men are more likely to be affected by dementia than women. It also indicates that women having dementia are diagnosed late, meaning that they show symptoms a little late as compared to men.

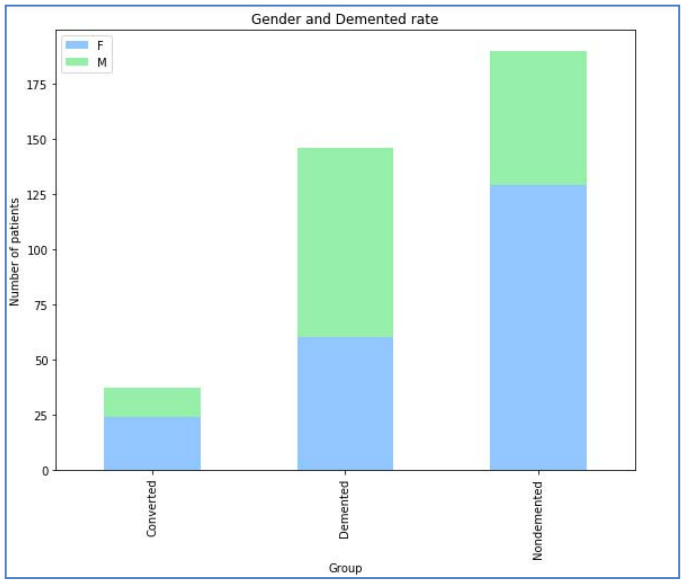


Fig. 1. Gender-wise distribution of dementia

The graph illustrated in Fig. 2 suggests that the Mini-Mental State Examination (MMSE) score of people who have dementia is lowest. In contrast, the people nondemented have high MMSE scores, and the people classified as converted have MMSE scores between those two classes.

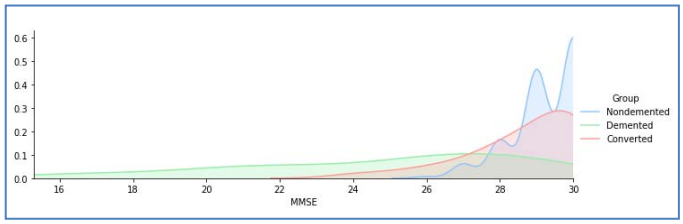


Fig. 2. Distribution of MMSE for nondemented, demented and converted Group

The graph illustrated in Fig. 3 suggests that Atlas Scaling Factor (ASF) value is higher for the converted Group than the demented Group and nondemented Group. This results in the addition of ASF in the feature set.

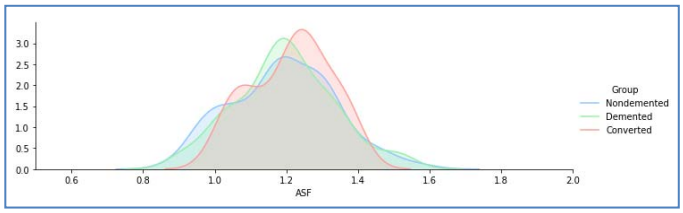


Fig. 3. Distribution of ASF for nondemented, demented and converted Group

The graph in Fig.4 proposes that the estimated Total Intracranial Volume (e-TIV) value is higher for the converted Group than the demented Group and nondemented Group. This results in the presence of e-TIV in the feature set.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:52:38 UTC from IEEE Xplore. Restrictions apply.

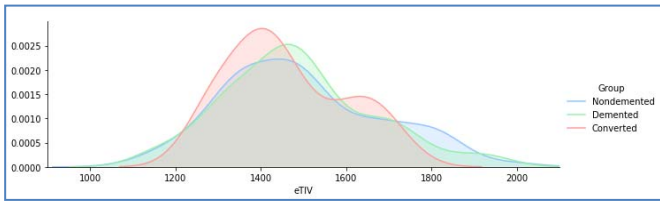


Fig. 4. Distribution of e-TIV for nondemented, demented and converted Group

The graph illustrated in Fig.5 suggests that the nondemented Group has a higher brain volume ratio than the demented and converted Group. This is assumed to be because of the dementia disease that affects the brain resulting in shrinking its tissue.

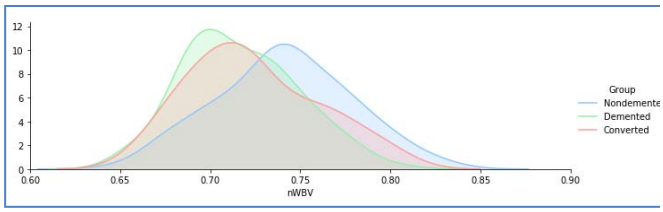


Fig. 5. Distribution of n-WBV for nondemented, demented and converted Group

The following table shows the min, max, mean, and median of the OASIS Dataset features.

TABLE I. OASIS Dataset min, max, mean, median

Feature	Min	Max	Mean	Median
MR Delay	0.0	2639.0	595.11	552.0
EDUC	6.0	23.0	14.6	15.0
SES	1.0	5.0	2.41	2.0
MMSE	4.0	30.0	27.34	29.0
CDR	0.0	2.0	0.29	0.0
eTIV	1106.0	2004.0	1488.13	1470.0
nWBV	0.64	0.84	0.73	0.73
ASF	0.88	1.587	1.19	1.19

IV. DATA COLLECTION AND PROPOSED METHODOLOGY

A. Dataset

A longitudinal collection of OASIS - MRI data set, comprising of demented and nondemented subjects with right-hand (R) type aging from 60 to 96 is considered. A sample size of 150 subjects, including men and women, have attended scanning sessions more than two visits; sessions were separated by at least one year with 373MR Sessions [16]. The sample training data included with demographic values of Subject ID, MRI ID, Group, Visit, MR delay, Sex, Age, Social Economic Status (SES), Education level (EDUC), Mini- Mental State Examination (MMSE), Clinical Dementia Ratio (CDR), estimated Total Intracranial Volume (e-TIV), normalized Whole Brain Volume (n-WBV), and Atlas Scaling Factor (ASF). In particular, some subjects treated as demented at initial visit later transformed into the nondemented managed by converted type. If CDR value is equal to zero, the subjects were considered as mostly nondemented, simultaneously if CDR ≥ 1, the subjects will face the tendency to have dementia [9].

1. *Data collection*

The trained data set collected from the Open Access Series of Imaging Studies (OASIS) included longitudinal MRI data of 150 subjects.

2. *Data Preprocessing*

Real-world data is available more likely incomplete with missing entries. Therefore, data preprocessing is one of the data mining techniques to address this issue. Missing entries were filled-up by the median of particular attribute values.

3. *Attribute Selection*

Select a specific characteristics to predict the outcome to do mapping with input correspondence values. Group column is chosen as an output variable that corresponds to the dementia status based on other input variables.

4. *Input variable matching*

The performance of any ML model largely depends on the number of input attributes taken into consideration. To maintain better performance, the selection of the corresponding attributes, instead of selecting multiple ones is important. Features such as SUBJECT_ID, MRI_ID, Visit, and Hand were removed as they were not affecting the groups in any way. Hand Feature was removed because the dataset consists of MRI of Right-handed people only.

5. *Classifier*

Three groups of dementia are considered they are, demented, nondemented, and converted.

B. Methodology

In this paper, stacking is used, which is an ensemble learning technique that involves combining multiple machine learning models to improve the machine learning results [10]. Similarly, here two machine learning models are used for the classification of Demented, Nondemented, and Converted groups, and the result of one model is added as an additional feature to the existing dataset. This new dataset is fed to the other model to train.

The two models that are used here are Gradient Boosting Machine and Artificial Neural Network. The reason for selecting Gradient Boosting Machine is because of its fast processing and iterative nature to improve from a weak classifier to a robust classifier. It is an ensemble meta-algorithm that primarily reduces bias and also variance, which is an essential criterion to analyze our models and prevent them from over-fitting [13].

Artificial neural networks are found to be effective as general nonlinear models with applications to medical diagnosis, prognosis, and survival analysis. Dementia syndrome is very complicated. The relationship between various risk factors is not dependent; thus, most of the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:52:38 UTC from IEEE Xplore. Restrictions apply.

traditional statistical models are inapplicable. The second model ANN is chosen because of its nonlinear adaptive processing system made of a large number of artificial neurons in the right way to connect each other. The ANN can solve the problems of unambiguous and uncertain medical information more effectively and since it can monitor complex relations; therefore, ANN is selected as a second model here [4].

Our problem is a multi-class classification problem as it contains three classes Demented, Nondemented and Converted thus there are two heuristic methods for multi-class classification, they are:

1. *One-vs-Rest*
2. *One-vs-One*

Here the One-vs-Rest heuristic method is used on Gradient Boosting Machine.

Fig. 6 illustrates the idea of using three different Gradient Boosting Machine for each separate Group – nondemented, demented, and converted.

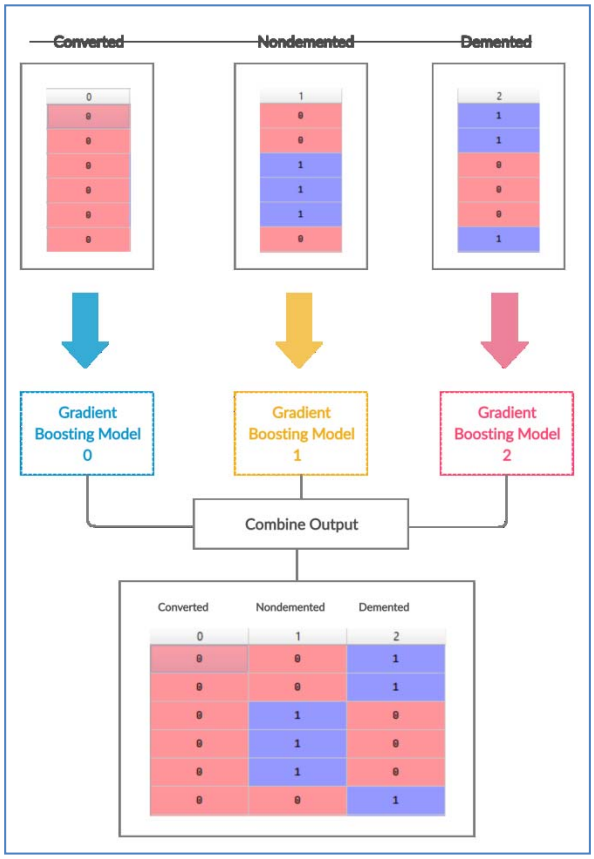


Fig. 6. Training 3 Gradient Boosting Machine

After training on the test data, these three Gradient Boosting Models predict the output for the whole dataset, and their predicted output is combined and added to the original dataset as another feature, which is later fed to the ANN Model for training and testing as illustrated in Fig.7.

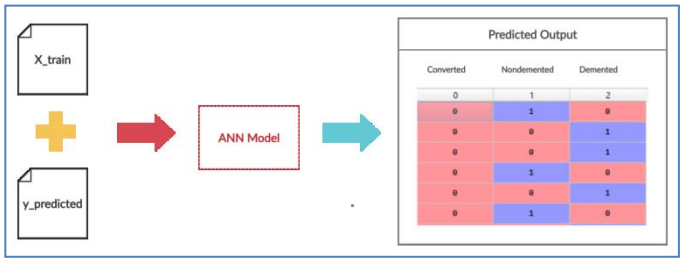


Fig. 7. Predicted Output + Original Dataset fed to ANN Model

V. RESULT

The same data is tested on the Gradient Boosting Model and ANN Model separately, and then the combined version of these two models. The following results are observed:

1. *Gradient Boosting Machine*

TABLE II. Confusion Matrix of given subjects Gradient Boosting Machine TND*: True Nondemented; TD*: True Demented; TC*: True Converted; PND*: Predict Nondemented; PD*: Predict Demented, and PC*: Predict Converted

	TC	TD	TND
PC	3	2	2
PD	1	29	0
PND	4	1	33

```
==> Classification Report -- Gradient Boosting

              precision    recall  f1-score   support

           0       0.38      0.43      0.40         7
           1       0.91      0.97      0.94        30
           2       0.94      0.87      0.90        38

    accuracy                           0.87        75
   macro avg       0.74      0.75      0.75        75
weighted avg       0.88      0.87      0.87        75
```

Fig. 8. Classification Report of Gradient Boosting Model – [0, 1, 2] stands for [Converted, Demented, Nondemented]

2. *ANN Model*

TABLE III. Confusion Matrix of given subjects ANN Model TND*: True Nondemented; TD*: True Demented; TC*: True Converted; PND*: Predict Nondemented; PD*: Predict Demented, and PC*: Predict Converted

	TC	TD	TND
PC	0	5	2
PD	0	30	0
PND	0	1	37

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:52:38 UTC from IEEE Xplore. Restrictions apply.

```
==> Classification Report -- ANN

              precision    recall  f1-score   support

           0       0.00      0.00      0.00         7
           1       0.83      1.00      0.91        30
           2       0.95      0.97      0.96        38

    accuracy                           0.89        75
   macro avg       0.59      0.66      0.62        75
weighted avg       0.81      0.89      0.85        75
```

Fig. 9. Classification Report of ANN Model – [0, 1, 2] stands for [Converted, Demented, Nondemented]

3. *Stack Model - Gradient Boosting Model + ANN Model*

TABLE IV. Confusion Matrix of given subjects Gradient Boosting Machine + ANN TND*: True Nondemented; TD*: True Demented; TC*: True Converted; PND*: Predict Nondemented; PD*: Predict Demented, and PC*: Predict Converted

	TC	**TD**	**TND**
PC	3	2	2
PD	1	29	0
PND	4	1	35

```
==> Classification Report -- GB + ANN

              precision    recall  f1-score   support

           0       0.50      0.43      0.46         7
           1       0.91      0.97      0.94        30
           2       0.95      0.92      0.93        38

    accuracy                           0.89        75
   macro avg       0.78      0.77      0.78        75
weighted avg       0.89      0.89      0.89        75
```

Fig. 10. Classification Report of Stack Model – [0, 1, 2] stands for [Converted, Demented, Nondemented]

From the above results, it is clearly observed that the ANN produced excellent results as it detects the highest number of demented and nondemented groups correctly. Still, the problem with that is its precision and recall for detecting converted Group is zero, as seen in the Classification Report of ANN Model in Fig. 9 and in the confusion matrix in TABLE III; hence this model is not appropriate to choose alone for dementia analysis. The solo Gradient Boosting Model also provides excellent results giving weighted average accuracy of 87% as seen in the Classification Report in Fig. 8 and confusion matrix in TABLE II, but it is observed that the stack model provides the highest weighted average accuracy of 89%, the best precision and recall for all the groups, as seen in its Classification Report in Fig. 10, and the confusion matrix in TABLE IV.

VI. CONCLUSION

This paper introduces a method of detecting dementia from auxiliary MRI reports and a few other features. Gradient Boosting Machine and ANN are exploited as well as their stacking with each other to achieve these tasks. The Neural Network is trained with the new dataset, including the predicted output of the Gradient Boosting Machine, produced the best results as compared to the results produced by both the models' individually.

Dementia is one of the significant health issues that has challenged health experts worldwide. Also, it mostly happened in older people with age > 60. For more than two decades, healthcare professionals are working continuously to find a method to forecast it in the earlier stages and to find a medicine that can cure. Unfortunately, yet there are no proper medicines to cure this disease completely. Many computer scientists are working along with medicinal institutions performing research activities on this disease. Still, there is an extreme need for the identification of relevant characteristics that can forecast the detection of dementia.

It tis then approached with the stacking of models for classification and prediction of dementia and achieved optimized results with efficient performance values. The future work involves the identification of a new set of attributes that can help in detecting early cognitive impairment and dementia.

REFERENCES

[1] Huddleston SH, Brown GG., "Machine learning. Informs analytics body of knowledge", 2018.

[2] Libbrecht MW., Noble WS., "Machine learning applications in genetics and genomics.", Nat Rev Genet 2015.

[3] Ichikawa D., Saito T., Ujita W., Oyama H., "How can machine-learning methods assist in virtual screening for hyperuricemia? A healthcare machine-learning approach.", J. Biomed Inform, 2016.

[4] Wang, Naibo & Chen, Jinghua & Xiao, Hui & Wu, Lei & Jiang, Han & Zhou, Yueping. "Application of artificial neural network model in diagnosis of Alzheimer's disease.", BMC Neurology, 2019.

[5] Kaur P., Sharma M., Mittal M., "Big data and machine learning based secure healthcare framework.", Procedia Comput Sci, 2018.

[6] Kruthika KR., Rajeswari, Maheshappa HD., "Multistage classifier-based approach for Alzheimer's disease prediction and retrieval.", Inf Med Unlocked 2019, Nov. 2018.

[7] Liu X., Tosun D., Weiner MW., Schuff N., "Locally linear embedding (LLE) for MRI based Alzheimer's disease classification.", Neuroimage 2013.

[8] XIE R., Khalil I., Badsha S., Atiquzzaman M., "Collaborative extreme learning machine with a confidence interval for P2P learning in healthcare.", Comput Network, 2019.

[9] Battineni, Gopi & Chintalapudi, Nalini & Amenta, Francesco, "Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM).", Informatics in Medicine Unlocked, 2019.

[10] Yun Yang, "Ensemble Learning" in *Temporal Data Mining Via Unsupervised Ensemble Learning*, Elsevier, 2017, pp. 35-56.[online]. Available: http://www.sciencedirect.com/science/article/pii/B978012811654800004X

[11] Yuvasri Raghavan, Shubham Agarwal, "Comparative Study of Machine Learning Models for Alzheimer's Detection.", IRJET, Oct. 2019.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:52:38 UTC from IEEE Xplore. Restrictions apply.

[12] Tohka, J., Moradi, E., Huttunen, H., & Alzheimer's Disease Neuroimaging Initiative, "Comparison of feature selection techniques in machine learning for anatomical brain mri in dementia.", Neuroinformatics, 2016.

[13] Deepika Bansal, Rita Chhikara, Kavita Khanna, Poonam Gupta. "Comparative Analysis of Various Machine Learning Algorithms for Detecting Dementia.", Procedia Computer Science, 2018.

[14] Yudong Zhang, Zhengchao Dong, Preetha Phillips, Shuihua Wang, Genlin Ji, Jiquan Yang, and Ti-Fei Yuan, "Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning," frontiers in Computational Neuroscience, vol 9, June 2015.

[15] Moradi E., Pepe A., Gaser C., Huttunen H., Tohka J., "Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects," NeuroImage, vol. 104, pp. 398-412, Jan. 2015.

[16] Jacob Boysen, "MRI and Alzheimers," August 16, 2017.[online]. Available: https://www.kaggle.com/jboysen/mri-and-alzheimers

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:52:38 UTC from IEEE Xplore. Restrictions apply.

JJCO Japanese Journal of Clinical Oncology

Japanese Journal of Clinical Oncology, 2020, 00(00)1–11
doi: 10.1093/jjco/hyaa130
Original Article

OXFORD

Original Article

Descriptive epidemiology of incidence and mortality of primary liver cancer in 185 countries: evidence from GLOBOCAN 2018

Rajesh Sharma *

University School of Management and Entrepreneurship, Delhi Technological University, Delhi, India

*For reprints and all correspondence: Rajesh Sharma, University School of Management and Entrepreneurship, Delhi Technological University, Vivek Vihar- Phase II, Delhi-110095, India, E-mail: rajesh.sharma@dtu.ac.in

Received 1 June 2020; Editorial Decision 2 July 2020; Accepted 3 July 2020

Abstract

Purpose: This study aims to examine the burden of primary liver cancer in 185 countries in 2018.
Methods: The estimates of incidence, mortality and prevalence of primary liver cancer were procured from GLOBOCAN 2018. The development status of a country was measured using the human development index—a composite indicator of income per capita, education and life expectancy.
Results: Globally, primary liver cancer resulted in an estimated 781 631 deaths at age-standardized mortality rate of 8.5/100 000, and 841 080 cases were estimated to be diagnosed in 2018. Males accounted for 596 574 cases and 548 375 deaths, which is more than twice the burden of primary liver cancer in females (cases: 244 506; deaths: 233 456). The global age-standardized incidence rate was 9.3/100 000 in 2018, varying from Morocco (1.1/100 000) to Mongolia (93.7/100 000). There were remarkable variations in terms of age-standardized mortality rate, too, which ranged from 1/100 000 in Nepal to 75.4/100 000 in Mongolia. East Asia was the top region contributing 55.6% of global cases and 54.7% of global deaths.
Conclusions: Since majority of the primary liver cancer burden pertains to hepatocellular carcinoma and screening approaches are yet to be fully proven, the policy focus must be on prevention approaches through the hepatitis-B vaccine, early detection of hepatitis-C infection, reduced alcohol consumption, obesity control, reduced aflatoxin exposure and containment of other modifiable risk factors.

Key words: liver cancer, incidence, mortality, mortality-to-incidence ratio, GLOBOCAN

Introduction

Primary liver cancer (PLC) is the fourth leading cause of cancer death among all cancer groups worldwide [1], and its incidence and mortality have increased in the last three decades [2]. PLC has one of the worst 5-year survival rates among 18 cancer groups examined in Concord-3 [3]. Histologically, hepatocellular carcinoma (HCC) constitutes 70–85% of PLC, while 10–25% is intrahepatic cholangiocarcinoma (ICC) [4,5]. Hepatitis-B virus (HBV) and hepatitis-C virus (HCV) contributes around 80% of HCC across the globe [6]. Aflatoxin exposure is another significant risk factor of HCC in Africa and East Asia [7]. The main source of Aflatoxin exposure is fungi, such as Aspergillus Flavas and Aspergillus Parasiticus [8]. Apart from HBV and HCV infections, other risk factors of HCC are alcohol consumption, obesity and non-alcoholic fatty liver disease (NAFLD) [9].

Liver cancer not only exerts substantial burden in adults, it also exerts significant burden in children and adolescents; around 67–80% of pediatric PLC is hepatoblastoma and the rest 20–33% is HCC [10]. The incidence of HCC is greater than hepatoblastoma in East Asia, while the converse is true in the Western countries [11,12]. In hyperendemic countries such as China, HCC in younger children

© The Author(s) 2020. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permission@oup.com.

Downloaded from https://academic.oup.com/jjco/article-abstract/doi/10.1093/jjco/hyaa130/5876941 by Delhi Technological University user on 03 August 2020

is mostly acquired through perinatal exposures [13–15]; in contrast, in low-endemic areas, the common causes of HCC are tyrosinemia and other metabolic disorders [10].

For cancer control policy formulation and implementation at the global, regional and national levels, it is pertinent to understand the epidemiology of PLC. This study envisages to systematically examine the epidemiology of PLC in 185 countries drawing upon estimates from GLOBOCAN 2018 [1,16,17]. This study provides a comprehensive and comparable assessment of PLC burden measured in terms of all-age and age-standardized incidence, mortality and prevalence at the global, regional and national levels. Following previous literature [18–21], the mortality-to-incidence ratio (MIR) is employed as a proxy for 5-year survival rates. I also studied the PLC epidemiology, both age- and sex-wise, in the light of the socio-economic status of countries measured using the human development index (HDI, [22]). While previous studies on PLC epidemiology either focussed on select countries or regions, this is the first study providing a comprehensive assessment of PLC burden in 185 countries for one of the most recent year (2018).

Materials and methods

GLOBOCAN 2018 estimates

The estimates of incidence and mortality of PLC for 185 countries were drawn from GLOBOCAN 2018 [1,16,17]. The estimates of PLC pertained to code C22 (liver and intrahepatic bile ducts cancer) of the international classification of disease (ICD) tenth revision (ICD-10). The GLOBOCAN estimates were available in terms of age-wise and sex-wise incidence, mortality and prevalence (1-year, 3-year and 5-year) for the year 2018. All-age and age-specific incidence and death counts are reported along with 95% uncertainty intervals (UIs) inside square brackets; these UIs reflect uncertainties associated with availability and quality of data from the cancer registry, death registration and timeliness of data reporting [1,16,17].

The age-wise estimates were examined for the 16 age groups: below-15, 15 to 19, 20 to 24, 25 to 29, 30 to 34, 35 to 39, 40 to 44, 45 to 49, 50 to 54, 55 to 59, 60 to 64, 65 to 69, 70 to 74, 75 to 79, 80 to 84 and 85-plus. The PLC estimates were provided by GLOBOCAN for 185 countries in 21 world regions as defined by the United Nations Population Division (UNPD). In addition to incidence, mortality and 5-year prevalence, I also examined the global, regional and national MIR, which is employed as a proxy for 5-year survival rates. The MIRs were calculated as the ratio of all-age death counts and all-age incidence counts of the mean estimates.

Human development index

The comparative performance of countries in terms of various indicators of PLC burden was evaluated vis-à-vis a country's socio-economic status measured by its HDI [22,23]. HDI, an indicator of socio-economic status, is proposed and developed by the United Nations Development Program (UNDP). HDI has been used as an indicator of socio-economic status of countries in previous epidemiological studies too [20,21]; the advantage with this indicator is that it is a composite indicator of three indicators of development, namely, gross national income (GNI) per capita, education (composite construct of mean years of schooling and expected years of schooling) and longevity (life expectancy at birth). The UNDP rescales each of these indicators on a 0 (lowest) to 1 (highest) scale, and the geometric mean of three indicators provides the final value of the HDI as per the following formula [22].

$$\mathrm{HDI} = \sqrt[3]{\mathrm{INCOME} * \mathrm{EDUCATION} * \mathrm{LE}}$$

INCOME denotes income index; EDUCATION denotes education index; LE denotes life expectancy index. HDI values were drawn only for the year 2018, which is consistent with GLOBOCAN estimates. Based on HDI values, countries were categorized by UNDP into four groupings: low HDI (HDI < 0.550), medium HDI (0.550–0.669), high HDI (0.700–0.799) and very high HDI (>0.800).

Results

Global and regional burden

Globally, 841 080 [95% UI, 817 635–865 198] cases of PLC were estimated to be diagnosed at an age-standardized rate of 9.3/100 000 in 2018 (Table 1). The PLC resulted in an estimated 781 631 [737 605–828 285] deaths at an age-standardized rate of 8.5/100 000. Globally, the 5-year prevalence of PLC was 675 210 in 2018. East Asia was the leading region in terms of both incidence and death counts, with 467 327 [459 452–475 337] cases and 427 932 [421 450–434 514] deaths, respectively, in 2018 (Table 1). South-east Asia was the second leading region with 89 010 [83 523–94 858] cases and 88 429 [82 662–94 599] deaths in 2018. The 5-year prevalence was again the highest in East Asia (383 221), followed by South-east Asia (69 441) and South-central Asia (35 246). East Asia also had the highest age-standardized incidence rate (ASIR) of 17.7/100 000, with the lowest ASIR estimated in South-central Asia (2.5/100 000). The age-standardized mortality rate (ASMR) was again the highest in East Asia at 16/100 000, and the lowest ASMR was recorded by South-central Asia at 2.3/100 000 (Table 1).

Country-wise burden

Country-wise, China, USA and Japan were the three leading countries in terms of new cases with 392 868 [386 346–399 500], 37 948 [37 032–38 887] and 35 535 [34 731–36 358] cases, respectively, in 2018 (Supplementary Table S1). In terms of death counts too, China was the leading country with 273 014 [268 132–277 985] deaths, followed by the USA with an estimated 20 564 [19 973–21 172] deaths in 2018. Vietnam was ranked third in terms of death counts with 19 708 [17 974–21 609] deaths, which is higher than Japan (18 539 [17 987–19 108]) even though Vietnam had lesser number of new cases (25 335 [23 628–27 166]) than Japan (35 535 [34 731–36 358]). In terms of 5-year prevalence, China (296 780), Japan (43 875), USA (31 023) and South Korea (25 015) were the leading countries (Supplementary Table S1).

The top three countries in terms of ASIR were Mongolia (93.7), Egypt (32.2) and the Gambia (23.9) with Algeria (1.4), Morocco (1.1) and Nepal (1.1) recording the lowest ASIR (Fig. 1a). The ASMR mostly followed the patterns of ASIR with Mongolia (75.4), Egypt (31.8) and the Gambia (26.2) as the three leading countries and Nepal (1/100 000) had the least ASMR (Fig. 1b). At the global level, the MIR was 0.93, varying from 0.44 in Samoa to 1.6 in Iceland (Fig. 1c).

Among the countries with >100 estimates cases, I also plotted the ASIRs, ASMRs and MIRs as per HDI categories (Fig. 2a–c); this figure illustrates that where ASIR and ASMRs displayed mixed patterns as per HDI levels, in terms of MIR, however, high HDI and very high HDI countries had comparatively lower MIR than low

Downloaded from https://academic.oup.com/jjco/article-abstract/doi/10.1093/jjco/hyaa130/5876941 by Delhi Technological University user on 03 August 2020

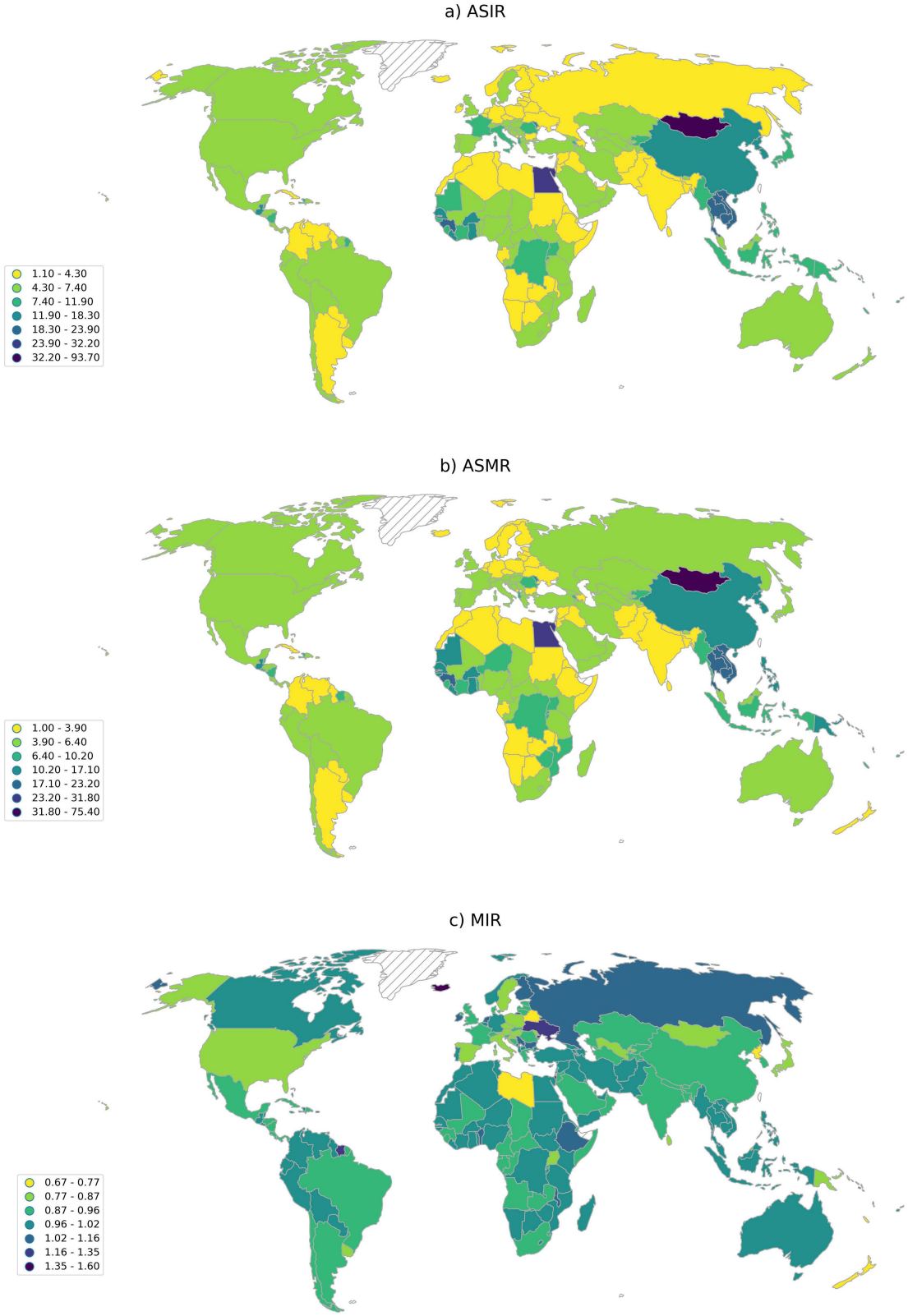


Figure 1. Geographical distribution of primary liver cancer statistics in 2018. (1a) ASIR (1b) ASMR (1c) MIR; ASIR: Age-standardized incidence rate (cases per 100 000); ASMR: Age-standardized mortality rate (deaths per 100 000); MIR: mortality-to-incidence ratio. The MIR was calculated by dividing all-age death counts with all-age incidence. The estimates of liver cancer burden were drawn from GLOBOCAN 2018.

Downloaded from https://academic.oup.com/jjco/article-abstract/doi/10.1093/jjco/hyaa130/5876941 by Delhi Technological University user on 03 August 2020

Table 1. Region-wise burden of primary liver cancer in 2018

Population	Incidence	ASIR	Deaths	ASMR	Prevalence	MIR
Australia and New Zealand	2921 [2796–3051]	5.7	2709 [2591–2833]	4.7	2354	0.93
Caribbean	2923 [2613–3270]	5	2791 [2513–3099]	4.7	2117	0.95
Central America	11 229 [10 870–11 600]	6.3	10 672 [10 376–10 976]	5.9	8453	0.95
Central and Eastern Europe	22 784 [22 150–23 437]	4	22 745 [22 181–23 323]	3.9	16 039	1.00
Eastern Africa	11 550 [8507–15 682]	4.8	11 251 [8067–15 692]	4.7	11 705	0.97
Eastern Asia	467 327 [459 452–475 337]	17.7	427 932 [421 450–434 514]	16	383 221	0.92
Melanesia	915 [789–1061]	11.4	819 [718.7–933.3]	10.6	894	0.90
Micronesia	85 [70–104]	15.2	68 [54–86]	12	24	0.80
Middle Africa	6010 [4325–8351]	6.5	5853 [4248–8064]	6.5	6115	0.97
North America	41 851 [41 406–42 301]	6.6	34 339 [33 925–34 759]	4.8	34 107	0.82
Northern Africa	27 935 [25 427–30 691]	14.1	27 505 [24 873–30 415]	13.9	21 466	0.98
Northern Europe	10 997 [10 693–11 309]	4.7	9997 [9701–10 302]	3.8	7878	0.91
Polynesia	67 [49–92]	9.2	54 [33–89]	7.4	40	0.81
South America	24 248 [21 806–26 964]	4.6	22 973 [22 393–23 569]	4.3	17 225	0.95
South-Central Asia	44 010 [41 162–47 055]	2.5	40 812 [38 208–43 594]	2.3	35 246	0.93
South-Eastern Asia	89 010 [83 523–94 858]	13.3	88 429 [82 662–94 599]	13.2	69 441	0.99
Southern Africa	2710 [2438–3012]	4.9	2597 [2476–2724]	4.7	2344	0.96
Southern Europe	25 026 [23 934–26 168]	6.8	21 996 [21 420–22 587]	5.3	17 262	0.88
Western Africa	16 574 13 547.8–20 276]	8.3	16 356 [13 212–20 248]	8.2	15 106	0.99
Western Asia	9249 [8248–10 372]	4	9096 [8404–9845]	4	6875	0.98
Western Europe	23 659 [22 949–24 391]	5.3	22 637 [22 095–23 193]	4.5	17 298	0.96
Global	841 080 [817 635–865 198]	9.3	781 631 [737 605–828 285]	8.5	675 210	0.93

Incidence: all-age new cases; deaths: all-age death counts; ASIR: age-standardized incidence rate (cases per 100 000); ASMR: age-standardized mortality rate (counts per 100 000); prevalence: 5-year prevalence; MIR: mortality-to-incidence ratio. The MIR was calculated by dividing all-age death counts with all-age incidence. Data source: GLOBOCAN 2018 (IARC).

and medium HDI countries (Fig. 2c).[1] For instance, Belarus (0.67), New Zealand (0.73) and South Korea (0.73), the very high HDI countries, recorded one of the lowest MIRs. Uganda (0.86) in the low HDI category, Zambia (0.89) in medium HDI category and Libya (0.77) and Mongolia (0.79) in the high HDI category recorded the lowest MIR in their respective categories.

Sex-wise burden

Globally, males accounted for 596 574 [576 976–616 838] cases and females accounted for 244 506 [231 861–257840] cases in 2018. Male ASIR (15.5/100 000) was again higher than the female ASIR (6.5/100 000). In terms of death counts too, males accounted for nearly two-third of fatalities with 548 375 [511 583–587 813] deaths

[1] I also plotted ASIR, ASMR and MIR as per world bank country classification in low, lower-middle, upper-middle and high income countries (Supplementary Figure S1). The results were almost similar as to that of HDI (Fig. 2) classifications with few exceptions. Country classification was based on GNI per capita available at https://datahelpdesk.worldbank.org/knowledgebase/articles/906519 (accessed on 13 May, 2020).

Downloaded from https://academic.oup.com/jjco/article-abstract/doi/10.1093/jjco/hyaa130/5876941 by Delhi Technological University user on 03 August 2020

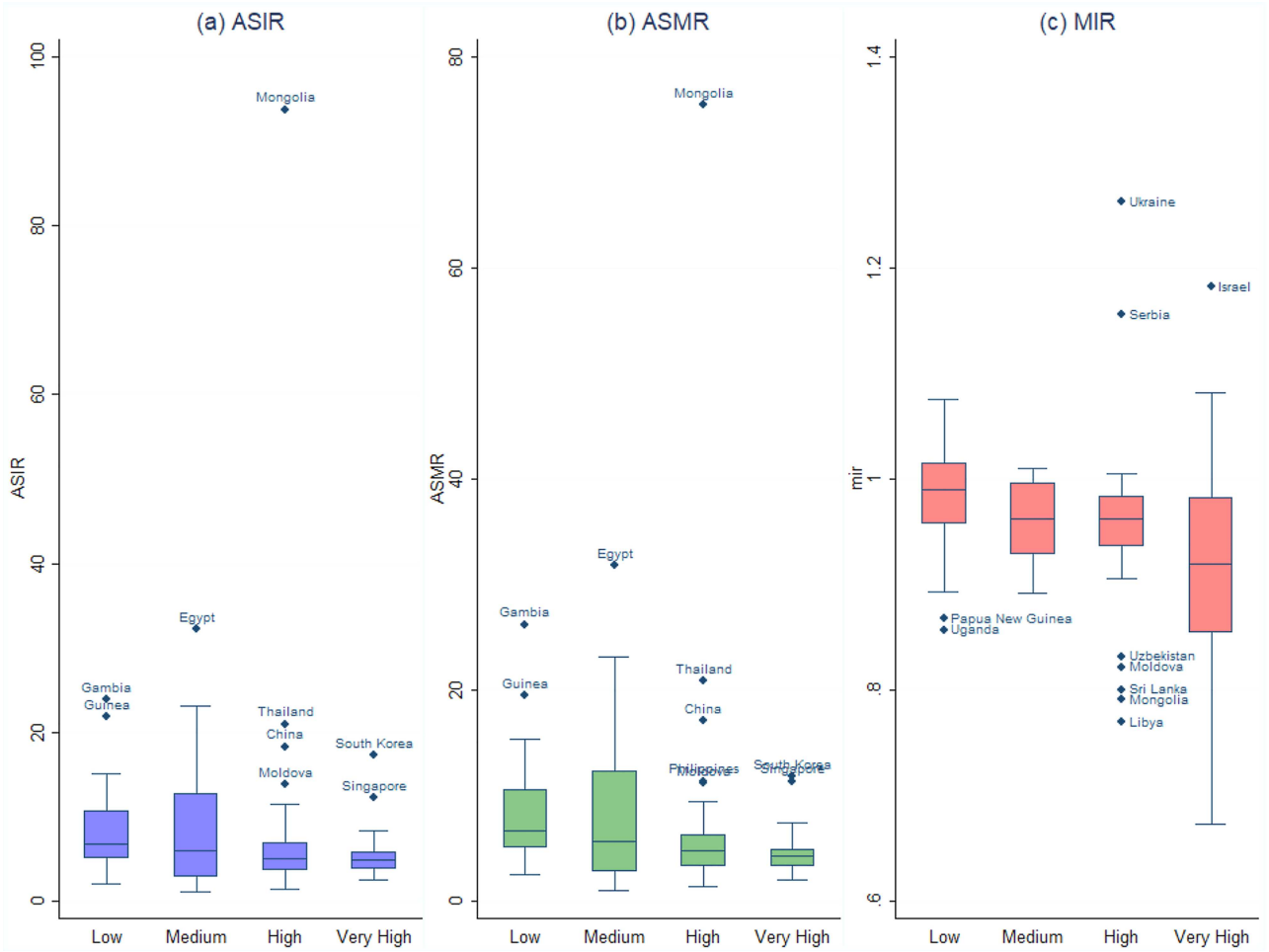


Figure 2. Boxplot of HDI category wise primary liver cancer statistics. (a) ASIR (b) ASMR (c) MIR. The MIR was calculated by dividing all-age death counts with all-age incidence. The MIR was calculated by dividing all-age death counts with all-age incidence. The countries were categorized by United Nations Development Program (UNDP) as low (HDI < 0.550), medium (0.550–0.669), high (0.700–0.799) and very high (>0.800). The estimates of liver cancer burden were drawn from GLOBOCAN 2018.

and the rest of 233 256 [209 922–259 184] deaths due to PLC occurred in females. Male ASMR (14.2/100 000) was also more than double than that of female ASMR (6.2/100 000). The male–female difference in incident cases and death counts was also preserved in regions (Fig. 3a–d). Few regions such as East Asia and South-east Asia had greater disparities in males and females, registering 3-fold difference or more between males and females (Fig. 3a–d). Similar differences were observed across regions in terms of ASIR and ASMR. Few regions in America (e.g. Central America and South America) had lesser male–female disparities in terms of counts as well as age-standardized rates (Fig. 3a–d).

Age-wise distribution

Globally, PLC incidence and deaths peaked in the 64–69 age group with a worsening ratio of death counts and cases in older age groups (Fig. 4). The age-wise distribution of incidence and mortality differed remarkably between different HDI groupings with higher incidence and mortality in older age groups for countries higher in the development spectrum (Supplementary Figure S2). For instance, in low and medium HDI categories, cases and deaths plateaued in 54–59 age groups and declined thereafter, whereas very high HDI category had maximum estimated deaths in the older-most age group (85-plus). Moreover, the low HDI group of countries had a non-negligible portion of cases and deaths in child and adolescent age groups (below the age 20); in contrast, the proportion of PLC cases and deaths in other HDI groupings was much lower in those aged below 20 (Supplementary Figure S2).[2]

Discussion

The PLC cases and deaths increased from 748 000 and 695 000, respectively, in 2008 (GLOBOCAN 2008, [24]) to 841 080 and 781 631, respectively, in 2018. The age-standardized rates underwent a reduction with ASIR falling from 10.8/100 000 in 2008 [24] to 9.8/100 000 in 2018 and the ASMR falling from 10.0/100 000 in

2 Similar results were observed when countries were categorised geographically in 21 world regions. Supplementary Figure S3 illustrates percentage of cases and deaths in four non-overlapping age groups (below-20, 20–29, 50–69 and 70-plus) in 21 world regions. It shows that nearly 30–40% of PLC cases and deaths occurred in the below-50 age groups in all regions of Africa (except North Africa); in contrast, in nearly all European regions, >90% cases and deaths occurred in above-50 age groups (Supplementary Figure S3).

Downloaded from https://academic.oup.com/jjco/article-abstract/doi/10.1093/jjco/hyaa130/5876941 by Delhi Technological University user on 03 August 2020

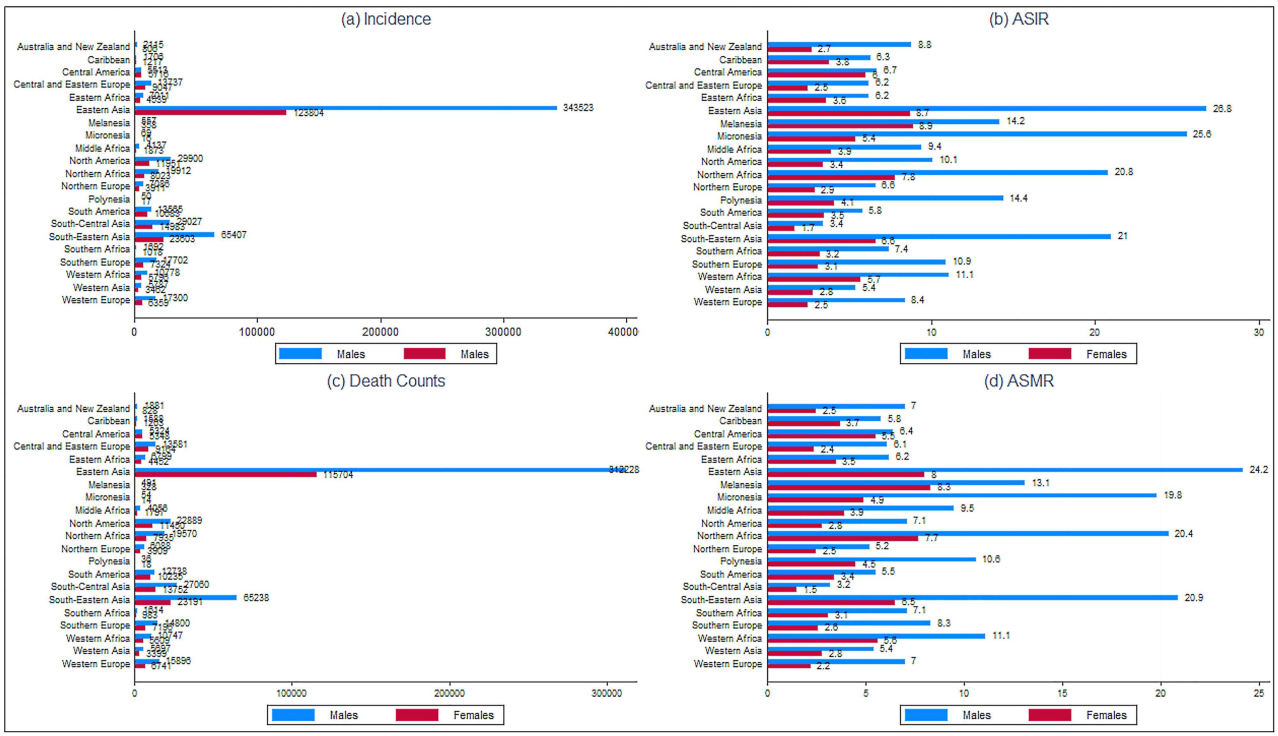


Figure 3. Sex-wise burden of primary liver cancer in 21 regions. Incidence: all-age new cases of liver cancer; deaths: all-age deaths due to liver cancer. The estimates of liver cancer burden were drawn from GLOBOCAN 2018.

2008 to 8.5/100 000 in 2018. PLC has one of the worst survival rates [3], which is also evident from very high MIR of PLC in comparison with other cancer groups. The MIR of PLC was 0.93 in 2018 as per GLOBOCAN 2018 estimates [1], which is only better than that of pancreatic cancer (0.94). In comparison, the MIR of PLC was worse than that of other frequent cancer sites: breast cancer (0.30), prostate cancer (0.28), colorectum cancer (0.48), stomach cancer (0.76) and lung cancer (0.84).[3] East Asia was the top region contributing 55.6% of global cases and 54.7% of global deaths. The age-standardized rates were again the highest in East Asian countries led by Mongolia with ASIR of 93.7/100 000 and ASMR of 75.4/100 000 in 2018.

Concord-3 [3] demonstrated that 5-year survival rates of liver cancer ranged 5–30% across countries, and the highest survival rates ranging between 20 and 29% were observed in South Korea, Singapore and Taiwan. Following previous studies [18–21], MIR was employed as the proxy for 5-year survival rates, and there were marked differences in MIR at the country level. I found that low HDI countries generally recorded high MIRs (i.e. lower 5-year survival rates) and lower MIRs (or higher 5-year survival rates) were recorded by developed countries such as South Korea, Singapore and Japan, which corroborates previous findings of higher 5-year survival rates in developed countries than the developing ones [3].

Males accounted for nearly two-third PLC cases and deaths in 2018, with several regions recording male age-standardized rates triple or more than females (Fig. 3). Such gender-wise disparities were observed in previous epidemiological studies [2] as well as in rodent HCC models, too [25,26]. The male–female differences in PLC incidence are commonly associated with anticarcinogenic effects of female sex hormones [27,28] and lower prevalence of alcohol consumption in females than males [29].

A vast majority of PLC incidence pertains to HCC arising from HBV or HCV infections with HBV infection as the dominant risk factor of HCC in South-east Asia and Africa and HCV infection identified as the leading risk factor of HCC in developed countries such as Japan [30]. HCV infection, for instance, contributes to 70% of HCC cases in Japan [31], which has decreased to 60% recently [32]. World Health Assembly held in May 2016 endorsed Global Health sector strategy 2016–21, which calls for the elimination of viral hepatitis, 96% of which comprises of HBV and HCV, as public health target by 2030 through reduction of new infections by 90% and associated mortality by 65% [33]. Due to increased HBV vaccination coverage, HBV prevalence has decreased across the world in percentage terms over 1990–2005, yet the number of HBsAg positive persons has increased from 223 million to 240 million with high endemicity in Sub-Saharan Africa [34].

Regarding the HCC burden due to HCV infection, the development of direct-acting antiviral drugs (DAAs) has significantly boosted the treatment paradigm of HCV infections [35]. The DAAs, taken orally, target specific proteins, and improves sustained virologic response rate (SVR) to >90% and ~100% in a few instances [35,36]. These DAAs hold enormous potential in developing countries such as India, where HCV treatment rates are low [37]. Moreover, the low-cost generic versions of DAAs make these cost-effective in developing countries such as India [38].

Diabetes mellitus [39,40] and obesity [41] are other identified risk factors of PLC. Obesity and diabetes are not only associated with HCC but are associated with increased risk of ICC, too [42]. The

[3] The MIR for different cancer groups is calculated by the author using GLOBOCAN 2018 estimates [1,16,17].

Downloaded from https://academic.oup.com/jjco/article-abstract/doi/10.1093/jjco/hyaa130/5876941 by Delhi Technological University user on 03 August 2020

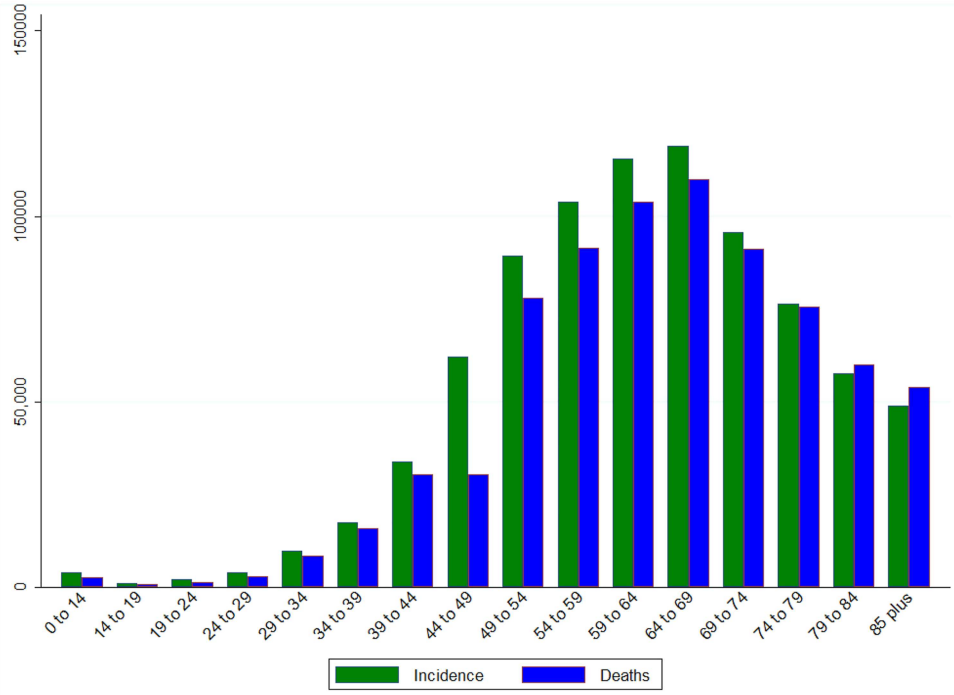


Figure 4. Age-group wise distribution of incidence and mortality of primary liver cancer in 2018. Incidence: new cases; deaths: deaths due to liver cancer. The estimates of liver cancer were drawn from GLOBOCAN 2018.

unabated incidence and prevalence of diabetes mellitus is expected to further result in an increasing burden of chronic liver diseases, including PLC worldwide [43]. In addition to HBV and HCV, the incidence of NAFLD related to HCC has been increasing worldwide, with prevalence varying from 6 to 35% in different countries [44,45].

Moreover, NAFLD has been identified as the fastest-growing cause of cancer-related deaths [46]. The prevalence of NAFLD has increased in parts of East Asia, USA, Europe and Australia [47]. Although the focus on HBV infection and vaccination is expected to result in the downward trend of HCC in Asia, the increasing trend of other risk factors such as obesity, diabetes mellitus and NAFLD is expected to negate the gains from HBV vaccination [31].

China is the leading country in terms of both incidence and death counts. Although China has employed tremendous efforts to shed its leadership position in liver diseases through disease prevention, diagnosis, management and therapeutics [48], still, China was ranked first as per GLOBOCAN 2018 estimates contributing 46.7% cases and 34.9% deaths in 2018. HBV vaccination is fully embedded into the neonatal immunization program in China since 2002 [49]. Despite expanded access to HBV vaccine in China, 20% of children aged below 5 years, and 40% of those aged 5–19 remain unprotected from HBV infection [50]. As described earlier, China may have worked tremendously for expanded access to HBV vaccination; the growth of other risk factors such as obesity may contribute to a greater burden in the future. For instance, China experienced a 414% increase in overweight and obesity between 1982 and 2002 [51].

In several low and medium HDI countries, PLC exerted a significant burden in child and adolescent age groups (Supplementary Figure S2). Pediatric HCC is different from adult HCC in a few key aspects [52,53]. For instance, in adults, HCC arises along with prolonged chronic hepatitis and cirrhosis [54]; in children and adolescents, however, HBV infection is acquired perinatally or due to inherited metabolic disorders [55,56]. Moreover, HCC in children and adolescents is more advanced than in adults [57]. Thus, screening of children whose mothers are HBV carriers is essential in the early detection of pediatric HCC [58].

As far as early screening and detection of liver cancer in adults is concerned, few approaches using ultrasonography, α-fetoprotein (AFP) levels in serum, and surveillance of cirrhosis patients have been considered. The American Association for liver diseases (AASLD) in 2005 issued level 1 recommendation for screening of liver cancer using ultrasonography at six-month intervals for high-risk patients [59]. The AASLD's recommendation, however, was based on a randomized control trial (RCT) conducted in China that demonstrated mortality-reduction benefits of liver cancer screening in the treatment group [60]. The reported benefits of screening, however, were put into question [61]. Currently, major US agencies, which provide recommendations on cancer screening, do not endorse standard or routine screening for liver cancer, as is the case with mammographic screening for breast cancer and colonoscopy for colorectal cancer.[4]

Although, screening for liver cancer hasn't been recommended, United States Preventive Services Task Force (USPSTF) recommends screening for HCV infection in adults aged 20–79 years [64]. It was done due to two reasons. First, there is a sharp spurt in HCV infection in young adults aged 20–39 years in the USA, which coincides with

[4] United States Preventive Services Task Force (USPSTF) has not recommended for liver cancer screening; the American cancer society found no convincing evidence for screening recommendation of PLC [62]; the national cancer institute (NCI) found no convincing beneficial effects of PLC screening [63].

Downloaded from https://academic.oup.com/jjco/article-abstract/doi/10.1093/jjco/hyaa130/5876941 by Delhi Technological University user on 03 August 2020

nation-wide surge in the rate of injection drug use [65]; second, nearly half of HCV patients are unaware of the disease; thereby leading to late diagnosis, which results in life-threatening complications such as cirrhosis and HCC [66].

In summary, apart from worse survival rates, the PLC epidemiology is profoundly different from other frequent cancer groups in a few key aspects. First, geographically other cancer groups such as breast cancer [21], colorectal cancer [67] and stomach cancer [68] are quite dispersed across the world; the PLC, however, is highly concentrated in Asia. Second, while screening modalities are available for breast cancer, colorectal cancer and prostate cancer [20,21,67], there are less-than-ideal or proven screening modalities available for early screening and detection of PLC. Third, a non-negligible number of cases and deaths occurred due to liver cancer in children and adolescents, which is quite different from other frequent cancer groups such as prostate cancer [20], breast cancer [21] and colorectal cancer [67], which predominantly affect comparatively older age groups.

Limitations

There are two limitations of the study. First, the reliability of GLOBOCAN estimates crucially depends upon data availability from population-based cancer registries (PBCRs). In several low- and medium-income countries, however, either PBCRs are not available or lack sufficient coverage. In the absence of reliable and high-quality data from PBCRs, the GLOBOCAN estimates draw strength from data available either from regional registries or from neighbouring countries; the estimates reported and examined in this paper, therefore, are expected to be biased. Second, I employed HDI as an indicator of socio-economic status of a country, which is one of the widely employed indicators of socio-economic status of a country. For a set of countries, with data unavailability, the incidence to mortality ratios (method 3a of incidence) and MIRs (method 3 of mortality) were scaled as per HDI of countries and HDI of respective regions [16].[5] The general patterns of results, however, were similar if GNI per capita was considered as the indicator of development.

Conclusions

In the absence of proven screening devices for PLC, primary prevention through containment of risk factors seems to be the apt choice as a large percentage of PLC cases can be prevented through HBV vaccination, containment of alcohol consumption and obesity control. A large portion of PLC is HCC, and there exists proven vaccination against HBV; it is time to identify high-risk areas and scale up the vaccination in high-risk adults who would have missed the compulsory doses of the HBV vaccine. Moreover, as effective treatment of HCV infection is available through DAAs, early detection of HCV infection must be ensured to prevent HCV-induced sequelae such as cirrhosis and HCC. Lastly, policy makers must keep a strict vigil on the increasing prevalence of risk factors such as NAFLD, obesity and diabetes, which can reverse the gains achieved through HBV vaccination and DAA treatment regimens.

[5] As per Ferlay et al [16], the formula of estimation of such countries is modified as below.Method 3b: $I_{\text{National}} = M_{\text{National}} * \left(\frac{I_{\text{Regional}}}{M_{\text{Regional}}}\right) * \left(\frac{\text{HDI}_{\text{National}}}{\text{HDI}_{\text{Regional}}}\right)$
Method 3: $M_{\text{National}} = I_{\text{National}} * \left(\frac{M_{\text{Regional}}}{I_{\text{Regional}}}\right) * \left(\frac{\text{HDI}_{\text{Regional}}}{\text{HDI}_{\text{National}}}\right)$

Compliance with ethical standards

Disclosure of potential conflicts of interest

The author declare that they have no conflict of interest.

Research involving human participants and/or animals

This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent

The research was conducted using data available in the public domain and did not include any human participants or animals.

Supplementary material

Supplementary material can be found at *Japanese Journal of Clinical Oncology* online.

Acknowledgement

I thank International Agency for Research on Cancer for providing GLOBOCAN 2018 data in the public domain.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394–424.
2. Global Burden of Disease Liver Cancer Collaboration. The burden of primary liver cancer and underlying Etiologies from 1990 to 2015 at the global, regional, and National Level: results from the global burden of disease study 2015. *JAMA Oncol*. 2017;3:1683–91.
3. Allemani C, Matsuda T, Di Carlo V, *et al*. Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet* 2018;391:1023–75.
4. Sia D, Villanueva A, Friedman SL, Llovet JM. Liver cancer cell of origin, molecular class, and effects on patient prognosis. *Gastroenterol* 2017;152:745–61.
5. Massarweh NN, El-Serag HB. Epidemiology of hepatocellular carcinoma and intrahepatic cholangiocarcinoma. *Cancer Control* 2017;24:1073274817729245.
6. de Martel C, Maucort-Boulch D, Plummer M, Franceschi S. Worldwide relative contribution of hepatitis B and C viruses in hepatocellular carcinoma. *Hepatology* 2015;62:1190–200.
7. Bosetti C, Turati F, La Vecchia C. Hepatocellular carcinoma epidemiology. *Best Pract Res Clin Gastroenterol* 2014;28:753–70.
8. Kurtzman CP, Horn BW, Hesseltine CW. Aspergillus nomius, a new aflatoxin-producing species related to Aspergillus flavus and Aspergillus tamarii. *Antonie van leeuwenhoek* 1987;53:147–58.
9. Fattovich G, Stroffolini T, Zagni I, Donato F. Hepatocellular carcinoma in cirrhosis: incidence and risk factors. *Gastroenterol* 2004;127:S35–50.
10. Khanna R, Verma SK. Pediatric hepatocellular carcinoma. *World J Gastroenterol*. 2018;24:3980–99.
11. Zhang XF, Liu XM, Wei T, Liu C, Li MX, Long ZD, Lv Y. Clinical characteristics and outcome of hepatocellular carcinoma in children and adolescents. *Pediatr Surg Int* 2013;29:763–70.
12. Hung GY, Horng JL, Yen HJ, Lee CY, Lin LY. Changing incidence patterns of hepatocellular carcinoma among age groups in Taiwan. *J Hepatol* 2015;63:1390–6.

Downloaded from https://academic.oup.com/jjco/article-abstract/doi/10.1093/jjco/hyaa130/5876941 by Delhi Technological University user on 03 August 2020

13. Chang MH, Chen DS, Hsu HC, Hsu HY, Lee CY. Maternal transmission of hepatitis B virus in childhood hepatocellular carcinoma. *Cancer* 1989;64:2377–80.
14. Liang X, Bi S, Yang W, *et al*. Epidemiological serosurvey of hepatitis B in China—declining HBV prevalence due to hepatitis B vaccination. *Vaccine* 2009;27:6550–7.
15. Lok AS, McMahon BJ. Chronic hepatitis B. *Hepatol* 2007;45:507–39.
16. Ferlay J, Colombet M, Soerjomataram I, *et al*. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer* 2019;144:1941–53.
17. Ferlay J, Ervik M, Lam F, *et al*. *Global Cancer Observatory*. https://gco.iarc.fr/. (28 June 2020, date last accessed).
18. Vostakolaei F, Karim-Kos HE, Janssen-Heijnen ML, Visser O, Verbeek AL, Kiemeney LA. The validity of the mortality to incidence ratio as a proxy for site-specific cancer survival. *Eur J Public Health* 2010;21:573–7.
19. Choi E, Lee S, Nhung BC, Suh M, Park B, Jun JK, Choi KS. Cancer mortality-to-incidence ratio as an indicator of cancer management outcomes in Organization for Economic Cooperation and Development countries. *Epidemiol Health* 2017;39:e2017006.
20. Sharma R. The burden of prostate cancer is associated with human development index: evidence from 87 countries, 1990–2016. *EPMA J* 2019a;10:137–52.
21. Sharma R. Breast cancer incidence, mortality and mortality-to-incidence ratio (MIR) are associated with human development, 1990–2016: evidence from global burden of disease study 2016. *Breast Cancer* 2019b;26:428–45.
22. UNDP. *Human Development Statistical Update 2018*. http://hdr.undp.org/sites/default/files/2018_human_development_statistical_update.pdf (24 June 2019, date last accessed).
23. Human Development Database: http://hdr.undp.org/en/data# (20 June 2020, date last accessed).
24. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010;127:2893–917.
25. Ghebranious N, Sell S. Hepatitis B injury, male gender, aflatoxin, and p53 expression each contribute to hepatocarcinogenesis in transgenic mice. *Hepatology* 1998;27:383–91.
26. Maeda S, Kamata H, Luo JL, Leffert H, Karin M. IKKβ couples hepatocyte death to cytokine-driven compensatory proliferation that promotes chemical hepatocarcinogenesis. *Cell* 2005;121:977–90.
27. De Maria N, Manno M, Villa E. Sex hormones and liver cancer. *Mol Cell Endocrinol* 2002;193:59–63.
28. Yeh SH, Chen PJ. Gender disparity of hepatocellular carcinoma: the roles of sex hormones. *Oncology* 2010;78:172–9.
29. Wilsnack RW, Wilsnack SC, Kristjanson AF, Vogeltanz-Holm ND, Gmel G. Gender and alcohol consumption: patterns from the multinational GENACIS project. *Addiction* 2009;104:1487–500.
30. Mittal S, El-Serag HB. Epidemiology of hepatocellular carcinoma: consider the population. *J Clin Gastroenterol* 2013;47:S2–6.
31. Choo SP, Tan WL, Goh BK, Tai WM, Zhu AX. Comparison of hepatocellular carcinoma in eastern versus western populations. *Cancer* 2016;122:3430–46.
32. Kudo M, Izumi N, Kubo S, *et al*. Report of the 20th Nationwide follow-up survey of primary liver cancer in Japan. *Hepatol Res* 2020;50:15–46.
33. WHO. *Global Hepatitis Report*. https://apps.who.int/iris/bitstream/handle/10665/255016/9789241565455-eng.pdf (6 May 2020, date last accessed).
34. Ott JJ, Stevens GA, Groeger J, Wiersma ST. Global epidemiology of hepatitis B virus infection: new estimates of age-specific HBsAg seroprevalence and endemicity. *Vaccine* 2012;30:2212–9.
35. Centers for Disease Control and Prevention. *Viral Hepatitis Surveillance United States*. https://www.cdc.gov/hepatitis/statistics/2014surveillance/pdfs/2014hepsurveillancerpt.pdf (26 June 2020, date last accessed).
36. Asselah T, Marcellin P, Schinazi RF. Treatment of hepatitis C virus infection with direct-acting antiviral agents: 100% cure? *Liver Int* 2018;38:7–13.
37. Blach S, Zeuzem S, Manns M, *et al*. Global prevalence and genotype distribution of hepatitis C virus infection in 2015: a modelling study. *Lancet Gastroenterol Hepatol* 2017;2:161–76.
38. Aggarwal R, Chen Q, Goel A, Seguy N, Pendse R, Ayer T, Chhatwal J. Cost-effectiveness of hepatitis C treatment using generic direct-acting antivirals available in India. *PloS one* 2017;12:e0176503.
39. Adami H-O, Chow W-H, Nyrén O, *et al*. Excess risk of primary liver cancer in patients with diabetes mellitus. *J Natl Cancer Inst* 1996;88:1472–7.
40. El-serag HB, Tran T, Everhart JE. Diabetes increases the risk of chronic liver disease and hepatocellular carcinoma. *Gastroenterol* 2004;126:460–8.
41. Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. *N Engl J Med* 2003;348:1625–38.
42. Petrick JL, Thistle JE, Zeleniuch-Jacquotte A, *et al*. Body mass index, diabetes and intrahepatic cholangiocarcinoma risk: the liver cancer pooling project and meta-analysis. *Am J Gastroenterol* 2018;113:1494.
43. Wild SH, Morling JR, McAllister DA, *et al*. Type 2 diabetes and risk of hospital admission or death for chronic liver diseases. *J hepatol* 2016;64:1358–64.
44. Bellentani S. The epidemiology of non-alcoholic fatty liver disease. *Liver Int* 2017;37:81–4.
45. Said A, Ghufran A. Epidemic of non-alcoholic fatty liver disease and hepatocellular carcinoma. *World J Clin Oncol* 2017;8:429–36.
46. Zoller H, Tilg H. Nonalcoholic fatty liver disease and hepatocellular carcinoma. *Metabolism* 2016;65:1151–60.
47. Farrell GC, Wong VW, Chitturi S. NAFLD in Asia—as common and important as in the west. *Nat Rev Gastroenterol Hepatol* 2013;10:307–18.
48. Wang FS, Fan JG, Zhang Z, Gao B, Wang HY. The global burden of liver disease: the major impact of China. *Hepatology* 2014;60:2099–108.
49. Qu C, Duan Z, Chen K, Zou H. Reducing liver cancer risk beginning at birth: experiences of preventing chronic hepatitis B virus infection in China. *Hepatoma Res* 2017;3:228–40.
50. Hutton DW, So SK, Brandeau ML. Cost-effectiveness of nationwide hepatitis B catch-up vaccination among children and adolescents in China. *Hepatology* 2010;51:405–14.
51. Asia Pacific Cohort Studies Collaboration. The burden of overweight and obesity in the Asia–Pacific region. *Obes Rev* 2007;8:191–6.
52. Czauderna P. Adult type vs. childhood hepatocellular carcinoma–are they the same or different lesions? Biology, natural history, prognosis, and treatment. *Med Pediatr Oncol* 2002;39:519–23.
53. Chen JC, Chen CC, Chen WJ, Lai HS, Hung WT, Lee PH. Hepatocellular carcinoma in children: clinical review and comparison with adult cases. *J Pediatr Surg* 1998;33:1350–4.
54. El-Serag HB. Hepatocellular carcinoma. *N Engl J Med* 2011;365:1118–27. doi: 10.1056/nejm ra 1001683.
55. McAteer JP, Goldin AB, Healey PJ, Gow KW. Surgical treatment of primary liver tumors in children: outcomes analysis of resection and transplantation in the SEER database. *Pediatr Transplant* 2013;17:744–50.
56. Darbari A, Sabin KM, Shapiro CN, Schwarz KB. Epidemiology of primary hepatic malignancies in US children. *Hepatology* 2003;38:560–6.
57. Wang J, Mao Y, Liu Y, Chen Z, Chen M, Lao X, Li S. Hepatocellular carcinoma in children and adolescents: clinical characteristics and treatment. *J Gastrointest Surg* 2017;21:1128–35.
58. Zhang XF, Liu XM, Wei T, Liu C, Li MX, Long ZD, Lv Y. Clinical characteristics and outcome of hepatocellular carcinoma in children and adolescents. *Pediatr Surg Int* 2013;29:763–70.
59. Bruix J, Sherman M. Management of hepatocellular carcinoma: an update. *Hepatol* 2011;53:1020–2.
60. Zhang BH, Yang BH, Tang ZY. Randomized controlled trial of screening for hepatocellular carcinoma. *J Cancer Res Clin Oncol* 2004;130:417–22.
61. Lederle FA, Pocha C. Screening for liver cancer: the rush to judgment. *Ann Intern Med* 2012;156:387–9.

Downloaded from https://academic.oup.com/jjco/article-abstract/doi/10.1093/jjco/hyaa130/5876941 by Delhi Technological University user on 03 August 2020

62. American Cancer Society. *Can Liver Cancer Be Found Early?* https://www.cancer.org/cancer/liver-cancer/detection-diagnosis-staging/detection.html (28 June 2020, date last accessed).
63. National Cancer Institute. *Liver (Hepatocellular) Cancer Screening–Health Professional Version*. https://www.cancer.gov/types/liver/hp/liver-screening-pdq (28 June 2020, date last accessed).
64. Owens DK, Davidson KW, Krist AH, *et al*. Screening for hepatitis C virus infection in adolescents and adults: US preventive services task force recommendation statement. *JAMA* 2020;323:970–5.
65. Zibbell JE, Asher AK, Patel RC, Kupronis B, Iqbal K, Ward JW, Holtzman D. Increases in acute hepatitis C virus infection related to a growing opioid epidemic and associated injection drug use, United States, 2004 to 2014. *Am J Public Health* 2018;108:175–81.
66. Kapila N, Muir AJ. A big step forward in hepatitis C screening. *Clin Liver Dis* 2020;15:149.
67. Sharma R. An examination of colorectal cancer burden by socioeconomic status: evidence from GLOBOCAN 2018. *EPMA J* 2020;11:95–117.
68. Etemadi A, Safiri S, Sepanlou SG, *et al*. The global, regional, and national burden of stomach cancer in 195 countries, 1990–2017: a systematic analysis for the global burden of disease study 2017. *Lancet Gastroenterol Hepatol* 2020;5:42–54.

Downloaded from https://academic.oup.com/jjco/article-abstract/doi/10.1093/jjco/hyaa130/5876941 by Delhi Technological University user on 03 August 2020

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Detecting Intrusions and Attacks in the Network Traffic using Anomaly based Techniques

Vinod Kumar
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
vinod_k@dtu.ac.in

Vinay Choudhary
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
choudharyvinay92@gmail.com

Vivek Sahrawat
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
viveksahrawat7@gmail.com

Vinay Kumar
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
vinayverma9650@gmail.com

***Abstract*— Technology has become the backbone of today's Information and Communication Technology. Today a large number of transactions are carried out online and thus possess a risk of information stealing and misuse. Intrusion Detection techniques are thus developed and applied to the networks and individual host computers to detect such intrusive actions and alarm the administrators. A secure machine-controlled anomaly detection system is a more practical procedure to help in network analysis. An anomaly detection system (ADS) examine the network flow and focuses on detecting uncommon network behaviour, and classify them into attacks. The propsoed research work present a comparison of some of these intrusion detection techniques using different machine learning algorithms and ensemble learning. A subset of attributes (significant) is chosen from the primitive set of attributes using a random forest regressor technique, and then, the chosen set of significant features is used to train different classifiers. Moreover, the proposed work has been analyzed by experimenting on one of the publically available dataset CICIDS2017, available from the Canadian University of Cybersecurity.**

***Keywords*— Classification, anomaly detection system (ADS), CICIDS2017, machine learning.**

I. INTRODUCTION

Established institutions and ample folks communicate with one another over the internet every day. Within the past twenty years, where the number of individuals using the net services has been hyperbolic. Today this range has exceeded four billion and is increasing exponentially. Moreover, any business, government enterprise, or industry relies heavily on technology to conduct their day to day activities, transactions, or communicate. The data today is not restricted to have a physical presence, the e-money, photos, texts, id passwords, videos, and many more forms of virtual or online data also exists. And there is a need to preserve and safeguard this online data. There is an improvement in infrastructures and the protocols of the wireless devices and IoT applications. Online services and networks are often used for communicating and exchanging sensitive information. Attackers often peep into these networks and get access to sensitive and confidential information which suggests the requirement for some mechanism to detect such actions and then alert the authorities about the same.

The recent COVID-19 outbreak further forced the world to adopt the concept of work from home to keep their services and businesses running, and this remote working mechanism comes with a whole new and wider dimension of internet and network security, with making their networks more vulnerable to cyber-attacks. With the remote working concept, a greater amount of sensitive and confidential information has to be shared via the network, which brings dozens of opportunities for the hackers and the intruders to steal this sensitive information and data. Mobile office and work from home concept allow the employees, who are authorized to access the corporate network and have to share sensitive information over the mobile or computer networks.

Every aspect of cybersecurity reviles around the CIA triad. The C here stands for confidentiality, which means keeping the data hidden from unwanted access. I stand for integrity means keeping the data unaltered throughout the transmission process. The sent and received data over the net should be the same. Finally, the A stands for availability, meaning the data should be available to the intended use when required. And security aspects enforces us to be ready with backups whenever the data is lost or is delayed. There are some major network attacks:

- DNS cache poisoning attack, it tricks the DNS

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:59:53 UTC from IEEE Xplore. Restrictions apply.

servers and makes it accept a fake DNS address of the corresponding IP address of the requested legitimate websites. This attack is primarily used for malicious banking credential stealing.

- Another major network attack is the man-in-the-middle attack. An intruder places itself between two hosts, who think that they are directly communicating with each other but the communication is happening between the middle attacker with each of the two hosts simultaneously. This attack has the potential to access and modify the information passing to and fro the network and hence compromising the CIA triad principles.
- Denial of service attack challenges the availability principle of security goals. The hosts or the severs are overwhelmed by bombarding them with huge unnecessary requests, which results in the crashing of the systems and making legitimate users starve for accessing the required resources. This may delay the entire service of an enterprise and severely affect the time-bound services, which can cause slow network performance. The DoS attack may also be made in a distributed fashion and is known as distributed denial of service attack, where a combined network of physically separate machines are used to overwhelm the server with the request.

It is a challenge for the researchers to develop efficient algorithms to detect such types of network attacks before harm is caused to the victim organization. For this purpose, some benchmarking datasets are required to evaluate the developed algorithms and techniques.

In this paper, some tasks are implemented for detecting an anomaly in networks by making use of effective machine learning techniques. Also, for evaluating the most useful features or attributes and consider them for further study, a random forest regressor algorithm is used during pre-processing.

The rest of the research paper is structured into various sections. The related work that has been done in the field of anomaly detection is highlighted in section 2. Section 3 presents the background details regarding the dataset and algorithms used for our study. Section 4 consists of the details of the pre-processing stage and the experimental setup. Section 5 then compares the results of various algorithms and analyses them. Section 6 concludes our work and Section 7 presents the scope for future work.

II. RELATED WORK

For anomaly detection in networks by machine learning techniques, a large quantity of harmful and harmless network traffic data is required for training and testing steps, However, real-time network traffic can't be used in public due to security problems [1] and thus to develop efficient intrusion detection systems, several datasets have been simulated in almost real environments. DARPA 98, KDD 99, CAIDA (Centre of Applied Internet Data Analysis), NSL-KDD, ISXC 2012 and, CICIDS 2017 are various datasets for network anomaly detection. The initially available datasets were DARPA dataset [2], which was made publically available in 1998, and several improvements were made to it after that."Contests were organized in the initial years for introducing effective new techniques for intrusion detection. With all the advancements and new emerging attacks, there was a need for more comprehensive, balanced, up to date and perfect datasets" [3]–[4]. Recently, the Canadian University of Cybersecurity launched a new dataset, containing novel attacks. This dataset is named as CICIDS-2017. Moreover, this dataset also incorporates all the 11 criteria of a good dataset, as suggested by the authors of [8].

For anomaly detection, these datasets are used as a benchmark in performance evaluation. There is no single "Fit-For-All" algorithm for detecting various attacks differing enormously based on their source, target, and mode of application. The anomaly detection systems are often classified based on the type of environment in which they are deployed as "Host Anomaly Detection Systems" and "Network Anomaly Detection Systems" [6]. The proposed work aim to detect attacks by finding anomalies in the "Network Anomaly Detection Systems" in this paper. Another classification of ADS can be based on techniques employed in implementing them. There are mainly two categories: "Signature-based ADS" and "Anomaly-based ADS". Signature-based ADS were prevalent in real-world applications since the last decade. But due to their inability to detect zero-day attacks, they are replaced by anomaly-based ADS. Signature-based ADS were based on concepts like string matching, regular expressions, and finite state machines [7].

III. BACKGROUND

A. Network Intrusion Detection System

The network of computers was first secured by Network Hardening Principle. Network hardening refers to the process of securing networks by making configuration changes to reduce vulnerabilities. But these changes are often a burden on the administration and hence other mechanisms were evaluated. One such mechanism evolved was of firewall, where one can disable and filter the unnecessary services and restrict access to them. This was achieved but creating manual rules for traffic, similar to the white-listing process. This came with its challenge of adding or deleting new rules, every time some new service was desired to work or restrict and thus causing inconvenience for the administrators.

All these challenges were resolved with the advancement in the field of artificial intelligence. Machine learning concepts

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:59:53 UTC from IEEE Xplore. Restrictions apply.

provided an important and efficient way to monitor and analyze the traffic flowing across the networks. These techniques let us make a baseline of what our normal network traffic looks like and then any deviation from this baseline can be treated as an anomaly and users can be alarmed regarding the suspicious network traffic presence. The requirements for these anomaly detection systems are :

- an architecture to collect and extract the traffic data,
- a good dataset schema to store the data along with its attributes, and
- efficient algorithms to analyze this stored data to deduce outliers and intrusions from it.

Today most of the systems employ more efficient and effective anomaly-based systems using the machine learning algorithms which can generalize the type of attacks from real-time network traffic. Machine learning algorithms generally consists of two phases – the "learning phase" and the "inference phase" [8]. The learning phase enables us to build the network traffic profile, which generalizes the benign and normal traffic passing through the network. During the inference phase, any abnormal behaviour is treated as an anomaly and indicates the presence of attack, and the system administrators are alerted about the same.

For anomaly detection in networks by machine learning techniques, a large quantity of harmful and harmless network traffic data is required for training and testing steps. The anomaly detection-based approach focuses on detecting uncommon network behaviour by examining network flow. This anomaly detection technique has been booming in detecting attacks that's not encountered yet, thus it is effective against zero-day attacks.
There are various difficulties in the implementation of NIDS:

- First, the detection rate (DR) is needed to be enhanced and False Prediction Rate should be reduced.
- Second, extracting, recording, and selecting the important features is a very crucial process as the classifier's performance is completely dependent on the choice of input features. Hence, a dataset with good quality and quantity of features is desired.
- Third, unbalanced and skewed datasets can become a significant hurdle in the wholesome analysis of the performance of proposed algorithms, and results may be skewed. Therefore, the dataset must be well balanced and contain records for all attacks and labels in the right prevalence ratio.

B. Description of CICIDS 2017 Dataset

At the University of New Brunswick, the Canadian Institute for Cybersecurity created CICIDS2017 [9]. This dataset consists of a 5-day (3rd July- 7th July 2017) data stream on a network created by computers having up-to-date operating systems like Windows7 / 8.1 / 10, Ubuntu 12/16, Kali and Mac. DoS (Denial of Service) HULK, DDoS (Distributed DoS), FTP-Patator, DoS Slowloris, SSH-Patator, DoS SlowHTTPTest, DoS Goldeneye, Botnet, PortScan, Web Attack, Infiltration, Heartbleed are various types of attacks present in the datasets available. The 2017 CICIDS dataset has the subsequent benefits over other datasets:

- The obtained information is the real-world data; obtained from a testbed consisting of real computers.
- Data streams are collected from computer systems with the up-to-date operating system. There's operating system diversity (Windows, Linux, and Mac) between each victim computer.
- Datasets are labeled. To apply the machine learning techniques, feature extraction, which is an important step, was applied and 85 features (see Appendix A for the feature list) were obtained.
- Both raw information (pcap files - captured network packets files) and processed data (CSV files- comma-separated data files) are out there to work on.
- In the course of deciding which attack to take place, the 2016 McAfee security report was used, therefore there's a large and up-to-date assortment of attacks.
- It is more abundant than other datasets regarding protocols used. It conjointly includes the HTTPS (Hypertext Transfer Protocol Secure) protocol additionally to FTP, HTTP, SSH, and e-mail protocols.

On the other hand, CICIDS2017 dataset has subsequent disadvantages:

- Raw data files and processed data files are giant (97.9 GB and 1147.3 MB respectively).
- Unlike the KDD99 and NSL-KDD datasets, CICIDS2017 doesn't have separate files dedicated to training and testing. These sections ought to be created by users. The way to do that is handled within the creation of the training and testing section.
- The formation of DARPA98, KDD99 then NSL-KDD is analogous to an evolutionary process. At every step, faults and deficiencies within the previous dataset have been discovered, and measures were taken to forestall them. On the other hand, CICIDS 2017 is incredibly new and has not been studied much yet, therefore it's possible to contain some minor mistakes. What these mistakes are and the way to repair them is processed within the Data Pre-processing section.

Dataset is divided into a training dataset and a testing dataset. 80% of original dataset instances are selected randomly and referred to as a training dataset and the remaining instances are referred to as testing datasets.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:59:53 UTC from IEEE Xplore. Restrictions apply.

C. Random Forest As Feature Selection Technique

Random forest works on the concept of the fair election as result is based on the prediction of several independent decision trees and a group of decision trees having the same prediction will outperform individual decision trees. It uses a random subset of a given dataset and a random subset of features for creating an individual decision tree so that individual decision tree has little amount of correlation with each other.

To ensure that individual decision tree has little amount of correlation with each other random forest uses the following two methods:

- Bagging (Bootstrap Aggregation) – Decision trees are very sensitive to changes in the dataset. The random forest takes advantage of this property of decision trees and allows every individual decision tree to select a subset of the dataset from the available dataset in a random way. This results in independent and different decision trees. This process of allowing random selection of data is known as bagging.
- Randomness in feature – When a node is splitted in a decision tree, every possible feature is used to select the path to be taken in the process of making predictions. In contrast, random forest allows every individual decision tree to select a subset of features from the available feature list in a random way. This results in even more variations among the different decision trees and lower correlation across trees and tress become more independent from each other.

For feature selection, random forest technique assigns weight to different features according to how much a particular feature reduces the accuracy in the classification of the individual decision tree. Dependent variables have a low weight assigned as the Random forest technique ensures that there is a little amount of correlation between individual decision trees.

D. Ensemble Learning

The ensemble method's goal is to mix the prediction results of multiple learning algorithms. Many base classifiers/regressors designed and trained a different learning algorithm for better predictions than individual learning models. Ensemble Methods are divided into two groups

- Sequential ensemble methods – Here base learners are geerated sequentially. The dependence of base learners on each other is exploited here. AdaBoost is a suitable example.
- Parallel ensemble methods – Here base learners are generated parallely. The independence of the base learners is used here. Random Forest is an example.

The concept behind the ensemble classifier is to mix different machine learning algorithms and based on majority vote also called hard voting or the average prediction chances also called soft voting to make predictions. Such a classifier is often useful for a group of equally well-performing classifier/regressor to balance out their weaknesses. In the majority vote, the output for a selected sample is that category label that represents the bulk of the category labels classified by each regressor/classifier. e.g. Different predictions for a given sample (To classify whether the given image is of a hot dog or not).

- model 1 -> yes
- model 2 -> yes
- model 3 -> no

then using the ensemble learning model will predict "yes" based on the majority class prediction. Soft vote predicts the category label as an average of the sum of predicted probabilities. Specific weights are often assigned to each classifier/regressor through the weight parameter. Once weights are assigned, the expected class label probabilities for each classifier/regressor are collected, multiplied by the allotted classifier's weight, and averaged. The final class label is then predicted from the individual category label with the best average likelihood.

IV. EXPERIMENTS

A. Data pre-processing

Various techniques for data pre-processing are performed on the dataset downloaded from the official CICIDS2017 website [10]. Dataset is pre-processed to get rid of the mistakes and defects present. It's necessary to pre-process the dataset before training and testing. For this, some defects within the CICIDS2017 dataset are processed and corrected. One major is the class imbalance problem [12], which is corrected by following the pre-processing stage.

The dataset contains 3119345 records, collected, and analyzed using CICFlowMeter [11]. Once the records are examined, it is seen that around 2.5 lacks records are incomplete or incorrect. The primary step is to delete the surplus records. The next error within the dataset is within the columns that form up the features. The dataset contains 86 columns that define the properties e.g. Flow ID, Source IP, Source Port, etc. The Fwd Header Length feature (which tells about the forward direction data flow for total bytes used) was repeated two times one in the 41st column and second time in the 62nd column. This error is removed by deleting the repeating columns. Table 1 contains the description of the dataset.

Another editing that needs to be done in the dataset is to convert the value of the features including string and categorical values (Flow ID, Source IP, Timestamp, External IP, Destination IP) into numerical data to be easily understandable by different machine learning algorithms. This

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:59:53 UTC from IEEE Xplore. Restrictions apply.

is often be done using LabelEncoder() function provided by the Sklearn package. In this way, all string values that don't seem to be understandable by machine learning operations can get integer values and become more efficient for further processing. The "Label" tagged column is also modified, if there is an attack then it's value will be 1 and if there is no attack then it's value will be 0. Finally, some structural changes are done to the dataset, including:

- In the Label feature, the character "–" (Unicode Decimal Code &#8211) used to identify the web attack subtypes (Brute Force and SQL Injection) should get replaced with the character "-" (Unicode Decimal Code &#45), as utf-8, the default Pandas library, won't recognize it. Otherwise, the Pandas library won't recognize this character and can fail.
- "Flow Bytes/s", "Flow Packets/s" features have valued to like "Infinity" and "NaN" with numerical values, that are changed to -1 and 0 respectively to make them efficient for various machine learning algorithms.

Table 1. Description of CICIDS2017 Dataset

Recording Day	pcap file size (GB)	Recording Hours	CSV file size (MB)	Attack
Mon	10.0	Full Day	257	-
Tues	10.0	Full Day	166	SSH-Patator, FTP-Patator
Wed	12.0	Full Day	272	DoS Hulk, GoldenEye, Slowloris, Slowhttptest, Heartbleed
Thurs	7.7	Morning	88	Web Attacks
		Afternoon	103	Infiltration
Fri	8.2	Morning	72	Bot
		Afternoon	190	PortScan, DDos

B. Experimental Setup

The technical specifications of the computer used in the implementation:

- CPU: "Intel(R) Core (TM) i5-7200U CPU @ 2.70GHz"
- RAM: 8 GB
- OS: "Windows 10 Pro 64-bit"
- GPU: "Nvidia 940 mx 8gb ddr4"

Jupyter-Notebook 5.7.8 on Anaconda 2019.03 with Python 3.7.3 is used for implementing data pre-processing steps and various machine learning algorithms. Various python libraries are used: pandas, matplotlib, and sklearn. The preprocessed training data set having optimal features yielded by the pre-processing stage are used for the comparison of various machine learning classifiers, with the ensemble learning approach. Machine learning classifiers and ensemble learning classifiers are trained for 1000 epochs.

C. Performance Measures

The results of the algorithms are evaluated and compared according to four measures i.e. accuracy, recall, precision, and F1-Score. These measures have a value between 0 and 1. If the value is close to 1, then the performance of the algorithm used increases, and if the value is close to 0, then it decreases.

- Accuracy – The ratio of correctly predicted instances to a total number of instances in the dataset.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

- Recall (Sensitivity) – The ratio of instances predicted as an attack to all the instances corresponding to attacks.

$$Recall(Sensitivity) = \frac{TP}{TP + FN} \quad (2)$$

- Precision – The ratio of correctly predicted data instances as an attack to all data instances predicted as an attack.

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

- F-measure (F-score/F1-score) – The harmonic-mean of recall/sensitivity and precision. This measure is the most important. Hence, more focus is given on F-measure.

$$F1 - Score = \frac{2 * (Sensitivity) * (Recall)}{Sensitivity + Recall} \quad (4)$$

Where,
TP: Actual attack data predicted as an attack.
FP: Actual no-attack data predicted as an attack.
FN: Actual attack data predicted as no-attack.
TN: Actual no-attack data predicted as no-attack.

V. RESULTS AND DISCUSSION

In this section, the results of the study carried out are presented. The performance evaluation criterion is based on the values of the F1-measure. Histograms are created to visualize the comparison between different features and different machine learning algorithms. A random forest regressor [13] is used for necessary feature selection. First, a simple principal component analysis (PCA) is used for feature selection, but PCA is not efficient when there is an enormous amount of feature in the dataset. The below figure shows the importance of various features.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:59:53 UTC from IEEE Xplore. Restrictions apply.

Table 2. Important features with importance weights computed using random forest [9]

FEATURES	IMPORTANCE WEIGHT ($*10^{(-5)}$)
Standard Backward Packet Length	24656
Flow Byte/s	17876
Forward Packets Total Length	11721
Standard Forward Packet Length	6389
Standard IAT	1017
Minimum IAT	658
Total IAT	499
Flow Duration	413
Maximum Flow IAT	354
Maximum Backward Packet Length	339
Mean IAT	329
Backward Packets Total Length	135
Minimum Forward Packet Length	63
Flow Packet/s	58
Mean Backward Packet Length	53
Mean Forward Packet Length	51
Total Backward Packet/s	23
Total Forward Packet/s	13
Maximum Forward Packet Length	12
Minimum Backward Packet Length	8

Table 2 exhibits the importance weight of the top 20 important features computed using random forest after data pre-processing.

Four machine learning algorithms, ID3 (Decision Tree), MLP, ensemble learning, and Naïve Bayes, are used and compare them on the premise of various performance evaluation criterion i.e. precision, recall, accuracy, and F1-Score. Table 3 exhibits the value of precision, recall, accuracy, and F1-Score of various algorithms. The resulted performance based on F1-measure is as follows (F1-measure takes a value between 0 and 1): Ensemble Learning (NB, ID3 and, MLP): 0.89, ID3: 0.95, MLP: 0.84 and Naive Bayes: 0.82.

Table 3. Performance evaluation data for different algorithms

"ML Algorithm"	Precision	Recall	Accuracy	F1-Score
Naïve Bayes	0.82	0.82	0.8165	0.82
ID3	**0.95**	**0.95**	**0.9515**	**0.95**
Ensemble Learning	0.89	0.89	0.891	0.89
MLP	0.84	0.84	0.8387	0.84

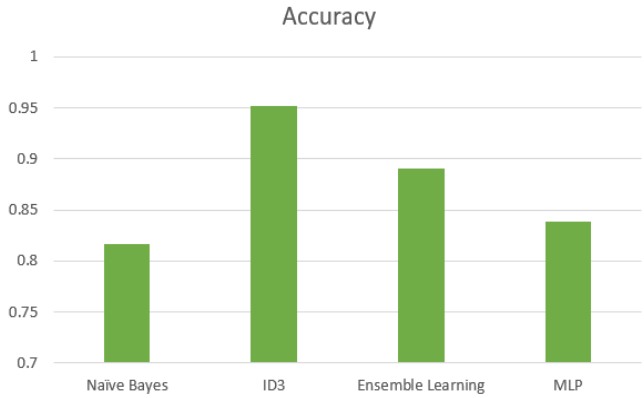


Fig. 1. Comparing Accuracy Parameter of Algorithms

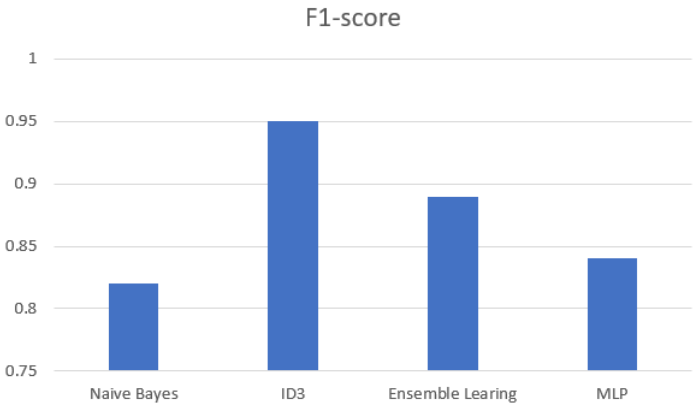


Fig. 2. Comparing F1 score of Algorithms

Fig. 1 and fig. 2 exhibit the accuracy comparison and F1-Score comparison respectively that are obtained after applying various machine learning algorithms. It is evident from fig. 1 and fig. 2 that ID3 and Ensemble Learning perform best out of the various algorithms applied after data pre-processing and feature selection.

Why ID3 come out with the best results?

- ID3 (Iterative Dichotomiser 3) is very intelligible, versatile, and easy to debug. It'll work with classification and regression problems. If you are attempting to predict a categorical value or to predict a continuous value ID3 will handle both problems, but it can handle categorical data efficiently than Naïve Bayes and MLP.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:59:53 UTC from IEEE Xplore. Restrictions apply.

- All Decision Trees is that they solely want a part of data, and they can build a classifier directly from that data. Whereas in other algorithms, choosing features is up to you. There's no way to just toss a part of data, and it picks the most important features that it'll use to classify.
- ID3 is quick regarding computation than other algorithms. ID3 does not produce classifiers using all datasets. It's an iterative method in which a rule is computed and every data record that satisfies the rule is going to be removed and additional rules are generated until all data records aren't covered.

VI. CONCLUSION

The study aims to implement the task for detecting an anomaly in networks by making use of effective machine learning techniques and compare their efficiencies. CICIDS2017 dataset has been used because of its wide attack diversity and up-to-datedness. A network flow is described by the features (over eighty) contained within the dataset. A subset of attributes is chosen from the primitive set of attributes using the Random Forest Regressor technique to make our mind up that which feature is used in machine learning algorithms. Finally, four machine learning algorithms i.e. Naive Bayes, ID3, MLP, and ensemble learning, is practiced on the dataset. The resulted performance based on F1-measure is as follows (F1-measure takes a value between 0 and 1): Ensemble Learning (NB, ID3 and, MLP): 0.89, ID3: 0.95, MLP: 0.84 and Naive Bayes: 0.82. The lower accuracy and huge false alarm rates were also tackled to a great extent. Also, the class imbalance problem was rectified to get better results.

VII. FUTURE SCOPE

In this paper, the primary aim is to implement numerous machine learning algorithms independently of one another for anomaly detection in networks. However, this technique has weak practical applicability in real life. To overcome this drawback, a multi-layered /hierarchical machine learning structure is often designed. Moreover, because of such a structure, it's attainable to save time, CPU power, and memory. There is no "fit-for-all" solution for intrusion detection task [14]. For example, in a two-tiered structure, the primary layer is often created from quick and computationally low-cost algorithms such as Naïve Bayes, therefore network traffic can be observed continuously and at a minimal cost. The primary step, when detecting an anomaly, is to transmit it to an upper layer composed of algorithms with higher performance such as ID3, ensemble learning, etc. The ultimate layer that creates up the determination mechanism takes the preventive decision to protect the network against various attacks. Also, the dataset used, consist of CSV (comma-separated values) files, was obtained from a network of computers created by the Canadian Institute for Cybersecurity. This work is open to future enhancements. Unluckily, in real-time systems, this technique isn't much viable. However, this drawback can be resolved by adding a module that catches real network data and makes it practicable with various machine learning algorithms.

References

[1] R. Sommer, V. Paxson, Outside the Closed World: On Using Machine Learning For Network Intrusion Detection, in: IEEE Symposium on Security and Privacy, IEEE, 2010, pp. 305–316. doi:10.1109/SP.2010.25.

[2] R. Lippmann, J. Haines, D. Fried, J. Korba and K. Das. "The 1999 DARPA off-line intrusion detection evaluation". Computer networks, vol. 34, no. 4, pp. 579 595, 2000. DOI http://dx.doi.org/10.1016/S1389- 1286(00)00139-0.

[3] Nehinbe, J. O. (2011), "A critical evaluation of datasets for investigating idss and ipss researches", In IEEE 10th International Conference on CIS, pages 92–97.

[4] Ali Shiravi, HadiShiravi, M. T. and Ghorbani, A. A. (2012), "Toward developing a systematic approach to generate benchmark datasets for intrusion detection. Computers and Security", 31(3):357 – 374.

[5] Gharib, A., Sharafaldin, I., Habibi Lashkari, A., and Ghorbani, A. A. (2016, "An evaluation framework for intrusion detection dataset", In 2016 International Conference on Information Science and Security (ICISS), pages 1-6.

[6] Abhishek Pharate , Harsha Bhat , Vaibhav Shilimkar "Classification of Intrusion Detection System" IJCS Volume 118 – No. 7, May 2015.

[7] Subaira AS, Anitha P., "Efficient classification mechanism for network intrusion detection system based on data mining techniques: a survey. In international conference on intelligent systems and control", 2014 (pp. 274-80).

[8] S. Ustebay, Z. Turgut, and M. A. Aydin, "Intrusion detection system with recursive feature elimination by using random Forest and deep learning classifier," in 2018 international congress on big data, deep learning and fighting cyber terrorism (IBIGDELFT), 2018, pp.71-76

[9] Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorba-ni, "Toward Generating a New Intrusion Detection Dataset and Intru-sion Traffic Characterization", 4th International Conference on In-formation Systems Security and Privacy (ICISSP), Purtogal, Janu-ary 2018

[10] Search UNB [Internet]. University of New Brunswick est.1785. [cited 2019May26]. Available from: https://www.unb.ca/cic/datasets/ids-2017.html

[11] CICFlowMeter (2017). Canadian institute for cybersecurity (cic).

[12] Panigrahi R, Borah S., "A detailed analysis of CICIDS2017 dataset for designing Intrusion Detection Systems", International Journal of Engineering & Technology 7 (3.24), 2018, 479-482.

[13] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python.

[14] Zhang, C., Pan, X., Li, H., Gardiner, A., Sargent, I., Hare, J., Atkinson, P.M., "A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification", ISPRS Journal of Photogrammetry and Remote Sensing 140 (2018) 133–144,doi:http://dx.doi.org/10.1016/j.isprsjprs.2017.07.014

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:59:53 UTC from IEEE Xplore. Restrictions apply.

The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/1741-0401.htm

Developing a framework for evaluating sustainability index for logistics service providers: graph theory matrix approach

Anchal Gupta
Delhi Technological University, Delhi, India, and
Rajesh Kumar Singh
Department of Operations Management, Management Development Institute, Gurgaon, India

Received 30 December 2019
Revised 30 March 2020
8 April 2020
14 May 2020
17 June 2020
Accepted 17 June 2020

Abstract

Purpose – The purpose of this study is to identify sustainability practices and to propose a framework for evaluating the sustainability index of logistics service providers (LSPs).

Design/methodology/approach – Sustainable practices followed by LSPs are identified through literature review and analysis of a case study. Thirteen such sustainable practices are identified. Thereafter, with expert inputs, nine sustainable practices are shortlisted and considered for the evaluation of the sustainability index in the proposed framework. Graph Theory Matrix Approach has been applied to evaluate the sustainability index of an LSP.

Findings – Major practices identified for evaluating sustainability index include the use of recyclable packaging, use of renewable energy sources, green procurement, reduction in carbon emissions, use of CNG/electric vehicles, rainwater harvesting and so on. The sustainability index of an LSP is evaluated by using the proposed framework.

Practical implications – LSPs can benchmark their sustainability index with respect to the best in the industry. Based on it, LSPs can also identify potential areas for improvement.

Originality/value – Novelty of the study lies in the proposed framework for evaluation of sustainability index which can be used to develop strategies for green logistics. LSPs can also improve their performance in terms of sustainability measures by adopting green logistics.

Keywords Logistics service providers, Sustainable practices, Sustainability index, Graph theory matrix approach, case study

Paper type Research paper

1. Introduction

"Sustainability" – a term first coined by the World Commission on Environment and Development (WCED) in the year 1987 (WCED, 1987) – was adopted gradually by most of the organizations around the world. In the year 2015, the United Nations has set up 17 sustainable development goals (SDGs) for improving sustainable development in business organizations as well as world economies, with a target to achieve these goals by 2030 (Szabo *et al.*, 2016; Salvia *et al.*, 2019; Modgil *et al.*, 2020). Due to this, organizations worldwide are seeking to implement sustainable solutions. With the increasing scarcity of resources as well as increasing needs of society and organizations, there is an urgent need to ensure sufficient availability for current use and conserve for meeting the needs of future generations. Moreover, the environment is continuously deteriorating (Seidel *et al.*, 2013). Therefore, many organizations including logistics service providers (LSPs) are adopting and implementing sustainable practices for optimum use of scarce resources and taking steps for a better global future.

Compared to other industry segments, LSPs contribute more toward environmental degradation in terms of noise and air pollution, carbon emissions, fuel consumption, global warming, resource depletion, and so on (Jazairy and Haartman, 2018). The logistics sector is



International Journal of Productivity and Performance Management
© Emerald Publishing Limited
1741-0401
DOI 10.1108/IJPPM-12-2019-0593

expected to grow exponentially, which will lead to even worse environmental conditions (Wang *et al.*, 2016). Therefore, LSPs are required to initiate sustainable logistics solutions for handling current and upcoming environmental challenges. Previous studies have explored the benefits of the adoption of green practices by various organizations (Perotti *et al.*, 2012; Dubey *et al.*, 2017; Majumdar and Sinha, 2018). Many authors have also discussed the impact of the implementation of sustainable practices on firms' performance (Perotti *et al.*, 2012; Green *et al.*, 2019). Ansari and Kant (2017) explored the positive impact of the sustainable practices of logistics service providers on their efficiency, productivity, and profitability. Many authors have also highlighted the importance of the triple bottom line approach in logistics for improving sustainability (Agrawal *et al.*, 2016; Agrawal and Singh, 2019; Halldorsson, 2019; Kumar and Goswami, 2019).

The existing literature focuses more on green logistics practices in developed countries compared to developing countries (Gandhi *et al.*, 2015; Mangla *et al.*, 2015; Agrawal and Singh, 2019). Gupta and Singh (2020) indicate the need for greater emphasis on developing countries to create a balance between environmental sustainability and the operational excellence of LSPs. Presently, due to environmental consciousness, even consumers in developing nations have started expecting green products. LSPs in developing economies have initiated sustainable practices but only on a small scale. In India, LSPs have started transforming their processes into green processes by changing their conventional fleet into CNG (Compressed Natural Gas) or hybrid fleets for reducing carbon emissions. LSPs have also initiated various activities such as planting trees in the surroundings of their facilities, storing rainwater for washing vehicles and installing solar panels for saving electricity (Gupta *et al.*, 2018). LSPs are also working toward a reduction in waste, which is created in different forms such as underutilized resources, replication of tasks, and inadequate planning of resource allocation (Qian *et al.*, 2019). Moreover, LSPs are exploring alternative ways of reducing carbon emissions, fuel consumption, and noise levels for long-term sustainability. Despite the measures being adopted by LSPs, the actual adoption rate is very low due to lack of continuous evaluation measures. Owing to the lack of appropriate frameworks in the literature, most LSPs struggle in evaluating their sustainability index (Chhabra and Singh, 2016). There is, therefore, a need to perform an in-depth analysis of sustainable practices followed by LSPs and to evaluate the sustainability index of such providers in a developing country. This study analyzes and evaluates the sustainability performance of LSPs in the Indian context. The main objectives of this study are as follows:

(1) To identify the sustainable practices followed by LSPs

(2) To analyze the sustainable practices specifically for LSPs in the Indian context

(3) To evaluate the sustainability index using a Graph Theory Matrix Approach (GTMA)

(4) To propose a framework for evaluating the sustainability index of LSPs

This study is further elaborated under the following sections. Section 2 comprises the literature review on green initiatives and sustainable practices followed by LSPs. In Section 3, the case study method is used to identify and analyze sustainable practices in developing economies. In Section 4, the sustainability index has been evaluated by using GTMA. The implications of the study along with future scope are discussed in Section 5. The conclusions of the study are concluded in Section 6.

2. Literature review

There is a significant body of literature available on evolution, adoption, and implementation of sustainable practices in both the manufacturing and service sectors (Perotti *et al.*, 2012;

Anwer *et al.*, 2018; Majumdar and Sinha, 2018; Mensah and Casadevall, 2019). Sustainability means the conservation of natural resources and judicious use of resources for long term business continuity. Due to continuous environmental degradation and scarcity of resources, resource sustainability has become an important concern for all organizations including LSPs (Coaffee, 2008; Zhang *et al.*, 2017). LSPs have started redefining their conventional processes into green processes as a contribution towards society and the environment (Ferreira *et al.*, 2015). Previous studies discussed the adoption and importance of sustainability in the field of logistics (Gandhi *et al.*, 2015; Mangla *et al.*, 2017). Centobelli *et al.* (2017) conducted a systematic literature review on green and sustainable practices followed by LSPs by exploring more than 415 articles from 1960 to 2014. Evangelista *et al.* (2018) conducted a comprehensive literature review on the environmental sustainability of LSPs for the period of 2000–2016.

Researchers have not only conducted comprehensive literature reviews on environmental sustainability but also discussed the benefits that organizations can obtain by implementing sustainable practices. Various authors captured different aspects of logistics and linked them with sustainability measures. For example, Wilding *et al.* (2012) highlighted the importance of supplier assessment and collaboration in improving sustainability. Martinez-Jurado *et al.* (2014) explored the relationship between lean management, supply chain management, and sustainability. Marques (2019) conducted a detailed and systematic literature review on sustainable supply network management from a knowledge perspective. Many practitioners have already proved the strong relationship among the implementation of sustainable practices, an increase in profitability, and improvement in organizational performance (Perotti *et al.*, 2012; Singh *et al.*, 2016; Ansari and Kant, 2017). Several researchers have established the relationship among implementation of green sustainable practices and organizational culture, service quality, and operational performance (Zhu *et al.*, 2008; Bask *et al.*, 2018; Garcia-Dastugue; Eroglu, 2019; Green, 2019; San-Ong *et al.*, 2019).

Despite the extensive literature on logistics and sustainability, LSPs, specifically those in developing countries, are yet to implement green practices in all their processes (Agrawal and Singh, 2019). As a result, it is imperative for LSPs in emerging economies to practice resource conservation and ensure long-term business continuity (Singh and Agrawal, 2018). Moreover, organizations, including LSPs, are expected to follow strict guidelines for environmental protection, issued by the government and the United Nations (UN). The UN has set sustainable development goals (SDGs-7, 12 and 13) for the implementation of sustainable practices in the organizations (The Sustainable Development Goals Report, 2019). The UN has also emphasized the development of metrics and indicators for measuring sustainable development at the environmental, social, and economic levels. On the other end, consumers are also becoming proactive and demanding eco-friendly products (Green *et al.*, 2015). These requirements by the government, the UN, and customers are building pressure on LSPs.

Many authors have discussed the hurdles in the adoption of sustainable practices and highlighted methods for measuring the performance of sustainable supply chains (Taticchi *et al.*, 2013; Baird, 2017; Thanki; Thakkar, 2018). Pazirandeh and Jafari (2013) critically evaluated the impact of sustainability strategies on logistics efficiency using structural equation modeling (SEM). Tajbakhsh and Hassini (2015) proposed a performance measurement framework for measuring the sustainability of various industries based on seven sustainability dimensions (economic, environmental, social, reputable, valuable, equitable, and sustainable). Trianni *et al.* (2019) analyzed 26 small and medium manufacturing organizations of Germany and Italy and observed that the organizations accorded more importance to economic indicators over environmental and social indicators. They developed a novel framework for measuring sustainability indicators. Gardas *et al.* (2019) analyzed sustainable supply chain determinants for measuring the business performance of the oil and gas supply chain. Chowdhury and Paul (2020) reviewed the

applications of multi-criteria decision-making methods (MCDM) for measuring performance, including enablers of, and barriers to, corporate sustainability.

Most LSPs struggle in evaluating their sustainability index due to the lack of appropriate frameworks (Chhabra and Singh, 2016). It is crucial to measure the performance of LSPs in terms of their sustainability. This study proposes a framework for measuring the sustainability index of LSPs. This will help other LSPs understand their current position and take corrective measures if required. This provides a significant opportunity to conduct an in-depth analysis of the sustainable practices followed by LSPs in developing countries and to evaluate their sustainability index to understand the current situation of LSPs in the Indian market. Therefore, the following research questions have been developed:

RQ1. What are the sustainable practices followed by LSPs in developing countries to fulfill the UN's sustainable development goals?

RQ2. How can the sustainability of Indian LSPs be measured?

RQ3. What suggestions can be made to other LSPs for better achievement of SDG goals?

The sustainable practices followed by LSPs are identified through a literature review as discussed in the next section and a case study is also conducted on Indian LSP to understand their sustainable practices closely as discussed in Section 3.

2.1 Identification of sustainable practices from literature review

Previous studies have explored sustainable practices adopted by LSPs specifically in context to developing countries. It has been observed that LSPs have emphasized on the reduction of carbon emissions and fuel consumption as an initiative towards the environmental protection (Yao *et al.*, 2015; Aldakhil *et al.*, 2018; Liu *et al.*, 2018). Agrawal *et al.* (2016) have explored the importance of reverse logistics to assist return management and to make optimum use of resources. Renewable resources like solar and hydro products are more preferred nowadays over expensive single time use products for long term sustainability of products (Dubey *et al.*, 2017; Sureeyatanapas *et al.*, 2018). Reuse and recycling of packaging material and pallets are also done through plastic totes and heavy containers in order to save resources (Boubeta *et al.*, 2018). The adoption of green practices in the logistics sector has also become evident by converting manual processes (paper-work) into e-processes (paper-less) through digitalization (Kayikci, 2018). The following sustainable practices are identified through the literature review:

(1) *Use of reusable or recyclable packaging:* To promote green practices, several LSPs have started using plastic totes and heavy-duty containers which are cost-efficient and can be used for multiple trips (Centobelli *et al.*, 2018; Regattieri *et al.*, 2019). Paper, cardboard, corn starch, and bubble wrap, etc. are biodegradable and preferred for packaging (Gupta *et al.*, 2020). In India, Transport Corporation of India (TCI) and Agarwal Packers and Movers Ltd. are actively reusing and recycling packaging materials for making cost-effective and sustainable logistics system (Bhushan and Malviya, 2018)

(2) *Use of renewable energy sources:* The best way to conserve limited resources is to maximize the use of renewable energy sources. LSPs have initiated the use of solar energy and hydropower for many logistical activities instead of single-time use electric ones. Installation of solar panels for producing electricity (Fontes, and Freires, 2018; Gupta *et al.*, 2018), transparent roofs for natural light to enter during daytime (Tascioglu and Keser, 2019) and storage of rainwater for gardening and cleaning vehicles (Piecyk and Björklund, 2015; Gupta *et al.*, 2018) indicate the prominent use of natural resources by LSPs. All Cargo Logistics sets a wonderful example by providing renewable energy transportation from end to end logistics (Choudhary, 2019).

(3) *Green procurement:* Green procurement is the foremost step towards green logistics (Teixeira *et al.*, 2018). LSPs prefer to purchase products or services, which will cause a minimum adverse effect on the environment. The use of alternative fuel, proper planning for unutilized material, waste reduction, and use of recyclable and reusable products can support green procurement initiatives (Meehan and Bryde, 2011; Centobelli *et al.*, 2018). DHL has adopted green logistics to make our planet more sustainable by reducing emissions and moving towards the circular economy.

(4) *Reduction in carbon emissions:* LSPs have taken initiatives for low carbon supply chain integrations. This initiative has a positive impact on LSPs' environmental and financial performance (Qian *et al.*, 2019). LSPs are required to take appropriate solutions for carbon emission reduction and to develop strategies for a sustainable environment (Guiffrida *et al.*, 2011; Tang *et al.*, 2015; Aldakhil *et al.*, 2018; Liu *et al.*, 2018). Damco, DHL and TNT are offering green services and reducing carbon footprints through various services (Das and Kulkarni, 2018).

(5) *Use of CNG/Electric fleet:* Many LSPs are making use of CNG vehicles to reduce cost, pollution, and fuel consumption (Gupta *et al.*, 2018). The use of CNG/electric vehicles (fuel-free) helps in reducing the noise level and making the air clean. LSPs are also looking forward to adopting e-vehicles in the coming years (McKinnon, 2010; Liu *et al.*, 2018). Almost all Indian LSPs including Safe Express, CTC freight carriers Ltd. and Sal Logistics Pvt. Ltd., etc. are using more of the CNG vehicles to save cost and fuel.

(6) *Resource optimization:* Due to the scarcity of resources, LSPs are looking forward to the alternative options to use limited resources in an optimum way (Centobelli *et al.*, 2017). Conservation of resources will help in the smooth functioning of logistics operations in the future. LSPs are optimizing resources by making maximum use of natural resources over man-made products (Choudhury *et al.*, 2018; He *et al.*, 2018; Winkler and Zinsmeister, 2019). Om Logistics Ltd. conserves its natural resources by making maximum use of solar light and rainwater harvesting in their warehouses.

(7) *Reduction in fuel consumption:* Less fuel consumption by LSPs is considered as economic, social and environmental initiative (Blanco and Sheffi, 2017). Automobile companies are adding fuel-saving features in new vehicles by using new engine technologies. Moreover, the transition from fuel to electric vehicles is a great initiative for reducing fuel consumption and environmental pollution (Tang *et al.*, 2015; Yao *et al.*, 2015). Although, in India, there are many infrastructural bottlenecks to implement e-vehicles entirely, but the government is working in this direction. Jayem Logistics has significantly contributed towards eco-friendly transportation, reduction in fuel consumption, and pollution.

3. Study of sustainable practices in context to an Indian LSP

Most of the identified sustainable practices from the literature review are in context to both developing and developed countries. There is, therefore, a need to understand and identify the sustainable practices of Indian LSPs more closely to make this study more relevant in the Indian context. A real case illustration on Indian LSP has been discussed to have a thorough understanding of sustainable practices followed by them. The information is collected based on multiple visits and discussions with middle and senior management of the case organization.

3.1 Profile of case organization (ABC ltd.)

ABC Ltd. (given name) started its business as a multi-modal transportation organization in the mid-1980s in India. Its objective is to provide logistics support to the manufacturing organizations in the nation as well as across boundaries. With an annual turnover of around INR 1,500 crores (the US$ 225.47 million), it serves all the significant businesses sectors such as automobile, hardware, telecom, retail and pharmaceuticals.

The organization has a wide distribution network of more than 3,500 client base of domestic, transnational, and international presence majorly in the corporate B2B sphere. It provides service to most of the automobile manufacturers and more than 2000 auto segment part producers in India. Their customer base includes all significant brand names like Mahindra and Mahindra, Maruti Suzuki India Ltd., SUZUKI Motors, India Yamaha Motors Pvt. Ltd., Toyota, Daimler, Honda Motors India, Heromoto Corp, Tata Motors, Samsung Electronics India Limited, VIVO Mobiles India, Bharati Airtel, Cadila Pharmaceuticals, Arvind Fashions, etc. ABC Ltd. is conferred with many awards, recognitions, and excellence awards for its outstanding contribution to the logistics sector both in domestic and international services.

3.2 Organizational assets

ABC Ltd. ensures timely delivery and assured services to the clients through its technologically advanced assets. The organization offers modernized multi-client and dedicated warehousing spaces over more than 100 locations with its captive warehousing strength of more than 20 lac square meters in PAN India. It manages hub and spoke distribution through more than 40 Cross-Dock centers and 700 branches for Part Truck Load (PTL) distribution. This Indian LSP delivers more than 14,000 pin codes pan India and is serving over 85 countries through its own offices, joint ventures, exclusive partners, etc.

The case organization offers door to door customized deliveries through more than 4,500 modern technologically advanced, all-weather containerized vehicles of varied capacities ranging from 0.5 metric tons to 40 metric tons. It has more than 5,000 trained workforce, who work sincerely 24 by 7 for committed services, and customer query resolution. ABC Ltd. is equipped with advanced in-house created applications and SAP software along with ORACLE based databases for providing timely and accurate information to all stakeholders. It also uses modern advanced information technology tools and real-time applications to track and trace consignments through owned servers.

3.3 Organizational processes

The organization's order processing is made effective with invoice creation on behalf of their customers, while also creating and managing receipts for storage and dispatch of shipments, and sales tax or excise duties. It also incorporates key parameters like time and motion studies for optimization of resources, designing of warehouse layout, 5S support, and adaptable stock piling arrangements by using the latest and coordinated IT tools for the smooth functioning of warehouses. For managing inventory, the organization prefers to arrange their stocks by using concepts of LIFO (Last in First Out), FIFO (First in First Out), etc. This LSP also encourages the usage of mathematical models along with IT tools for analyzing the inventory costs, transportation costs, and stock management.

An ideal mix of roadways, railways, and the airways are offered by the organization to handle customer consignments and to serve them as per commitments. It offers flexibility to the customers in terms of selection of cargo, selection of location, the type of product and the size of shipments. This LSP also provides value-added services which include services like labeling and kitting of goods, packaging with bubble wrap or directly in carton boxes, and

sorting of goods. Document maintenance, packaging redesigning, consultancy on supply chain optimization, and return logistics are also provided.

3.4 Sustainability measures

In order to conserve natural resources, ABC Ltd. is carrying out numerous activities to manage resources at the ecological, societal and economic levels. The organization is also certified with ISO 14001:2015 certifications for maintaining green practices for sustainability.

Some of the sustainable practices that it follows are as under:

(1) *Reduction in carbon emissions:* All the trucks are outfitted with GPS technology which helps in tracking the exact location, while also enabling reduction in carbon emissions. The organization takes exceptional consideration for vehicles which have valid pollution check and carbon emission certificates.

(2) *Green practices:* ABC Ltd. plants more than 5,000 trees annually to enhance green cover around their warehouses. There is a culture of planting tree saplings on various occasions as a step towards environmental conservation. Air sanitizing plants are also grown around warehouses or distribution centers to purify the air and make surroundings clean.

(3) *Digitalization of processes:* The case organization is adopting process automation, in order to change its manual transactions to digital processes. This initiative not just reduces the use of paper and makes document handling easier but also brings transparency to the system.

(4) *Use of solar energy:* To optimize resources, ABC Ltd. has adopted proper conservation of natural resources. Additionally, the organization has installed solar panels on the roofs of most of the warehouses to generate power and to give back to society by saving electricity.

(5) *Use of natural/Renewable resources:* With preference to use natural light, the organization has built transparent roofs at their facilities. This reduces carbon emission and helps in saving electricity.

(6) *Rainwater harvesting:* Committed to green practices, ABC Ltd. utilizes rainwater stored in its underground tanks for further use in washing vehicles, cleaning the floors, or gardening purposes.

(7) *Education initiatives:* In line with PMKVY (Prime Minister Kaushal Vikas Yojna, Government of India scheme), the organization is running its non-profit education and training institute to increase employability of individuals in the logistics industry. Under this initiative, it offers various short term courses related to logistics and supply chain along with practical training sessions. The meritorious candidates are offered to work with the organization according to their abilities and domain knowledge.

(8) *Healthy working environment:* Offering employee friendly policies and environment, this LSP provides its employees with flexible working hours. In order to provide a comfortable and sound environment, residential and mess services are provided to the employees.

3.5 Selection of sustainable practices

After identifying thirteen sustainable practices from the literature review and case analysis, a dedicated team of three experts with over ten years of experience and working with leading

Indian LSPs, helped in finalizing the sustainable practices in Indian context. In order to achieve this, a list of all identified sustainable practices was shared with experts and their inputs were solicited on application of these practices in their respective organizations and ensuing impact on performance. The level of application of these sustainable practices in their organizations helped in understanding the importance of sustainability in Indian context. After expert inputs, nine sustainable practices were finalized. Out of these nine, five were identified from literature, one was from the case study and three were found to be common in both case study and literature review. The final list of sustainable practices after discussing with Indian experts is given in Table 1. As the objective of this study is to evaluate the sustainability index, therefore, the GTMA is applied for evaluating the sustainability index. The detailed research methodology will be discussed in the next section.

4. Evaluation of sustainability index- Graph Theory Matrix Approach (GTMA)

After identifying sustainability practices, a framework for the evaluation of the sustainability index of an LSP is proposed in this study. The Graph Theory Matrix Approach has been applied for the evaluation of the sustainability index, which considers any number of quantitative and qualitative sustainable practices simultaneously. The conventional methods like cause and effect diagram, process flow charts and block diagrams are not

Table 1. Sustainable practices followed by logistics service providers

S. No	Sustainable practices	Description	References	Source
1	Use of recyclable packaging (*S*1)	Using recyclable and reusable raw material for packaging to promote green practices	Gupta *et al.* (2020), Regattieri *et al.* (2019), Centobelli *et al.* (2018), Jilani *et al.* (2018), Lew *et al.* (2018), Sureeyatanapas *et al.* (2018), Chhabra *et al.* (2017)	Literature review
2	Use of renewable energy sources (*S*2)	Making use of renewable resources for long term sustainability	Taşçıoglu and Keser (2019), Fontes, and Freires (2018), Gupta *et al.* (2018), Piecyk and Björklund (2015), McKinnon (2010)	Literature review and Case study
3	Green procurement (*S*3)	Using green practices for purchasing	Centobelli *et al.* (2018), Teixeira *et al.* (2018), Meehan and Bryde (2011)	Literature review
4	Reduction in carbon emissions (*S*4)	Reducing carbon emission to protect the environment	Aldakhil *et al.* (2018), Centobelli *et al.* (2018), Gupta *et al.* (2018), Lew *et al.* (2018), Tang *et al.* (2015), Guiffrida *et al.* (2011)	Literature review and Case study
5	Use of CNG/ Electric/ hybrid fleet (*S*5)	Using Eco-friendly fleet to reduce the pollution and consumption of fuel	Liu *et al.* (2018), Gupta *et al.* (2018), McKinnon (2010)	Literature review
6	Resource optimization (*S*6)	Optimum use of available resources for long term business continuity	Winkler and Zinsmeister (2019), He *et al.* (2018), Choudhury *et al.* (2018), Centobelli *et al.* (2017)	Literature review
7	Digitization of processes (*S*7)	Adoption of paperless processes and use of latest technologies	Singh *et al.* (2019)	Literature review and Case study
8	Rainwater harvesting (*S*8)	Using rainwater for gardening or washing vehicles	Gupta *et al.* (2018)	Case study
9	Reduction in fuel consumption (*S*9)	Reducing the consumption of fuel for conserving resources	Blanco and Sheffi (2017), Yao *et al.* (2015), Tang *et al.* (2015)	Literature review

suitable for further mathematical analysis whereas this method establishes the directional relationship among variables and also take care of their interdependencies by mathematical analysis (Raj *et al.*, 2010; Jain and Raj, 2016; Hudnurkar; Ambekar, 2019).

It is a more preferred approach over conventional methods for representation and quantification. Digraph assists in the graphical representation of variables, which supports better visual analysis and provides valuable inputs for mathematical modeling. Many researchers have applied the GTMA in different fields. Franceschini *et al.* (2006) have used this approach for finding the quality index. Darvish *et al.* (2009) applied GTMA for the ranking of contractors. Attri *et al.* (2014) used this method to compute the intensity of TPM (Total Productive Maintenance) barriers. Apart from manufacturing, many authors have applied this approach in diversified areas such as to model the various mechanisms required for supply chain coordination (Kaur *et al.*, 2006), to evaluate the technology-based self-service banking service quality (Sindhwani and Goel, 2015), to analyze disposition decisions in reverse logistics (Agrawal *et al.*, 2016), to find the agility index of supply chains (Singh *et al.*, 2017), for quantifying the barriers in the implementation of technologies (Bhandari *et al.*, 2019) and to measure the flexibility index for supply chains (Singh and Kumar, 2019), etc.

It is found to be a very important decision-making tool that helps in creating associations among variables and evaluating by forming an index. Based on the literature review and case study, nine sustainable practices have been identified in this study as shown in Table 1. This approach includes various steps of developing a digraph between variables depending upon their mutual correlations. The research methodology framework is shown in Figure 1. Further, a permanent function is computed by developing a permanent matrix by using Equation (1).

4.1 Variables digraph

The sustainable practices adopted by LSPs are represented in the form of nodes and edges by developing a variable digraph as shown in Figure 2. The nodes represent the sustainable practices and the edges represent their linkages. S_i shows the inheritance of practices and s_{ij} shows the degree of dependence of jth practice on ith practice. It also indicates the directed edge from node i to node j. The linkages among all sustainable practices taken in this study are shown in Figure 2. The digraph for detailed analysis has been discussed in Figure 3.

The permanent function of sustainable practices (Perm S) shown in Figure 3 is written as:

$$\text{Perm } S = \begin{vmatrix} S1 & s_{12} & s_{13} & s_{14} & s_{15} & s_{16} & s_{17} & s_{18} & s_{19} \\ s_{21} & S2 & s_{23} & s_{24} & s_{25} & s_{26} & s_{27} & s_{28} & s_{29} \\ s_{31} & s_{32} & S3 & s_{34} & s_{35} & s_{36} & s_{37} & s_{38} & s_{39} \\ s_{41} & s_{42} & s_{43} & S4 & s_{45} & s_{46} & s_{47} & s_{48} & s_{49} \\ s_{51} & s_{52} & s_{53} & s_{54} & S5 & s_{56} & s_{57} & s_{58} & s_{59} \\ s_{61} & s_{62} & s_{63} & s_{64} & s_{65} & S6 & s_{67} & s_{68} & s_{69} \\ s_{71} & s_{72} & s_{73} & s_{74} & s_{75} & s_{76} & S7 & s_{78} & s_{79} \\ s_{81} & s_{82} & s_{83} & s_{84} & s_{85} & s_{86} & s_{87} & S8 & s_{89} \\ s_{91} & s_{92} & s_{93} & s_{94} & s_{95} & s_{96} & s_{97} & s_{98} & S9 \end{vmatrix}$$

4.2 Quantification of S_is and s_{ij}s

To find out the value of the permanent function (Perm S), there is a need to quantify S_is and s_{ij}s. The qualitative measure of inheritance of practices (S_i) is rated on a scale of 1–5 by experts by considering 1 to be the lowest and 5 to be the highest. The qualitative measures of interdependencies among sustainable practices (s_{ij}) have been quantitatively measured by finding the relative dependence on each other based on the value assigned by experts and

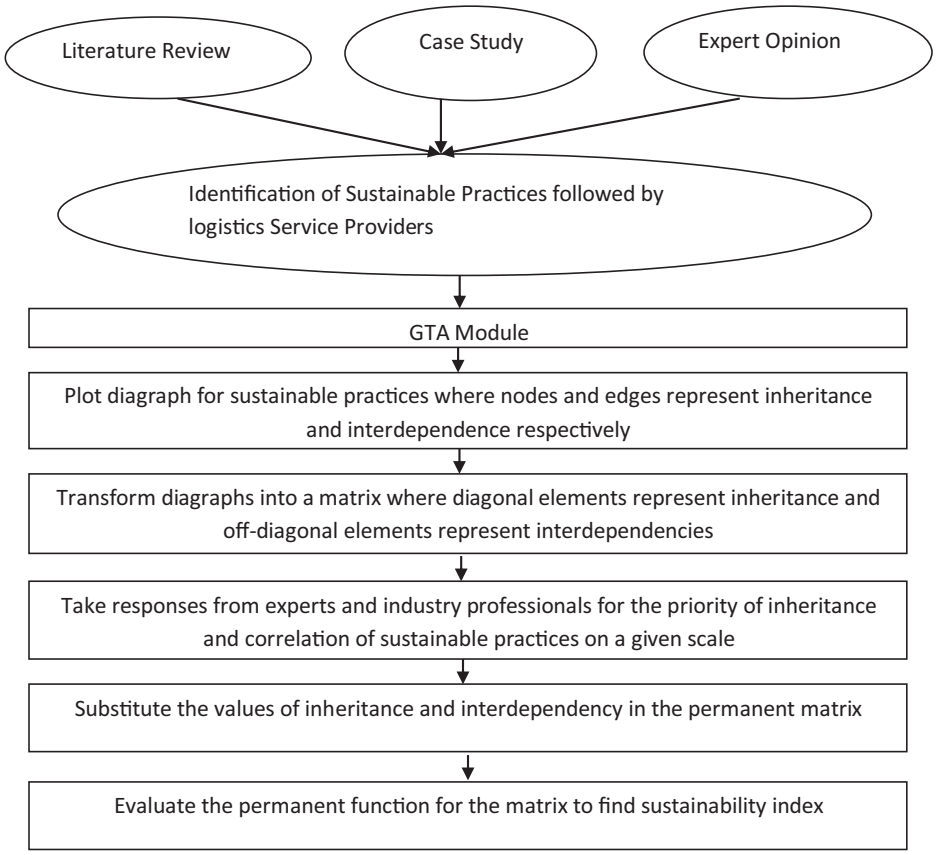


Figure 1.
Research methodology framework

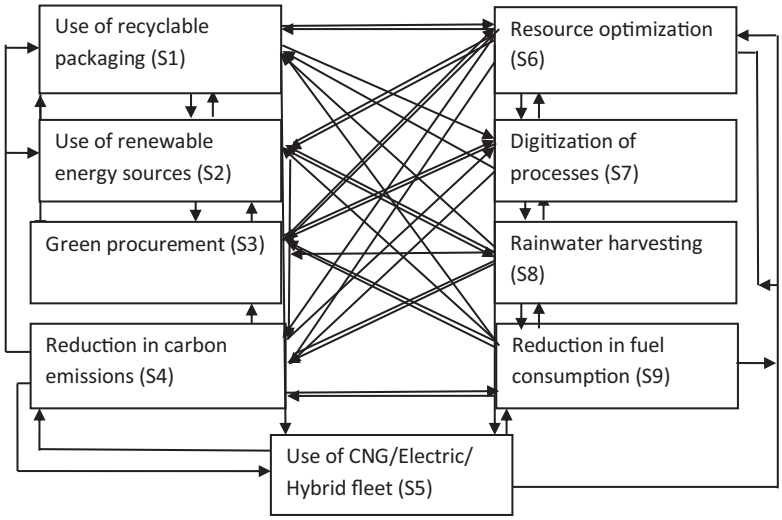


Figure 2.
Interaction among sustainable practices followed by LSPs

discussed in Table 2. A structured questionnaire was designed and sent to all concerned. The experts have been asked to respond on the importance given to sustainable practices (*S*1 to *S*9) and also to rate the relative dependence of one sustainable practice on another according to criteria given in Table 2. For instance, s_{12}- dependence of sustainable practice 2 on sustainable practice 1 which means the degree of dependence of use of energy renewable resources (*S*2) on use of recyclable packaging (*S*1) and similarly, the relative dependence will find for all other possible combinations.

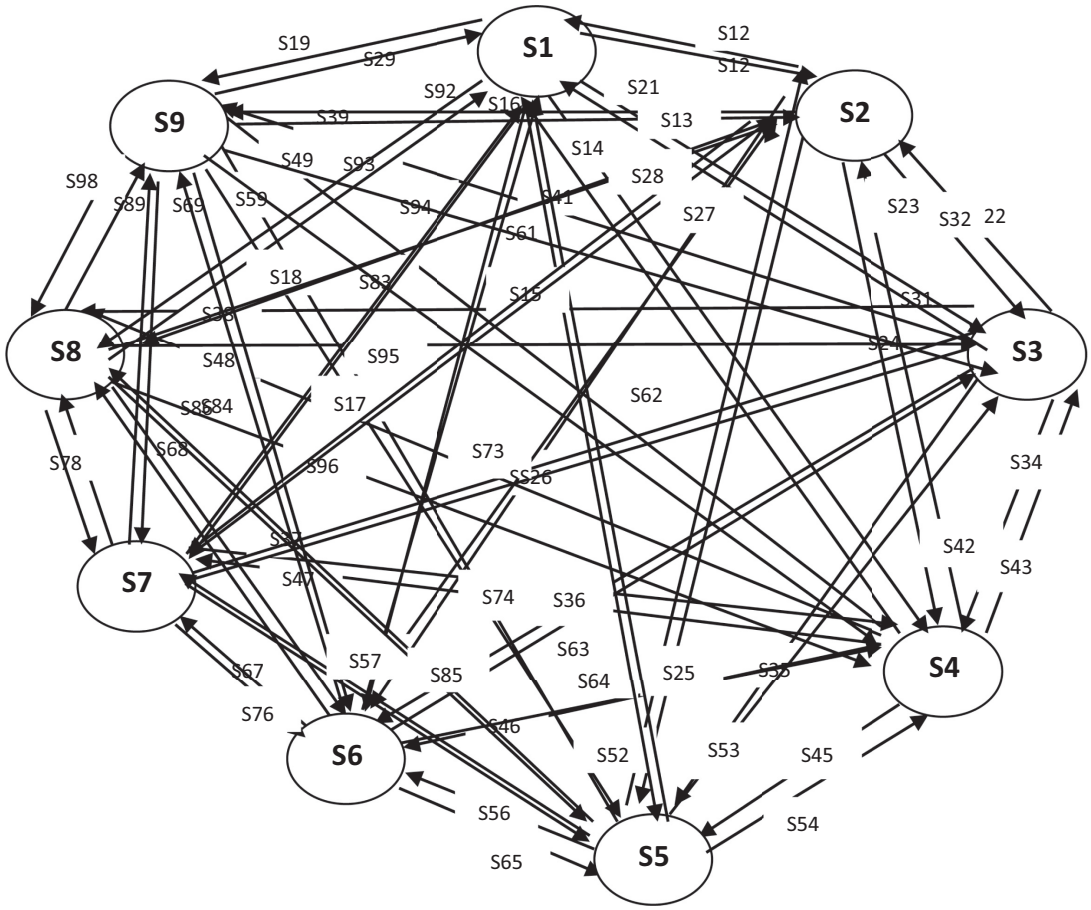


Figure 3.
Digraph representation of elements of permanent matrix

Qualitative measures of interdependencies	Relative dependence of sustainable practices s_{ij}	s_{ji} = 10-sij
Two attributes are of equal importance	5	5
One attribute is slightly dependent on the other	6	4
One attribute is very dependent on the other	7	3
One attribute is most dependent on the other	8	2
One attribute is extremely dependent on the other	9	1
One attribute is exceptionally dependent on the other	10	0

Source(s): Muduli *et al.* (2013)

Table 2.
Relative dependence of sustainable practices (s_{ij}s)

4.3 Variables' permanent matrix

The permanent function is a mathematical expression which helps to determine the sustainability index of LSPs. The calculation for the permanent function is almost the same as finding the determinant of the matrix but considers only positive terms. The value of the permanent function of sustainable practices (Perm *S*) is computed by using Equation (1) which is given as follows.

$$
\begin{aligned}
perm S = {} & \prod_{i=1}^{9} Si + \sum_{i=1}^{8}\sum_{j=i+1}^{9}\sum_{k=j+1}^{3}\sum_{l=k+1}^{4}\sum_{m=l+1}^{5}\sum_{n=m+1}^{6}\sum_{o=n+1}^{7}\sum_{p=o+1}^{8}\sum_{q=p+1}^{9}(s_{ij}s_{ji})S_kS_lS_mS_nS_nS_pS_q \\
& + \sum_{i=1}^{7}\sum_{j=1+1}^{8}\sum_{k=j+1}^{9}\sum_{l=k+1}^{4}\sum_{m=l+1}^{5}\sum_{n=m+1}^{6}\sum_{o=n+1}^{7}\sum_{p=o+1}^{8}\sum_{q=p+1}^{9}(s_{ij}s_{jk}s_{ki} + s_{ik}s_{kj}s_{ji})S_lS_mS_nS_oS_pS_q + \\
& \left[\begin{array}{l}
\sum_{i=1}^{6}\sum_{j=1+1}^{9}\sum_{k=j+1}^{8}\sum_{l=i+2}^{9}\sum_{m=1}^{8}\sum_{n=m+1}^{9}\sum_{o=n+1}^{9}\sum_{p=o+1}^{9}\sum_{q=p+1}^{9}(s_{ij}s_{jk})(s_{kl}s_{lk})S_mS_nS_oS_pS_q \\
+\sum_{i=1}^{6}\sum_{j=i+1}^{8}\sum_{k=j+1}^{9}\sum_{l=j+1}^{9}\sum_{m=1}^{8}\sum_{n=m+1}^{9}\sum_{n=m+1}^{9}\sum_{o=n+1}^{9}\sum_{p=o+1}^{9}\sum_{q=p+1}^{9}(s_{ij}s_{jk}s_{kj}s_{li} + s_{il}s_{lk}s_{kj}s_{ij})S_mS_nS_oS_pS_q
\end{array}\right] \\
& + \left[\begin{array}{l}
\sum_{i=1}^{7}\sum_{j=1+1}^{8}\sum_{k=j+1}^{9}\sum_{l=1}^{8}\sum_{m=l+1}^{9}\sum_{n=m+1}^{9}\sum_{o=n+1}^{9}\sum_{p=o+1}^{9}\sum_{q=p+1}^{9}(s_{ij}s_{jk}s_{ki} + s_{ik}s_{kj}s_{ji})(s_{lm}s_{ml})S_nS_oS_pS_q \\
+\sum_{i=1}^{5}\sum_{j=1+1}^{8}\sum_{k=i+1}^{9}\sum_{l=j+1}^{9}\sum_{m=1}^{8}\sum_{n=m+1}^{9}\sum_{o=n+1}^{9}\sum_{p=o+1}^{9}\sum_{q=p+1}^{9}(s_{ij}s_{jk}s_{kl}s_{lm}s_{mi} + s_{im}s_{ml}s_{lk}s_{kj}s_{ij})S_nS_oS_pS_q
\end{array}\right] \\
& + \left[\begin{array}{l}
\sum_{i=1}^{6}\sum_{j=1}^{8}\sum_{k=i+1}^{9}\sum_{l=j+1}^{9}\sum_{m=1}^{8}\sum_{n=m+1}^{9}\sum_{o=n+1}^{9}\sum_{p=o+1}^{9}\sum_{q=p+1}^{9}(s_{ij}s_{jk}s_{kl}s_{li} + s_{ik}s_{lk}s_{lk}s_{kj}s_{ji})(s_{mn}s_{nm})S_oS_pS_q \\
+\sum_{i=1}^{4}\sum_{j=i+1}^{8}\sum_{k=i+1}^{9}\sum_{l=j+1}^{9}\sum_{m=1}^{8}\sum_{n=m+1}^{9}\sum_{o=n+1}^{9}\sum_{p=o+1}^{9}\sum_{q=p+1}^{9}(s_{ij}s_{jk}s_{ki} + s_{ik}s_{kj}s_{ji})(s_{lm}s_{mn}s_{nl} + s_{ln}s_{nm}s_{ml})S_oS_pS_q \\
+\sum_{i=1}^{4}\sum_{j=i+1}^{9}\sum_{k=j+1}^{6}\sum_{l=i+1}^{9}\sum_{m=k+1}^{8}\sum_{n=k+2}^{9}\sum_{o=n+1}^{9}\sum_{p=n+2}^{9}\sum_{q=p+1}^{9}(s_{ij}s_{ji})(s_{kl}s_{lk})(s_{mn}s_{nm})S_oS_pS_q \\
+\sum_{i=1}^{4}\sum_{j=i+1}^{8}\sum_{k=j+1}^{9}\sum_{l=i+1}^{9}\sum_{m=i+1}^{9}\sum_{n=j+1}^{9}\sum_{o=j+1}^{9}\sum_{p=j+1}^{9}\sum_{q=j+1}^{9}(s_{ij}s_{ji})(s_{ij}s_{jk}s_{kl}s_{lm}s_{mn}s_{ni} + s_{in}s_{nm}s_{lk}s_{kj}s_{ji})(s_{mn}s_{nm})S_oS_pS_q + \ldots
\end{array}\right]
\end{aligned}
\tag{1}
$$

Substituting the values of the variables in the permanent matrix and compute the value of the permanent matrix (Perm S) by putting values in Equation (1):

$$
\text{Perm } S = \begin{vmatrix}
S1 & s_{12} & s_{13} & s_{14} & s_{15} & s_{16} & s_{17} & s_{18} & s_{19} \\
s_{21} & S2 & s_{23} & s_{24} & s_{25} & s_{26} & s_{27} & s_{28} & s_{29} \\
s_{31} & s_{32} & S3 & s_{34} & s_{35} & s_{36} & s_{37} & s_{38} & s_{39} \\
s_{41} & s_{42} & s_{43} & S4 & s_{45} & s_{46} & s_{47} & s_{48} & s_{49} \\
s_{51} & s_{52} & s_{53} & s_{54} & S5 & s_{56} & s_{57} & s_{58} & s_{59} \\
s_{61} & s_{62} & s_{63} & s_{64} & s_{65} & S6 & s_{67} & s_{68} & s_{69} \\
s_{71} & s_{72} & s_{73} & s_{74} & s_{75} & s_{76} & S7 & s_{78} & s_{79} \\
s_{81} & s_{82} & s_{83} & s_{84} & s_{85} & s_{86} & s_{87} & S8 & s_{89} \\
s_{91} & s_{92} & s_{93} & s_{94} & s_{95} & s_{96} & s_{97} & s_{98} & S9
\end{vmatrix}
$$

$$
= \begin{vmatrix}
4 & 5 & 9 & 8 & 5 & 9 & 5 & 5 & 9 \\
5 & 3 & 7 & 9 & 6 & 9 & 5 & 8 & 7 \\
1 & 3 & 2 & 9 & 9 & 7 & 9 & 6 & 8 \\
2 & 1 & 1 & 2 & 9 & 8 & 8 & 7 & 9 \\
5 & 4 & 1 & 1 & 4 & 7 & 5 & 5 & 9 \\
1 & 1 & 3 & 2 & 3 & 4 & 9 & 9 & 9 \\
5 & 5 & 1 & 2 & 5 & 1 & 2 & 5 & 5 \\
5 & 2 & 4 & 3 & 5 & 1 & 5 & 2 & 6 \\
1 & 3 & 2 & 1 & 1 & 1 & 5 & 4 & 3
\end{vmatrix}
$$

Perm $S^* = 2,30,595,214,680$

The best possible value of the index can be computed by taking the highest possible value of all variables. The value of inheritance of all sustainable practices is kept maximum, that is, 5 and the diagonal cells of the permanent matrix are kept 5 as maximum. This maximum index can be used as a reference for benchmarking of sustainability performance. Substituting all values in Equation (1) and the maximum theoretical index can be:

$$\text{Perm } S = \begin{vmatrix} 5 & 5 & 9 & 8 & 5 & 9 & 5 & 5 & 9 \\ 5 & 5 & 7 & 9 & 6 & 9 & 5 & 8 & 7 \\ 1 & 3 & 5 & 9 & 9 & 7 & 9 & 6 & 8 \\ 2 & 1 & 1 & 5 & 9 & 8 & 8 & 7 & 9 \\ 5 & 4 & 1 & 1 & 4 & 7 & 5 & 5 & 9 \\ 1 & 1 & 3 & 2 & 3 & 5 & 9 & 9 & 9 \\ 5 & 5 & 1 & 2 & 5 & 1 & 5 & 5 & 5 \\ 5 & 2 & 4 & 3 & 5 & 1 & 5 & 5 & 6 \\ 1 & 3 & 2 & 1 & 1 & 1 & 5 & 4 & 5 \end{vmatrix}$$

Perm $S^* = S^*_{\max} = 3,63,706,047,040$

S^* represents the sustainability index of LSPs. This value of sustainability index can be also used for comparison with other LSPs. It can help them in positioning themselves in terms of the adoption and implementation of sustainable practices. LSPs can analyze the importance and relationships among sustainable practices and can be implemented in their respective organizations for improving sustainability. The maximum value $S^*_{\max}$ represents the maximum index of sustainability that LSPs can achieve by adopting and implementing maximum sustainable measures.

5. Discussion

In this study, the GTMA is used to compute the sustainability index of LSPs. When the resources are limited and the environment is degrading at a very fast pace in the current market situation, the evaluation of sustainability index can help LSPs in finding suitable solutions to adopt and implement sustainable practices in their inbound and outbound processes. It can help LSPs to devise some strategies for improving the adoption and implementation of sustainable practices, in turn, positively impacting the performance of LSPs. The implementation of sustainable practices by LSPs will also help in reducing wastage of resources which further encourages the circular economy.

From the above case illustration, it has been observed that the sustainability index (S) of the given LSP is comparatively lesser than the best possible value of the sustainability index (S^*). This reflects that there is a scope of improvement for an LSP by implementing more sustainable practices in their logistical operations. LSPs can develop strategies for revision in their processes and can plan for improvement in terms of sustainability. On similar lines, Mangla *et al.* (2019) explored the agro-food supply chain by analyzing four Indian dairy-based organizations and developed strategies for improving the supply chain and logistics performance in emerging economies using the GTMA approach.

The sustainability index can also assist LSPs in positioning themselves and comparing them with their peer service providers. LSPs can develop their strategies for improving their performance through increased use of green practices. Several articles on the implementation of sustainable practices have discussed their importance in developing countries. For the successful implementation of sustainable practices, LSPs should come forward to identify the issues related to sustainability implementation and should formulate a proper strategy for handling them effectively. Similarly, Agrawal *et al.* (2016) evaluated the outsourcing index, which is very useful for managers and practitioners in formulating strategies for outsourcing reverse logistics decisions. In Bangladesh, Moktadir *et al.* (2018) applied graph theory for

identifying barriers for implementing sustainable manufacturing practices in the leather industry. They have also suggested that the managers should formulate strategies for optimum utilization of limited resources to reduce wastage. It also validates our findings in context to Indian LSP.

5.1 Managerial implications

The major implication of this study is that LSPs should select factors for evaluating the sustainability index depending upon the prevailing business environment. These factors may vary as per region and sector. Based on the sustainability index, LSPs may choose to find potential areas for improvement and a green logistics strategy can be developed accordingly. Organizations with low sustainability index can identify their weak areas and can work upon them to increase their sustainability index. Organizations with high sustainability index can compare their values with maximum possible value and can further work in the direction to achieve that level. This framework will be helpful for LSPs in benchmarking their performance with global standards in the era of the circular economy.

5.2 Limitations and future scope

In this study, a case analysis has been done for understanding the sustainable practices of LSPs. Findings from a case may vary with change in sector and location. Therefore, for the generalization of these findings, this study can be further validated by taking more case studies on LSPs in different contexts. This work can be extended for small LSPs as the future scope of the existing work. Apart from considered criteria, other sustainability criteria in context to an emerging circular economy can be also considered for further modification of framework. Further, the results can be compared after evaluating the sustainability index by different methods to check the robustness of the framework. The study may be further extended to comparing the sustainability index of logistics companies from developing and developed countries. It will help in the identification and adoption of advanced sustainability practices by LSPs of developing countries for continuous improvement. The study will not only encourage the organizations but also the society to conserve resources and make proper utilization of available resources.

6. Conclusion

In today's business environment of the circular economy, almost every organization is taking initiatives for reducing environmental degradation and increasing the sustainable use of resources. LSPs are continuously exploring innovative and sustainable ways to contribute to society and the environment. Sustainable practices adopted and implemented by various LSPs have been identified from previous studies. The case study method is also adopted to make the current study more relevant and realistic for LSPs. A real case study on an Indian LSP has been presented to clarify their processes and adoption of sustainable practices in their logistical activities. The information is based on observations and discussions with middle and senior management during multiple visits to the case organization.

This study makes two major contributions. The first is the identification of major sustainable practices in the logistics sector. The second is a framework for the evaluation of the sustainability index. The use of recyclable packaging, renewable energy sources, green procurement, carbon emissions reduction, use of CNG fleets, resource optimization, digitalization of processes, rainwater harvesting and reduction in fuel consumption are the major sustainable practices. Although previous studies have explored many performance measurement frameworks, these frameworks are not so relevant in the era of the circular economy. This study included prominent sustainability practices while evaluating the

sustainability index. We have used GTMA approach for evaluation of the sustainability index of LSPs. The application of this framework is illustrated using the real case of Indian LSP. The sustainability index can also help LSPs in formulating strategies to achieve the UN's sustainability goals. It will motivate LSPs to implement green logistics practices in their operations. This framework will also be helpful for LSPs in comparing their performance against global standards.

References

Agrawal, S. and Singh, R.K. (2019), "Analyzing disposition decisions for sustainable reverse logistics: triple Bottom Line approach", *Resources, Conservation and Recycling*, Vol. 150, p. 104448.

Agrawal, S., Singh, R.K. and Murtaza, Q. (2016), "Outsourcing decisions in reverse logistics: sustainable balanced scorecard and graph theoretic approach", *Resources, Conservation and Recycling*, Vol. 108, pp. 41-53.

Aldakhil, A.M., Nassani, A.A., Awan, U., Abro, M.M.Q. and Zaman, K. (2018), "Determinants of green logistics in BRICS countries: an integrated supply chain model for green business", *Journal of Cleaner Production*, Vol. 195, pp. 861-868.

Ansari, Z.N. and Kant, R. (2017), "Exploring the framework development status for sustainability in supply chain management: a systematic literature synthesis and future research directions", *Business Strategy and the Environment*, Vol. 26 No. 7, pp. 873-892.

Anwer, A.S., Garza-Reyes, J.A. and Kumar, V. (2018), "Best supply chain management practices and high-performance firms", *International Journal of Productivity and Performance Management*, Vol. 67 No. 9, pp. 1482-1509.

Attri, R., Grover, S. and Dev, N. (2014), "A graph theoretic approach to evaluate the intensity of barriers in the implementation of total productive maintenance (TPM)", *International Journal of Production Research*, Vol. 52 No. 10, pp. 3032-3051.

Baird, K. (2017), "The effectiveness of strategic performance measurement systems", *International Journal of Productivity and Performance Management*, Vol. 66 No. 1, pp. 3-21.

Bask, A., Rajahonka, M., Laari, S., Solakivi, T., Töyli, J. and Ojala, L. (2018), "Environmental sustainability in shipper-LSP relationships", *Journal of Cleaner Production*, Vol. 172, pp. 2986-2998.

Bhandari, D., Singh, R.K. and Garg, S.K. (2019), "Prioritization and evaluation of barriers intensity for implementation of cleaner technologies: framework for sustainable production", *Resources, Conservation and Recycling*, Vol. 146, pp. 156-167.

Bhushan, R. and Malviya, S. (2018), "Packaging solutions: reduce, recycle & don't reuse", available at: https://economictimes.indiatimes.com/industry/indlgoods/svs/packaging/reduce-recycle-don treuse-future-as-clear-as-glass-on-packaging-solutions/articleshow/64848893.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst (accessed on March 12 2020).

Blanco, E.E. and Sheffi, Y. (2017), *Green Logistics in Sustainable Supply Chains*, Springer, Cham, pp. 147-187.

Boubeta, I.G., Vázquez, M.F., Caamaño, P.D. and Prado, J.C.P. (2018), "Economic and environmental packaging sustainability: a case study", *Journal of Industrial Engineering and Management*, Vol. 11 No. 2, pp. 229-238.

Centobelli, P., Cerchione, R. and Esposito, E. (2017), "Developing the WH2 framework for environmental sustainability in logistics service providers: a taxonomy of green initiatives", *Journal of Cleaner Production*, Vol. 165, pp. 1063-1077.

Centobelli, P., Cerchione, R. and Esposito, E. (2018), "Environmental sustainability and energy-efficient supply chain management: a review of research trends and proposed guidelines", *Energies*, Vol. 11 No. 2, p. 275.

Chhabra, D. and Singh, R.K. (2016), "Justification of green logistic practices in Indian enterprises using analytic hierarchy process", *International Journal of Logistics Systems and Management*, Vol. 25 No. 3, pp. 295-312.

Chhabra, D., Garg, S.K. and Singh, R.K. (2017), "Analyzing alternatives for green logistics in an Indian automotive organization: a case study", *Journal of Cleaner Production*, Vol. 167, pp. 962-969.

Choudhary, S. (2019), "Renewable energy equipment transportation: an opportunity for end-to- end logistics", available at: https://energy.economictimes.indiatimes.com/energy-speak/renewable-energy-equipment-transportation-an-opportunity-for-end-to-end-logistics/3758 (accessed March 15 2020).

Choudhury, N., Raut, R.D., Gardas, B.B., Kharat, M.G. and Ichake, S. (2018), "Evaluation and selection of third party logistics services providers using data envelopment analysis: a sustainable approach", *International Journal of Business Excellence*, Vol. 14 No. 4, pp. 427-453.

Chowdhury, P. and Paul, S.K. (2020), "Applications of MCDM methods in research on corporate sustainability", *Management of Environmental Quality: An International Journal*, Vol. 31 No. 2, pp. 385-405.

Coaffee, J. (2008), "Risk, resilience, and environmentally sustainable cities", *Energy Policy*, Vol. 36 No. 12, pp. 4633-4638.

Darvish, M., Yasaei, M. and Saeedi, A. (2009), "Application of the graph theory *and matrix m*ethods to contractor ranking", *International Journal of Project Management*, Vol. 27 No. 6, pp. 610-619.

Das, M. and Kulkarni, V. (2018), "How green is your supply chain?", available at: https://www.thehindubusinessline.com/economy/logistics/how-green-is-your-supply-chain/article20318230.ece1 (accessed March 13 2020).

Dubey, R., Gunasekaran, A., Papadopoulos, T., Childe, S.J., Shibin, K.T. and Wamba, S.F. (2017), "Sustainable supply chain management: framework and further research directions", *Journal of Cleaner Production*, Vol. 142, pp. 1119-1130.

Evangelista, P., Santoro, L. and Thomas, A. (2018), "Environmental sustainability in third-party logistics service providers: a systematic literature review from 2000–2016", *Sustainability*, Vol. 10 No. 5, p. 1627.

Ferreira, L., Putnik, G.D., Lopes, N., Lopes, A. and Cruz-Cunha, M.M. (2015), "A cloud and ubiquitous architecture for effective environmental sensing and monitoring", *Procedia Computer Science*, Vol. 64, pp. 1256-1262.

Fontes, C.H.D.O. and Freires, F.G.M. (2018), "Sustainable and renewable energy supply chain: a system dynamics overview", *Renewable and Sustainable Energy Reviews*, Vol. 82, pp. 247-259.

Franceschini, F., Galetto, M., Grover, S., Agrawal, V.P. and Khan, I.A. (2006), "Role of human factors in TQM: a graph theoretic approach", *Benchmarking: An International Journal*, Vol. 13 No. 4, pp. 447-468.

Gandhi, S., Mangla, S.K., Kumar, P. and Kumar, D. (2015), "Evaluating factors in implementation of successful green supply chain management using DEMATEL: a case study", *International Strategic Management Review*, Vol. 3 No. 1, pp. 96-109.

García-Dastugue, S. and Eroglu, C. (2019), "Operating performance effects of service quality and environmental sustainability capabilities in logistics", *Journal of Supply Chain Management*, Vol. 55 No. 3, pp. 68-87.

Gardas, B.B., Raut, R.D. and Narkhede, B. (2019), "Determinants of sustainable supply chain management: a case study from the oil and gas supply chain", *Sustainable Production and Consumption*, Vol. 17, pp. 241-253.

Green, K.W., Toms, L.C. and Clark, J. (2015), "Impact of market orientation on environmental sustainability strategy", *Management Research Review*, Vol. 38 No. 2, pp. 217-238.

Green, K.W., Inman, R.A., Sower, V.E. and Zelbst, P.J. (2019), "Impact of JIT, TQM and green supply chain practices on environmental sustainability", *Journal of Manufacturing Technology Management*, Vol. 30 No. 1, pp. 26-47.

Guiffrida, A.L., Datta, P., Dey, A., LaGuardia, P. and Srinivasan, M. (2011), "Building sustainability in logistics operations: a research agenda", *Management Research Review*, Vol. 34 No. 11, pp. 1237-1259.

Gupta, A. and Singh, R.K. (2020), "Managing operations by a logistics company for sustainable service quality: Indian perspective", *Management of Environmental Quality: An International Journal*, Vol. ahead-of-print. No. ahead-of-print, doi: 10.1108/MEQ-11-2019-0246.

Gupta, A., Singh, R.K. and Suri, P.K. (2018), "Prioritizing critical success factors for sustainable service quality management by logistics service providers", *Vision*, Vol. 22 No. 3, pp. 295-305.

Gupta, A., Singh, R.K. and Suri, P.K. (2020), "Prioritizing best practices for logistics service providers", in Suri, P. and Yadav, R. (Eds), *Transforming Organizations through Flexible Systems Management*, Springer, Singapore, pp. 257-275.

Halldorsson, A. (2019), "Actionable sustainability in supply chains", in Zijm, H., Klumpp, M., Regattieri, A. and Heragu, S. (Eds), *Operations, Logistics and Supply Chain Management*, Springer, Cham, pp. 191-203.

He, M., Shen, J., Wu, X. and Luo, J. (2018), "Logistics space: a literature review from the sustainability perspective", *Sustainability*, Vol. 10 No. 8, p. 2815.

Hudnurkar, M. and Ambekar, S.S. (2019), "Framework for measurement of supplier satisfaction", *International Journal of Productivity and Performance Management*, Vol. 68 No. 8, pp. 1475-1492.

Jain, V. and Raj, T. (2016), "Modeling and analysis of FMS performance variables by ISM, SEM and GTMA approach", *International Journal of Production Economics*, Vol. 171, pp. 84-96.

Jazairy, A. and Haartman, R. (2018), "Analysing the institutional pressures on shippers and logistics service providers to implement green logistics practices", *International Journal of Logistics Research and Applications*, pp. 1-41.

Jilani, A., Ali, Y. and Khan, M.W. (2018), "Greening of humanitarian supply chain with focus on logistics", *International Journal of Business Performance and Supply Chain Modelling*, Vol. 10 No. 1, pp. 49-66.

Kaur, A., Kanda, A. and Deshmukh, S.G. (2006), "A graph theoretic approach for supply chain coordination", *International Journal of Logistics Systems and Management*, Vol. 2 No. 4, pp. 321-341.

Kayikci, Y. (2018), "Sustainability impact of digitization in logistics", *Procedia Manufacturing*, Vol. 21, pp. 782-789.

Kumar, G. and Goswami, M. (2019), "Sustainable supply chain performance, its practice and impact on barriers to collaboration", *International Journal of Productivity and Performance Management*, Vol. 68 No. 8, pp. 1434-1456.

Lew, A.F.R., Chew, B.C. and Hamid, S.R. (2018), "Green logistics implementation factors: a study on a global logistics provider", *Journal of Advanced Manufacturing Technology*, Vol. 12 No. 1, pp. 115-128.

Liu, X., McKinnon, A.C. and Wei, N. (2018), "An analysis of energy-related CO 2 emissions from China's logistics industry", in Liu, X. (Ed.), *Environmental Sustainability in Asian Logistics and Supply Chains*, Springer, Singapore, pp. 3-19.

Majumdar, A. and Sinha, S. (2018), "Modeling the barriers of green supply chain management in small and medium enterprises", *Management of Environmental Quality: An International Journal*, Vol. 29 No. 6, pp. 1110-1122.

Mangla, S.K., Kumar, P. and Barua, M.K. (2015), "Risk analysis in green supply chain using fuzzy AHP approach: a case study", *Resources, Conservation and Recycling*, Vol. 104, pp. 375-390.

Mangla, S.K., Govindan, K. and Luthra, S. (2017), "Prioritizing the barriers to achieve sustainable consumption and production trends in supply chains using fuzzy Analytical Hierarchy Process", *Journal of Cleaner Production*, Vol. 151, pp. 509-525.

Mangla, S.K., Sharma, Y.K., Patil, P.P., Yadav, G. and Xu, J. (2019), "Logistics and distribution challenges to managing operations for corporate sustainability: study on leading Indian diary organizations", *Journal of Cleaner Production*, Vol. 238, p. 117620.

Marques, L. (2019), "Sustainable supply network management: a systematic literature review from a knowledge perspective", *International Journal of Productivity and Performance Management*, Vol. 68 No. 6, pp. 1164-1190.

Martinez-Jurado, P.J. and Moyano-Fuentes, J. (2014), "Lean management, supply chain management and sustainability: a literature review", *Journal of Cleaner Production*, Vol. 85, pp. 134-150.

McKinnon, A. (2010), "Green logistics: the carbon agenda", *Electronic Scientific Journal of Logistics*, Vol. 6 No. 3.

Meehan, J. and Bryde, D. (2011), "Sustainable procurement practice", *Business Strategy and the Environment*, Vol. 20 No. 2, pp. 94-106.

Mensah, J. and Casadevall, S. (2019), "Sustainable development: meaning, history, principles, pillars, and implications for human action: literature review", *Cogent Social Sciences*, Vol. 5 No. 1, p. 1653531.

Modgil, S., Gupta, S. and Bhushan, B. (2020), "Building a living economy through modern information decision support systems and UN sustainable development goals", *Production Planning and Control*, pp. 1-21.

Moktadir, M.A., Ali, S.M., Rajesh, R. and Paul, S.K. (2018), "Modeling the interrelationships among barriers to sustainable supply chain management in leather industry", *Journal of Cleaner Production*, Vol. 181, pp. 631-651.

Muduli, K., Govindan, K., Barve, A. and Geng, Y. (2013), "Barriers to green supply chain management in Indian mining industries: a graph theoretic approach", *Journal of Cleaner Production*, Vol. 47, pp. 335-344.

Pazirandeh, A. and Jafari, H. (2013), "Making sense of green logistics", *International Journal of Productivity and Performance Management*, Vol. 62 No. 8, pp. 889-904.

Perotti, S., Zorzini, M., Cagno, E. and Micheli, G.J. (2012), "Green supply chain practices and company performance: the case of 3PLs in Italy", *International Journal of Physical Distribution and Logistics Management*, Vol. 42 No. 7, pp. 640-672.

Piecyk, M.I. and Björklund, M. (2015), "Logistics service providers and corporate social responsibility: sustainability reporting in the logistics industry", *International Journal of Physical Distribution and Logistics Management*, Vol. 45 No. 5, pp. 459-485.

Qian, C., Wang, S., Liu, X. and Zhang, X. (2019), "Low-carbon initiatives of logistics service providers: the perspective of supply chain integration", *Sustainability*, Vol. 11 No. 12, p. 3233.

Raj, T., Shankar, R., Suhaib, M. and Khan, R.A. (2010), "A graph-theoretic approach to evaluate the intensity of barriers in the implementation of FMSs", *International Journal of Services and Operations Management*, Vol. 7 No. 1, pp. 24-52.

Regattieri, A., Santarelli, G. and Piana, F. (2019), "Packaging logistics", in Zijm, H., Klumpp, M., Regattieri, A. and Heragu, S. (Eds), *Operations, Logistics and Supply Chain Management*, Springer, Cham, pp. 273-303.

Salvia, A.L., Leal Filho, W., Brandli, L.L. and Griebeler, J.S. (2019), "Assessing research trends related to Sustainable DevelopmentGoals: local and global issues", *Journal of Cleaner Production*, Vol. 208, pp. 841-849.

San Ong, T., Magsi, H.B. and Burgess, T.F. (2019), "Organizational culture, environmental management control systems, environmental performance of Pakistani manufacturing industry", *International Journal of Productivity and Performance Management*, Vol. 68 No. 7, pp. 1293-1322.

Seidel, S., Recker, J. and VomBrocke, J. (2013), "Sense making and sustainable practicing: functional affordances of information systems in green transformations", *MIS Quarterly*, pp. 1275-1299.

Singh, R.K. and Agrawal, S. (2018), "Analyzing disposition strategies in reverse supply chains: fuzzy TOPSIS approach", *Management of Environmental Quality: An International Journal*, Vol. 29 No. 3, pp. 427-443.

Singh, R.K. and Kumar, P. (2019), "Measuring the flexibility index for a supply chain using graph theory matrix approach", *Journal of Global Operations and Strategic Sourcing*, Vol. 13 No. 1, pp. 56-69.

Singh, R.K., Rastogi, S. and Aggarwal, M. (2016), "Analyzing the factors for implementation of green supply chain management", *Competitiveness Review*, Vol. 26 No. 3, pp. 246-264.

Singh, R.K., Koul, S. and Kumar, P. (2017), "Analyzing the interaction of factors for flexibility in supply chains", *Journal of Modeling in Management*, Vol. 12 No. 4, pp. 671-689.

Singh, R.K., Luthra, S., Mangla, S.K. and Uniyal, S. (2019), "Applications of information and communication technology for sustainable growth of SMEs in India food industry", *Resources, Conservation and Recycling*, Vol. 147, pp. 10-18.

Sindwani, R. and Goel, M. (2015), "Technology-based self-service banking service quality evaluation: a Graph Theoretic Approach", *International Journal of Advanced Science and Technology*, Vol. 80, pp. 1-18.

Sureeyatanapas, P., Poophiukhok, P. and Pathumnakul, S. (2018), "Green initiatives for logistics service providers: an investigation of antecedent factors and the contributions to corporate goals", *Journal of Cleaner Production*, Vol. 191, pp. 1-14.

Szabo, S., Nicholls, R.J., Neumann, B., Renaud, F.G., Matthews, Z., Sebesvari, Z., AghaKouchak, A., Bales, R., Ruktanonchai, C.W., Kloos, J. and Foufoula-Georgiou, E. (2016), "Making SDGs work for climate change hotspots", *Environment: Science and Policy for Sustainable Development*, Vol. 58 No. 6, pp. 24-33.

Tajbakhsh, A. and Hassini, E. (2015), "Performance measurement of sustainable supply chains: a review and research questions", *International Journal of Productivity and Performance Management*, Vol. 64 No. 6, pp. 744-783.

Tang, S., Wang, W., Yan, H. and Hao, G. (2015), "Low carbon logistics: reducing shipment frequency to cut carbon emissions", *International Journal of Production Economics*, Vol. 164, pp. 339-350.

Tascioglu, A. and Keser, H.Y. (2019), "Solar energy in the logistics sector: assessments on Turkey", *Journal of Business and Social Review in Emerging Economies*, Vol. 5 No. 2, pp. 225-236.

Taticchi, P., Tonelli, F. and Pasqualino, R. (2013), "Performance measurement of sustainable supply chains: a literature review and a research agenda", *International Journal of Productivity and Performance Management*, Vol. 62 No. 8, pp. 782-804.

Teixeira, C.R.B., Assumpcao, A.L., Correa, A.L., Savi, A.F. and Prates, G.A. (2018), "The contribution of green logistics and sustainable purchasing for green supply chain management", *Independent Journal of Management and Production*, Vol. 9 No. 3, pp. 1002-1026.

Thanki, S. and Thakkar, J. (2018), "A quantitative framework for lean and green assessment of supply chain performance", *International Journal of Productivity and Performance Management*, Vol. 67 No. 2, pp. 366-400.

The Sustainable Development Goals Report (2019), available at: https://www.un.org/sustainabledevelopment/sustainable-development-goals/ (accessed 30 December 2019).

Trianni, A., Cagno, E., Neri, A. and Howard, M. (2019), "Measuring industrial sustainability performance: empirical evidence from Italian and German manufacturing small and medium enterprises", *Journal of Cleaner Production*, Vol. 229, pp. 1355-1376.

Wang, X., Persson, G. and Huemer, L. (2016), "Logistics service providers and value creation through collaboration: a case study", *Long Range Planning*, Vol. 49 No. 1, pp. 117-128.

WCED, S.W.S. (1987), "World commission on environment and development", *Our common future*, Vol. 17, pp. 1-91.

Wilding, R., Wagner, B., Gimenez, C. and Tachizawa, E.M. (2012), "Extending sustainability to suppliers: a systematic literature review", *Supply Chain Management: An International Journal*, Vol. 17 No. 5, pp. 531-543.

Winkler, H. and Zinsmeister, L. (2019), "Trends in digitalization of intralogistics and the critical success factors of its implementation", *Brazilian Journal of Operations and Production Management*, Vol. 16 No. 3, pp. 537-549.

Yao, E., Lang, Z., Yang, Y. and Zhang, Y. (2015), "Vehicle routing problem solution considering minimising the fuel consumption", *IET Intelligent Transport Systems*, Vol. 9 No. 5, pp. 523-529.

Zhang, W., Shi, Y., Yang, M., Gu, X., Tang, R. and Pan, X. (2017), "Ecosystem evolution mechanism of manufacturing service system driven by service providers", *International Journal of Production Research*, Vol. 55 No. 12, pp. 3542-3558.

Zhu, Q., Sarkis, J. and Lai, K.H. (2008), "Confirmation of a measurement model for green supply chain management practices implementation", *International Journal of Production Economics*, Vol. 111 No. 2, pp. 261-273.

Further reading

Ansari, Z.N., Kant, R. and Shankar, R. (2019), "Prioritizing the performance outcomes due to adoption of critical success factors of supply chain remanufacturing", *Journal of Cleaner Production*, Vol. 212, pp. 779-799.

Liu, X., Ma, S., Tian, J., Jia, N. and Li, G. (2015), "A system dynamics approach to scenario analysis for urban passenger transport energy consumption and CO2 emissions: a case study of Beijing", *Energy Policy*, Vol. 85, pp. 253-270.

Sarkis, J., Helms, M.M. and Hervani, A.A. (2010), "Reverse logistics and social sustainability", *Corporate Social Responsibility and Environmental Management*, Vol. 17 No. 6, pp. 337-354.

About the authors

Anchal Gupta is doing PhD in Management from Delhi School of Management, Delhi Technological University, Delhi, India. Her interest is in Supply Chain Management, Logistics and Service Quality Management. Anchal Gupta is the corresponding author and can be contacted at: anchal.gupta1503@gmail.com

Rajesh Kumar Singh is Professor in Operations Management area at Management Development Institute, Gurugram, India. His areas of interest include Supply Chain Management, Small Business Management and Quality Management. He has published around 125 research papers in reputed international/national journals and conferences. His papers have appeared in journals such as *Production Planning and Control*, *Journal of Industrial and Production Engg*, *Industrial management and Data Systems*, *Singapore Management Review*, *International Journal of Productivity and Performance Management*, *Benchmarking: An International Journal*, *Journal of Modelling in Management*, *Competitiveness Review*, *Journal of Advances in Management Research*, *International Journals of Productivity and Quality Management* etc.

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Disease Detection in Coffee Plants Using Convolutional Neural Network

Manoj Kumar
Computer Engineering Dept.
Delhi Technological University
New Delhi, India
Email: mkumarg@dce.ac.in

Pranav Gupta
Computer Engineering Dept.
Delhi Technological University
New Delhi, India
Email: 7pranavgupta@gmail.com

Puneet Madhav
Computer Engineering Dept.
Delhi Technological University
New Delhi, India
Email: puneet.m98@gmail.com

Sachin
Computer Engineering Dept.
Delhi Technological University
New Delhi, India
Email: sachinrathore454@gmail.com

Abstract— Every year, the crop output of India is significantly affected due to delayed detection of diseases in crops. This research is a contribution to the farmers in their battle against coffee plants diseases. It will help in timely detection of diseases, resulting in increased coffee production output of India. Many coffee plant diseases like Leaf Rust, Cercospora Spots have clear visual symptoms and thus can be extracted and their classification can be done. Convolutional Neural Networks (CNNs) has proved its efficiency and accuracy in the field of image classification and pattern recognition. Hence it can act as a powerful tool in the diagnosis of coffee leaves diseases since these symptoms have clearly distinguishable patterns. Thus, a Convolution Neural Network model is proposed which utilizes the technique of Transfer Learning, reducing the training time taken by the model significantly. Further, to achieve a higher success rate, Data Augmentation technique is applied to enlarge the dataset used to train the network. The proposed model has achieved a high accuracy of 97.61%.

Keywords— Convolutional Neural Network, Coffee Plant Disease, Data Augmentation, Transfer Learning.

I. INTRODUCTION

Coffee is one of the most popular beverages and an important commercial product which is cultivated in more than 50 countries across the globe. Every day an estimated 2.25 billion coffee cups are consumed throughout the world.[1] India is among the top producers of coffee globally and has the finest quality of coffee grown in the shade rather than direct sunlight, anywhere in the world[2]. With an estimated 319,500 metric ton of coffee produced in the year 2018-19[3], of which almost 80% is exported throughout the globe. The amount of revenue generated from the coffee industry is projected to be at USD 987m in the year 2020 and its market is projected to increase annually at 7.4%. [4]

Coffee plants are highly vulnerable to various pests and diseases. Crop loss due to such diseases in coffee plants is estimated at 20-30% of the country's total production[5]. Chemical control's success on any disease depends on the disease pathogenesis stage at which the fungicide is applied [1]. Inexperienced and careless usage of pesticide may result in increased resistance of the pathogens in the long run, severely affecting the capability of plants to fight. Thus, the detection of plant diseases in an accurate and timely manner is an essential aspect of precision agriculture [2].

Soft computing can play a crucial role in the accurate and timely detection of such biotic stresses that attack the leaf of a coffee plant. Various techniques such as genetic algorithms [3], machine learning, use of computer vision and image processing [4], expert systems have been an area of active research to detect such diseases [5]. These methods have not been widely adopted due to low efficiency and success rate. In [6], the authors have used both image processing and artificial neural network to achieve a success rate of 90%. Some plant diseases do not have any conspicuous visual symptoms, however most common diseases found in coffee leaves have clear visual symptoms as shown in the Fig. 1. For example, the symptoms of coffee leaf rust (CLR) disease are oily spots on the upper surface of coffee leaf which are minute and yellowish in color, and expands further into a large circular spot that with time, turn to bright orange to red and at last, to brown with a yellow border[6].

Deep Learning have completely dominated the field of image classification for the last few years, thus the proposed Convolution Neural Network is used to detect coffee leaf diseases and categorize them into 5 classes – Healthy, Diseased Leaves with Cercospora spots, Phoma, Coffee Leaf Rust(CLR), and Leaf Miner with a high success rate.

[1] https://en.wikipedia.org/wiki/Economics_of_coffee
[2] https://en.wikipedia.org/wiki/Coffee_production_in_India
[3] https://www.indiacoffee.org/coffee-statistics.html
[4] https://www.statista.com/outlook/30010000/119/coffee/india#market-globalrevenue
[5] https://economictimes.indiatimes.com/news/economy/agriculture/india-to-get-594349-to-control-leaf-rust-in-coffee/articleshow/2466942.cms
[6] https://www.britannica.com/science/coffee-rust

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:48:41 UTC from IEEE Xplore. Restrictions apply.

The subsequent sections in the paper are organized in the following order: Section II describes the dataset used for training the model in detail. Section III explains the architectural components like Inception V3 model, Data augmentation technique and working of the proposed model. Section IV describes the various tests and experiments performed. In section V, the result is presented. Section VI contains a conclusion along with the future work.

II. DATASET

The pre-processed dataset [7] contains a wide variety of images of coffee leaves for disease detection. The pre-processing of data has been done to eliminate unwanted details and enhance the symptomatic features.

The images contain both disease-free and diseased leaves, affected by different types of biotic stresses. Dataset is divided in two parts – leaf dataset consisting of original images of the leaves and other one containing cropped images of symptoms. Fig. 1 shows the division of the dataset- disease categories with the corresponding number of leaf images.

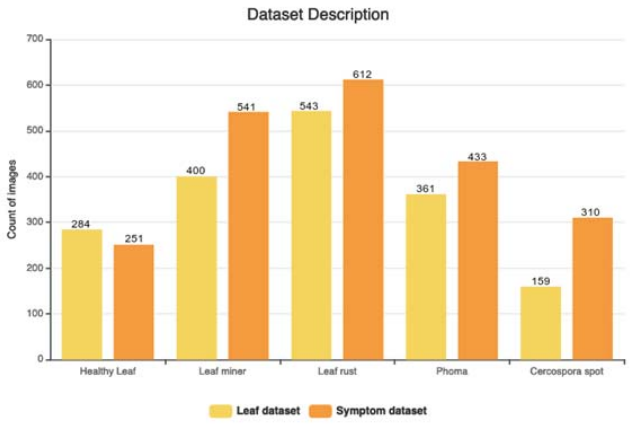


Fig. 1. Dataset Description

Details of each dataset is described below:

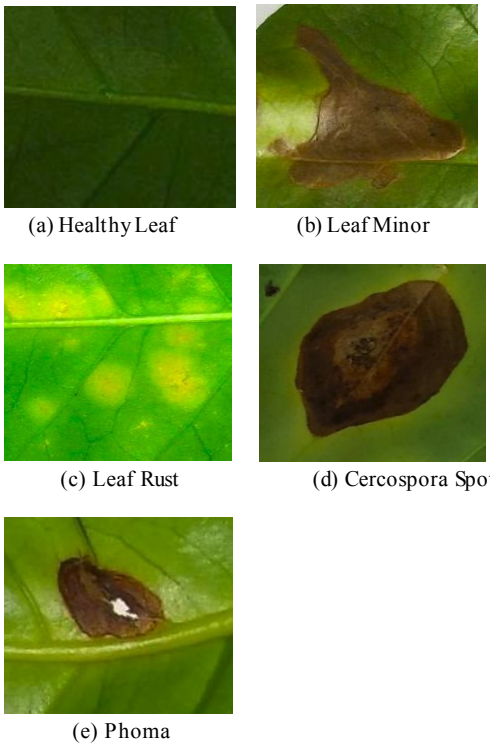
Leaf dataset: It contains images of coffee leaves and a CSV file with labels as different biotic stresses. All original images have the dimensions of 2048x1024 pixels.



Fig. 2. Sample of coffee leaves in Leaf Dataset

It consists of a total of 1747 images divided into 5 different classes - Healthy leaves, Infected leaves with Phoma, Leaf Miner, Coffee Leaf Rust and Cercospora Spots. Fig. 2 shows a sample of images found in this dataset.

Symptom dataset: A total of 2147 cropped images of the symptomatic parts of leaves are present. It is divided as per the named labelled classes - healthy leaves, leaf miner, leaf rust, phoma and cercospora. These are the 5 categories into which have classified our coffee leaves. Fig. 3 shows a sample of images available in this dataset. It is further divided into training, validation and test folders.



(a) Healthy Leaf (b) Leaf Minor

(c) Leaf Rust (d) Cercospora Spot

(e) Phoma

Fig. 3. Sample of cropped images showing only symptomatic leaf parts from Symptom Dataset

III. NETWORK ARCHITECTURE AND WORKING

The proposed network uses a Sequential model for classification. This Sequential model has been implemented using the Keras library in python language. The model has approximately 21 million parameters. Mini-Batch Gradient Descent (MBGD) is the optimization algorithm used to update the parameters. It performs much faster compared to conventional gradient descent algorithms. Fig. 4 represents the proposed neural network architecture. The images provided for training are input into the model in batches. The batch size value is kept as 32 with epoch value as 20. 1488 images belonging to 5 different classes is used for training the model and the image size has dimensions of 299x299 pixels.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:48:41 UTC from IEEE Xplore. Restrictions apply.

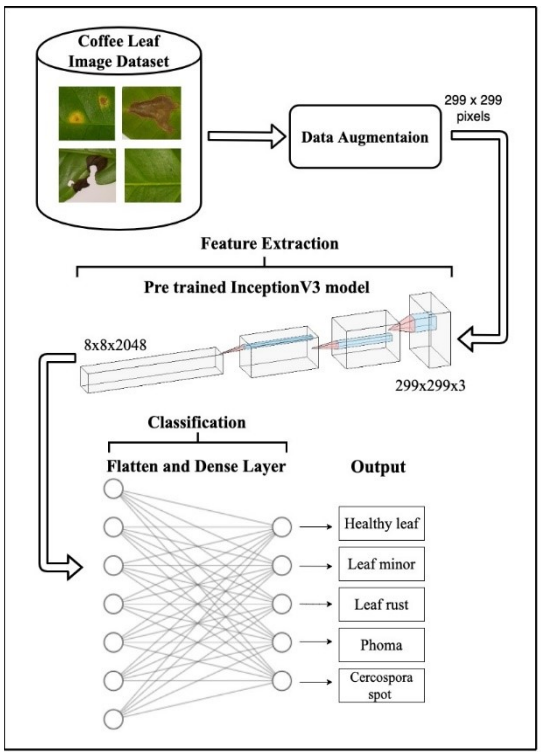


Fig. 4. Network Architecture

In the following sub-sections, explained the techniques of data augmentation and transfer learning and how it has enabled the proposed model to provide highly accurate testing results.

A. Data Augmentation

Data Augmentation is used to enhance the generalizability of the model. It applies augmentation techniques to produce a set of new images that are modified versions of our input data by application of image transformations, contrast changes, blurring images without changing their labels. It creates a varied and enlarged dataset and so the model is able to learn more features in a robust manner [8] and also prevents the problem of overfitting.

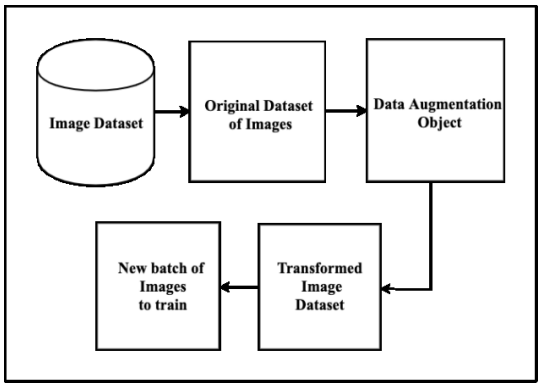


Fig. 5. Data Augmentation on the batch of images

Even though the technique enlarges the dataset but still cannot replicate the practical diversity found in nature. This is why only the image of symptomatic parts of leaves are used rather than considering the entire leaf [9]. In the proposed model and have used the in-place data augmentation technique using the Keras library in python. Implementation is called in-place because the transformation is applied on runtime i.e. it ensures that at each and every epoch new transformed image is input to the network which ultimately ensures high accuracy.

B. Transfer Learning

Transfer learning is a methodology that uses a pre-trained neural network. It makes use of knowledge gained while training on one dataset to be used on other datasets. It uses parametric values like weights and biases obtained while working on a large dataset and uses this knowledge to train on the new dataset. Image classification involves two processes, feature extraction and classification. With the help of transfer learning, one can reuse the part of the model involved in the feature extraction process and train the model for the classification part on a new dataset.

The model that used here for transfer learning is Inception v3 model. It is a 48-layer deep neural network. This model has been trained over a million images from ImageNet Database belonging to around 1000 different classes and have used TensorFlow hub library to perform transfer learning in our model and further added a flatten and a dense layer to our network. It changes the dimensionality of data and prevents the problem of overfitting of data. It has been used in the classification part.

C. Optimization Algorithm

Optimization algorithms help in the minimization (or maximization) of a given mathematical function F(x) which depends on the values of model's parameters and is used in carrying out computations for the output values. One such algorithm is Gradient Descent which falls under the umbrella of first-order optimization algorithms.

$$\theta = \theta - \alpha \nabla_\theta C(\theta; x_i, y_i) \qquad (1)$$

where θ is optimization parameter, C (θ; xi, yi) is the cost, (xi, yi) represents the data-point from the training set.

Mini-Batch Gradient Descent algorithm is used and the dataset is divided into batches of 32 images each. It combines the advantages of both Stochastic Gradient Descent and Batch Gradient Descent algorithms. The disadvantage of BGD is that it runs over the whole dataset to take a single step in the direction of minima. In contrast, SGD computes gradient for each randomly chosen datapoint which is faster but loses the advantages of vectorization and is error-prone also.

D. Categorical Cross-Entropy

Loss functions are used to evaluate the performance in the model. If our model predicts data that is highly deviated from the actual results, then loss function scales up to a large value.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:48:41 UTC from IEEE Xplore. Restrictions apply.

Thus, better predictions mean minimum value of loss function. There are many loss functions like Mean squared Error, Cosine Similarity, Binary Cross-entropy etc. One such loss function is Categorical Cross-Entropy. It is used for multi-class classification. It compares the distribution of our predictions with the actual labels. The probability of the true class is kept as 1 and for all other classes, the probability is kept 0. The equation is given as:

$$L(y, \hat{y}) = -\sum_{j=0}^{m} \sum_{i=0}^{n} (y_{ij} * log(\hat{y}_{ij})) \qquad (2)$$

where $\hat{y}$ is the predicted value and y representing the actual value, m and n represent the total number of data points and the total number of classes respectively.

IV. EXPERIMENTS

To implement this model, divided the dataset for training, validation and testing. 70% of images are used for training the model, 15% each for validation and testing. To create a varied dataset, data augmentation is implemented. The new output images are of dimensions 299x299 pixels. To decrease the training time and increase the performance, used InceptionV3 structure which is a pre-trained network. The convolutional layers that are present in it have pre-trained weights and biases. After each epoch, a new set of images are produced that are passed to the inception model. It forms the feature extraction part of the network.

Along with it, a dense and flatten layer are added to our sequential model. These layers serve in the classification part of the model. The flattening layer is used to convert a multidimensional output from a convolution layer to a linear output such that it can be directly fed into the neural network layer or to regular dense layer. Then, the dense layer performs the calculations and applies Softmax activation function.

Categorical cross-entropy is the loss function used and also tested the model for different values of learning rate as presented in Table I.

TABLE I. ACCURACY AND LOSS VALUE VARIATION WITH DIFFERENT LEARNING RATES

Learning Rate	Accuracy (%)	Loss Value
0.0001	95.82	0.48
0.001	95.22	0.42
0.005	97.61	0.35
0.01	97.31	0.36
0.1	74.03	1.73
1	32.01	10.1

Accuracy of the model increased with increase in learning rate at first, reaching a maximum at 0.005 with an accuracy of 97.61%, and started to decrease further. Hence, final value of the learning rate in the model is 0.005. Mini-batch gradient descent is used as the optimizer algorithm. Small batch size results in reduced accuracy [10]. Thus, batch size of 32 is used for maximum accuracy as shown in Table II.

TABLE II. BATCH SIZE VS ACCURACY

Batch Size	Accuracy (%)	Loss Value
4	94.03	0.49
8	95.22	0.49
16	97.31	0.36
32	97.61	0.35
64	97.01	0.37

V. RESULT

Using data-augmentation, transfer learning and MBGD as an optimizer, the proposed convolution neural network, has achieved an overall testing accuracy of 97.61% with a loss value of 0.35.

Table III represents how accuracy and loss improve with successive iterations. With every epoch, the accuracy increases and loss decrease significantly. After iteration 10, there is no significant improvement in the accuracy of our model although loss value continues to decrease slightly.

TABLE III. VARIATION IN ACCURACY AND LOSS VALUE WITH EPOCH

Epoch Number	Accuracy (%)	Loss Value
1	61.52	1.5752
2	73.43	0.7284
5	94.93	0.3632
8	97.01	0.3560
10	97.61	0.3535
12	97.61	0.3532

Further, Fig. 6 also clearly show that the value of loss and accuracy with varying epochs becomes nearly constant as more iterations are made.

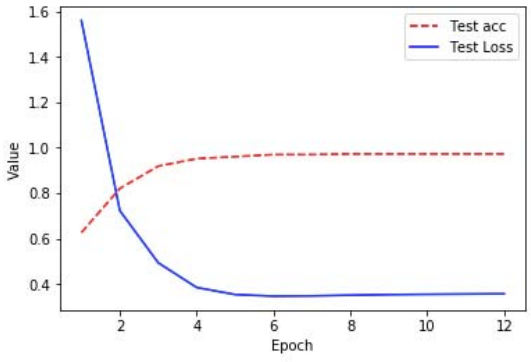


Fig. 6. Graph representing variation in accuracy and loss value with epoch

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:48:41 UTC from IEEE Xplore. Restrictions apply.

The coffee leaf images dataset are not used directly but rather used Data Augmentation technique to enhance the variety of training data, resulting into a much more reliable and accurate model and also avoids the problem of overfitting. The Transfer Learning method is applied which makes the training process faster and efficient. Moreover, flatten and dense layers are added to the network which helps in the classification process. Fig. 8 shows the confusion matrix based on the predicted results.

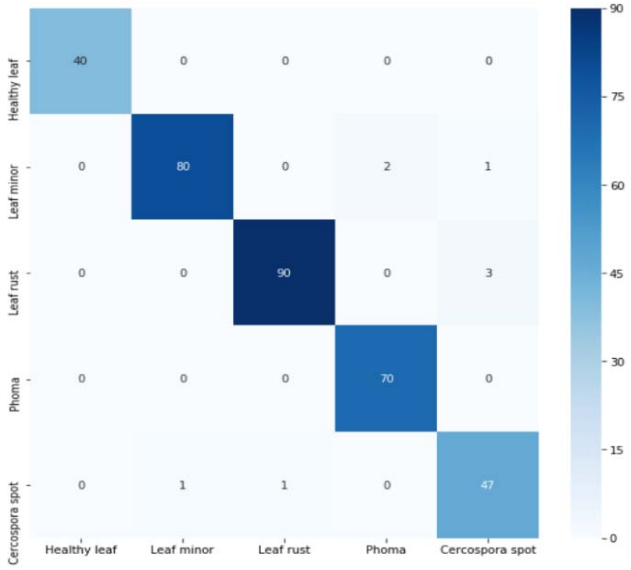


Fig. 8. Confusion matrix for proposed model

The average sensitivity and specificity computed for the proposed model is 0.974 and 0.992 respectively.

Table IV shows a comparison between various other techniques.

TABLE IV. COMPARISON OF DIFFERENT TECHNIQUES

S. No.	Technique Used	Accuracy (%)	Loss Value
1	Fuzzy Logic Based Expert System (using Decision Tree) [5]	85%	Not Available
2	Radial Basis Function Neural Network [11]	95%	Not Available
3	CNN with Data augmentation [12]	95%	0.10
4	CNN with data augmentation and transfer learning (using Inception V3 architecture) *	97.61%	0.35

*proposed model

The values of different hyper-parameters used in the proposed neural network are in Table V.

TABLE V. VALUES OF VARIOUS HYPER-PARAMETERS

Hyper-parameter	Value
Activation function	Softmax
Loss function	Categorical cross-entropy
Momentum	0.9
Batch size	32
Learning rate	0.005
Optimization function	Mini-batch gradient descent

VI. CONCLUSION AND FUTURE WORK

This research paper presents the use of deep convolutional neural networks in the detection and classification of healthy and unhealthy coffee plants. The unhealthy plant leaves are classified into 4 classes – Coffee Leaf Rust, Cercospora Spots, Leaf Miner, Phoma which are commonly occurring biotic stresses in coffee plants. One of the main contributions in our work is to propose a neural network for coffee plant disease detection with a very high level of accuracy.

In summary -

1. Identification and classification of coffee plant diseases by using the proposed neural network achieve a record-breaking accuracy of 97.61%.
2. With the help of this methodology, coffee plants farmer could be helped in timely identification of the diseases which will help in increasing the coffee production output of India.

As future work, aim to explore the requirements and necessary changes to our neural network to be able to accommodate various other new diseases in coffee plants and also account for variable environmental factors.

To make an even better prediction and achieve higher accuracy for a wide variety of diseases in coffee plants, additional variables such as temperature, air moisture, time of plantation and location weather are required.

REFERENCES

[1] Ricardo Balardin. 2010. *Factors affecting fungicide efficacy in the tropics, fungicides*, odile carisse, intechopen, doi: 10.5772/13583.

[2] S. A. Miller, F. D. Beed, and C. L. Harmon, "Plant disease diagnostic capabilities and networks," Annual Review of Phytopathology, vol. 47, pp. 15–38, 2009.

[3] P. Marcos, n. L. Silva rodovalho and a. R. Backes, "*coffee leaf rust detection using genetic algorithm,*" 2019 xv workshop de visão computacional (wvc), são bernardo do campo, brazil, 2019, pp. 16-20.

[4] L. Han, m. S. Haleem and m. Taylor, "*a novel computer vision-based approach to automatic detection and severity assessment of crop diseases,*" 2015 science and information conference (sai), london, 2015, pp. 638-644.

[5] Mettleq, A. S. A., & Abu-Naser, S. S. (2019). *A Rule Based System for the Diagnosis of Coffee Diseases.* International

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:48:41 UTC from IEEE Xplore. Restrictions apply.

Journal of Academic Information Systems Research (IJAISR), 3(3), 1-8.

[6] Suhartono, Derwin & Aditya, Wahyu & Lestari, Miranty & Yasin, Muhammad. (2013). *Expert System in Detecting Coffee Plant Diseases*. International Journal of Electrical Energy. 156-162. 10.12720/ijoee.1.3.156-162

[7] Krohling, Renato A.; Esgario , Guilherme J. M.; Ventura, José A. (2019), "BRACOL - A Brazilian Arabica Coffee Leaf images dataset to identification and quantification of coffee diseases and pests", Mendeley Data, v1. Available from: http://dx.doi.org/10.17632/yy2k5y8mxg.1

[8] . j. Shijie, w. Ping, j. Peiyi and h. Siping, "*research on data augmentation for image classification based on convolution neural networks,*" 2017 chinese automation congress (cac), jinan, 2017, pp. 4165-4170.

[9] Arnal Barbedo, J. G. (2019). *Plant disease identification from individual lesions and spots using deep learning*. Biosystems Engineering, 180, 96–107.

[10] Durmus, H., Gunes, E. O., & Kirci, M. (2017). *Disease detection on the leaves of the tomato plants by using deep learning*. 2017 6th International Conference on Agro-Geoinformatics.

[11] Rath, A. K., & Meher, J. K. (2020). *Disease detection in infected plant leaf by computational method*. Archives of Phytopathology and Plant Protection, 1–11.

[12] Marcos, a. P., rodovalho, n. L. S., & backes, a. R. (2019). *Coffee leaf rust detection using convolutional neural network*. 2019. Xv workshop de visão computacional (wvc). Doi:10.1109/wvc.2019.8876931.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:48:41 UTC from IEEE Xplore. Restrictions apply.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/342678861

Dual-Modal Emotion Recognition Using Discriminant Correlation Analysis

Conference Paper · July 2020

CITATIONS
0

READS
11

5 authors, including:



Utkarsh Bisaria
Delhi Technological University
1 PUBLICATION **0** CITATIONS
SEE PROFILE



Suhail Saifi
Delhi Technological University
1 PUBLICATION **0** CITATIONS
SEE PROFILE

All content following this page was uploaded by Suhail Saifi on 09 July 2020.

The user has requested enhancement of the downloaded file.

Proceedings of the International conference on Electronics and Sustainable Communication Systems (ICESCS 2020)
DVD Part Number: CFP20V66-DVD; ISBN: 978-1-7281-4107-7

Dual-Modal Emotion Recognition Using Discriminant Correlation Analysis

Dr. Rajiv Kapoor
Dept. of Electronics and Communication
Delhi Technological University
New Delhi. India
rajivkapoor@dce.ac.in

Suhail Saifi
Dept. of Electronics and Communication
Delhi Technological University
New Delhi, India
suhail_bt2k16@dtu.ac.in

Tarun Kapoor
Dept. of Electronics and Communication
Delhi Technological University
New Delhi, India
tarun_bt2k16@dtu.ac.in

Utkarsh Bisaria
Dept. of Electronics and Communication
Delhi Technological University
New Delhi, India
2utkarsh_bt2k16@dtu.ac.in

Nikhil Singh
Dept. of Electronics and Communication
Delhi Technological University
New Delhi, India
nikhil.gbpec2016@gmail.com

***Abstract*— A dual-modal framework is presented for human emotion recognition based on Discriminant Correlation Analysis. For visual features, key-frames are selected from the video sequences and features are extracted using a Residual Neural Network. The Tuning Network (T-Net) is used for fine-tuning of these visual features. From the audio channel, Intensity, MFCC, pitch, F0 envelope, Line Spectral Frequency (LSF), and Zero Crossing Rate (ZCR) is evaluated along with their delta regression coefficients. Discriminant Correlation Analysis based feature fusion is used to fuse audio and visual modalities.**

Keywords*— Dual-Modal Framework, Discriminant Correlation Analysis, T-Net, Residual Neural Network, Feature Fusion

I. INTRODUCTION

One of the key tasks for Human-Computer interaction is detecting and recognizing human emotions. Humans express emotions using different modalities, such as facial expressions, hand movements, speech, body posture, etc. Out of all these modalities, facial expression and speech are studied the most for human emotion recognition. For recognizing facial expression a lot of different challenges are faced due to head pose variations, illumination changes, 2D and 3D rotations, etc. For speech emotion recognition, major challenges include the extraction of significant features and finding appropriate classifiers. To improve the emotion recognition process, both audio and visual information extracted from the video clips are used . For detecting and recognizing facial expressions first, the face is detected in each frame using a Multi-Task CNN [7]. 2-D rotation is eliminated by aligning the face using facial landmarks. After face detection and alignment, key-frames are extracted from the frame sequence. The key-frames remove the redundant information inside video frames and make the process computationally less expensive by reducing the number of frames for processing.

For visual feature extraction, Xiaobai Li [1] used LBP histograms on three planes (LBP-TOP). To solve the problem of the variable number of key-frames, temporal interpolation was used, but it is computationally more expensive as compared to our approach. Junkai Chen [2] used Histograms of Oriented Gradients over three orthogonal planes (HOG-TOP) and concluded that compared with LBP-TOP, HOG-TOP is more robust and efficient. LBP-TOP and HOG-TOP, although maintains the temporal information, does not provide the extent of quality seen in features extracted using deep neural networks. In our framework, visual features are extracted from the key-frames using a Deep Residual Neural Network. The extracted features are fine-tuned using another Neural Network to reduce dimensionality and improve accuracy.

S. Sahoo [3] also proposed an audio-visual based emotion recognition system in which they used MFCC as audio and LBP as a visual feature and applied decision level fusion. The classification was done using HMM (Hidden Markov Model) and SVM classifier for audio and visual features respectively. For the audio part, the following features is extracted along with their delta regression coefficients: Intensity, MFCC, pitch, F0 envelope, Line Spectral Frequency (LSF), and Zero Crossing Rate (ZCR). Some statistical functions are applied to these features to extract a large and relevant amount of features from audio modality.

Junkai Chen [2] compared the accuracy of feature and decision level fusion and obtained better results with the former technique. In our framework, after extracting features from both modalities, Discriminant Correlation Analysis (DCA) [26] based feature level fusion is performed, which gives us highly information-rich fused data capable of representing both modalities together.

The final prediction is made on the feature vector obtained after the fusion of audio-visual modalities. The method was tested on SAVEE [4], CREMA-D [5], and eNTERFACE'05 [6] dataset and results were compared for audio, visual and fused audio-visual features.

II. FRAMEWORK

A. Visual Feature Extraction

1) Face Tracking and Face Alignment

One very important part of the emotion detection and recognition process is face detection and alignment. Significant challenges are faced due to 2D-3D rotation, different pose, lighting conditions, and different camera angles. Recently CNN's have proved to be very useful in solving a lot of computer vision tasks. A three-stage multi-task deep cascaded convolution network [7] is used for face detection and face alignment. In stage one, for the generation of candidate windows, fast Proposal Network is used. A refinement network (R-Net) is used in the second stage and the last stage The Output Network (O-Net) is used to produce the facial-key landmark position.

This 3-stage cascaded network has proved to be more accurate and less computationally expensive over the state-of-the-art techniques. The facial-key landmarks produced by the Output Network are used for face detection and alignment.

2) Key Frame Extraction

Rapid facial movements [8] between frames is analysed by studying the difference observed in LBP features.

a) Feature Difference analysis (FDA)

FDA is used to compare the difference in LBP features in sequential video frames within an interval.

The currently observed frame is referred to as the Target Frame (TF) when an interval of H frames is analysed. Start Frame (SF) and End Frame (EF) are the κth frames before and after TF respectively. Where κ is 1/2 × (*H*-1). The average of SF and EF is referred to as the Mean Feature Frame (MFF).

LBP histograms from each TF are compared to its MFF using the Chi-Squared distance. Any rapid facial movement results in high dissimilarity.

b) Threshold and peak estimation

The face is divided into 36 blocks and FDA is done for every block. Average of α largest FD values are used to describe each frame.

$$L = \frac{1}{\alpha}\sum_{j=1}^{\alpha}(value_j) \tag{1}$$

Contrasting is required to avoid errors due to local peak variations and noise:

$$L_{new\,i} = L_i - \frac{1}{2}(L_{i+k} + L_{i-k}) \tag{2}$$

Negative values are set to 0 as they represent that there are no sudden movements. Finally, the threshold is computed as:

$$th = L_{new\,mean} + \tau \times (L_{new\,max} - L_{new\,mean}) \tag{3}$$

Where τ is a percentage parameter and lies between 0 and 1. Subsequent peaks are separated by a distance of κ/2.

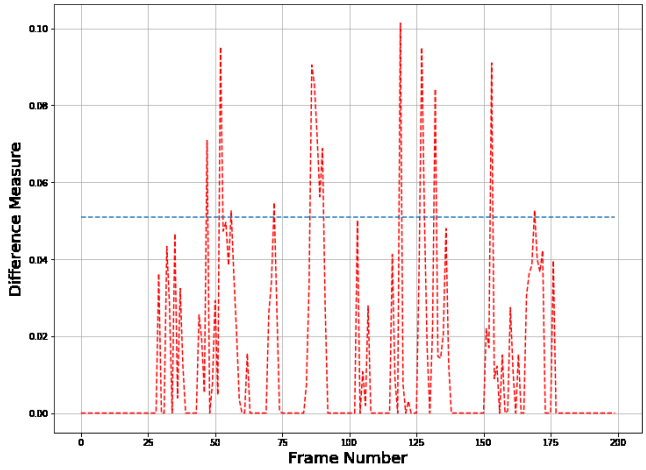


Fig.1. Plot showing FD vector vs. frame number

To obtain a constant number of key-frames from different videos, a new vector F is calculated as:

$$F_i = L_{new\,i} - th \tag{4}$$

λ frames with the highest corresponding values in F are selected as key-frames.

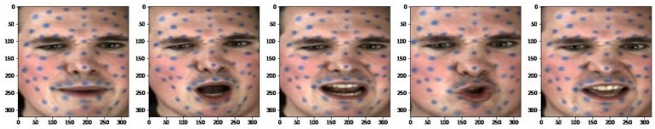


Fig.2. Key-frames extracted from a video of person JK in SAVEE expressing Anger

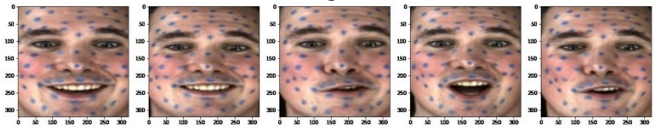


Fig.3. Key-frames extracted from a video of person JK in SAVEE expressing Happiness

3) ResNet-50

ResNet-50 is used [9], which is a 50 layer Residual Neural Network on each key-frame for feature extraction. The network can be trained on the emotion recognition datasets but better results are obtained using transfer learning on ImageNet [10] weight and bias.

Frames are fed to the Residual Network and feature vectors with 2048 dimensions are obtained. This method is repeated for each key-frame in a video sequence. For example, if there are λ key-frames, then the final feature vector with shape (λ×2048) is obtained for each video sequence.

4) Fine Tuning Video Features, T-Net

ResNet-50 operates on individual frames extracted from a video sequence, to represent each video with a single feature

vector, different approaches can be taken. The simplest way is to take mean overall the key-frames to obtain the average-feature set. This method is the least expensive in terms of computational power but the temporal information is not retained because the order of occurrence of frames inside the video sequence is not taken into consideration.

Temporal-information can be preserved using a feature descriptor that works in both the spatial and temporal domains. For example, Xiaobai Li [1] used LBP over three orthogonal planes (LBP-TOP) proposed by Zhao and Pietkäinen [11] as a feature descriptor, but according to previous studies, traditional LBP-TOP feature which uses all three histogram XY, XT, YT does not always yield the best performance. Xiaobai Li has used five different combinations of LBP histograms on three planes to increase performance. These methods, although maintains the temporal information does not provide the extent of quality seen in features extracted using deep neural networks like ResNet-50.

Taking mean overall key-frames not only results in loss of temporal information but also reduces the training data by a factor of λ, where λ is the number of key-frames per video. To solve this problem a method is proposed where instead of taking the mean overall key-frames, each key-frame is treated as an individual sample for training by assigning it the label of the video it is extracted from. This feature vector is referred to as the extended feature vector. The average and extended feature vector have same dimensions and only differ in the number of samples, in extended feature vector one video are split into λ samples and in the average feature vector, one video is represented by only one sample by taking mean overall the key-frames. A neural network is trained using the extended feature vector. This network is referred to as the tuning-network (T-NET). The output of the last dense layer of T-Net on the average feature set is used as the final visual feature vector.

B. Audio Feature Extraction

Audio have been used as our second modality for analysis. Human beings make use of both their facial expressions as well as their speech to express their emotions. Hence audio analysis is also very important which can give us some really important features to judge the emotion of a human being. For audio analysis, multiple features have been extracted that are quite relevant to express human emotions.

The feature set extracted consists of Intensity, MFCC, pitch, F0 envelope, Line Spectral Frequency (LSF), and Zero Crossing Rate (ZCR). It has been seen that by using time derivatives of these parameters, which are also called delta regression coefficients, the overall performance of the system can be improved hence delta regression coefficients have also been added for each of the above-mentioned parameters in our features set.

But there is a problem of variable feature-length of each audio sample depending on the duration of the audio clip (total number of frames will be different per audio clip). Thus, the coefficients of these features cannot be considered as it is, so to tackle this, some statistical functions have been applied on each of the parameters such that a fixed-length feature vector per audio clip can be obtained. The following statistical functions are applied: arithmetic mean, max/min value, range, linear regression coefficients, standard deviation, linear and quadratic error, kurtosis, skewness, and inter-quartile ranges.

Let us see what are the above-mentioned features mathematically are and how they are relevant features for detecting human emotions:

1) Pitch

It is the perceptual feature for a human listening to a particular sound. So in the emotion detection task, it is one of the important features to be determined. There are various time domain and frequency domain algorithms to determine pitch, also known as algorithms to find fundamental frequency(F0). Mathematically, the pitch for speech signal s is given as:

$$P(s) = D\{log|D(s.{w_n}^H\|s\|)|\} \quad (5)$$

Where D stands for DFT function, $\|s\|$ for the length of the signal and ${w_n}^H$ for hamming Window.

2) Intensity

Physically, the Intensity of a sound is defined as the amount of energy carried by an audio signal per unit time through a unit area. Mathematically, Intensity for a given sound signal s can be calculated as:

$$I_i(s) = \frac{\sum_{n=1}^{N}(s_{i+n}).{w_n}^H}{\sum_{n=1}^{N}{w_n}^H} \quad (6)$$

3) MFCC (Mel-frequency Cepstral Coefficient)

Mel-frequency Cepstral Coefficient (MFCC) is one of the most popular feature extractors in the field of audio analysis. It is the state-of-the-art audio feature extractor till now. The basic ideology behind this method is just to mimic the behaviour of human ears.

MFCC is calculated by a series of steps which have been explained in [12] very clearly. Finally, MFC coefficients are calculated by taking Discrete Cosine Transform (DCT) to decorrelate the filter bank coefficients, which is given by:

$$M_i = \sum_{k=1}^{N}\cos[i(k-1)\frac{\pi}{N}], \qquad i = 1,2,3 \dots P \quad (7)$$

Where P represents the number of cepstral coefficients $X_k, k = 1,2, \dots N$ is the log energy output of the k^{th} filter and N represent the number of triangular filter bank filters.

4) ZCR (Zero Crossing Rate)

The Zero-Crossing Rate (ZCR) of an audio frame is determined as the rate of change of sign of the signal in one frame duration.

Mathematically, ZCR for a frame $(Z(i))$ of length l can be calculated as:

$$Z(i) = \frac{1}{2l}\sum_{n=1}^{l}|sgn[s_i(n)] - sgn[s_i(n-1)]| \quad (8)$$

Where,

$$sgn[x_i(n)] = \begin{cases} 1, s_i(n) \geq 0 \\ -1, \ s_i(n) < 0. \end{cases}$$

C. Feature Fusion Technique

After extracting audio and visual features, the Discriminant Correlation Analysis (DCA) have been applied based feature fusion.

DCA is one of the robust fusion techniques. This technique is applied at the feature level, where a rich amount of information is available. In this fusion technique both of the feature, vectors are scaled in such a manner that inter-class correlation can be minimized within feature vectors and further scaled so that pair-wise correlation between two feature vectors can also be maximized. For more understanding, consider the following example:

Assume that η number of data samples have been collected that correspond to m different class labels. Now, let us divide this η columns matrix into m sections. In this matrix $u_{ij} \in U$ is j^{th} sample vector in the i^{th} class, $\overline{u_i}$ represents mean of the u_{ij} vectors belong to i^{th} class and $\bar{u}$ represents mean of the whole feature set i.e., $\overline{u_i} = \frac{1}{\eta_i}\sum_{j=1}^{\eta_i} u_{ij}$ and $\bar{u} = \frac{1}{\eta}\sum_{j=1}^{m} \eta_i \overline{u_i}$. Therefore, the inter-class scatter matrix can be described as :

$$S_{bu_{(p\times p)}} = \sum_{i=1}^{m} \eta_i (\overline{u_i} - \bar{u})(\overline{u_i} - \bar{u})^T = \boldsymbol{\varphi}_{bu}\boldsymbol{\varphi}_{bu}{}^T \quad (9)$$

Where p denotes the total number of features,

$$\boldsymbol{\phi}_{bu_{(p\times m)}} = \left[\sqrt{\eta_1}\,(\overline{u_1} - \bar{u}), \sqrt{\eta_2}\,(\overline{u_2} - \bar{u}), \ldots, \sqrt{\eta_m}\,(\overline{u_m} - \bar{u})\right]$$

So, if $p>>m$, which will be the case most of the time in real-world situations it will always be simple to compute $(\boldsymbol{\varphi}_{bu}{}^T\boldsymbol{\varphi}_{bu})_{(m\times m)}$ as covariance matrix instead of evaluating $(\boldsymbol{\varphi}_{bu}\boldsymbol{\varphi}_{bu}{}^T)_{(p\times p)}$. In [13] it has been shown that the most important eigenvectors of $\boldsymbol{\varphi}_{bu}\boldsymbol{\varphi}_{bu}{}^T$ can be acquired by simply mapping the eigenvectors of $\boldsymbol{\varphi}_{bu}{}^T\boldsymbol{\varphi}_{bu}$. Hence, just need to obtain eigenvectors of the matrix $(\boldsymbol{\varphi}_{bu}{}^T\boldsymbol{\varphi}_{bu})_{(m\times m)}$.

Now, to get well-separated classes, $\boldsymbol{\varphi}_{bu}{}^T\boldsymbol{\varphi}_{bu}$ should be a diagonal matrix. To diagonalize it, transformation can be used , given as:

$$A^T(\boldsymbol{\varphi}_{bu}{}^T\boldsymbol{\varphi}_{bu})A = \hat{\Delta} \quad (10)$$

Where A represents a matrix of orthogonal eigenvectors and $\hat{\Delta}$ represents a diagonal matrix where eigenvalues are sorted in descending order.

Now, consider $B_{(m\times c)}$ a matrix which consists of c eigenvectors obtained from A matrix by selecting c largest non-zero eigenvalues, such that:

$$B^T(\boldsymbol{\varphi}_{bu}{}^T\boldsymbol{\varphi}_{bu})B = \Delta_{(c\times c)} \quad (11)$$

Hence, the corresponding c most important eigenvectors for scatter-matrix in Eq. 7 can be acquired by simply mapping: $B \rightarrow \boldsymbol{\varphi}_{bu}B$:

$$(\boldsymbol{\varphi}_{bu}B)^T S_{bu}(\boldsymbol{\varphi}_{bu}B) = \Delta_{(c\times c)} \quad (12)$$

Therefore, $\theta_{bu} = \boldsymbol{\varphi}_{bu}B\Delta^{-1/2}$, is the weight which diagonalizes S_{bu} and lowers the dimensions of our matrix U from p to c, i.e.,

$$\theta_{bu}{}^T S_{bu}\theta_{bu} = I, \quad (13)$$

$$U'_{(c\times\eta)} = \theta_{bu_{(c\times p)}}{}^T U_{(p\times\eta)} \quad (14)$$

Where, U' is the transformed version of U in which classes are well-separated and inter-class correlation is minimized.

Similarly, another feature setcan be solved for, V and find transforming weight θ_{bv}, which minimize inter-class correlation in V feature vector and lowers the dimensions from q to c.

$$\theta_{bv}{}^T S_{bv}\theta_{bv} = I, \quad (15)$$

$$V'_{(c\times\eta)} = W_{bv_{(c\times q)}}{}^T V_{(q\times\eta)}\,. \quad (16)$$

Till now transformed U to U' and V to V', such that between class correlation has been minimized but now need to further transform U' and V' such that features of U' will have non-zero correlation only with corresponding features in another set, in V'. To accomplish this need to diagonalize another matrix, which is called the between-set covariance matrix, $S'_{uv} = U'V'^T$. For this task, singular value decomposition (SVD) can be used :

$$S'_{uv_{(c\times c)}} = X\,\Sigma Y^T \Rightarrow X^T S'_{xy} Y = \Sigma \quad (17)$$

Where, U' and V' have the rank of order c and $S'_{uv\ (c\times c)}$ is non-degenerate. Thus, Σ is a diagonal matrix in which only diagonal elements are non-zero. So, have $\theta_{du} = X\Sigma^{-1/2}$ and $\theta_{dv} = Y\Sigma^{-1/2}$ such that:

$$\left(X\Sigma^{-\frac{1}{2}}\right)^T S'_{uv}\left(Y\Sigma^{-\frac{1}{2}}\right) = I, \quad (18)$$

Where, S'_{uv} gets unitized. Therefore, transformed finally our feature vectors as:

$$U^* = \theta_{du}{}^T U' = \underbrace{\theta_{du}{}^T\theta_{bu}{}^T}U = \theta_u U \quad (19)$$

$$V^* = \theta_{dv}{}^T V' = \underbrace{\theta_{dv}{}^T\theta_{bv}{}^T}V = \theta_v V \quad (20)$$

Where $\theta_u = \theta_{du}{}^T\theta_{bu}{}^T$ and $\theta_v = \theta_{dv}{}^T\theta_{bv}{}^T$ are the final transformation weights for U and V, respectively.

Hence after transforming our audio feature vector, say U to U' and visual feature vector, say V to V', feature level

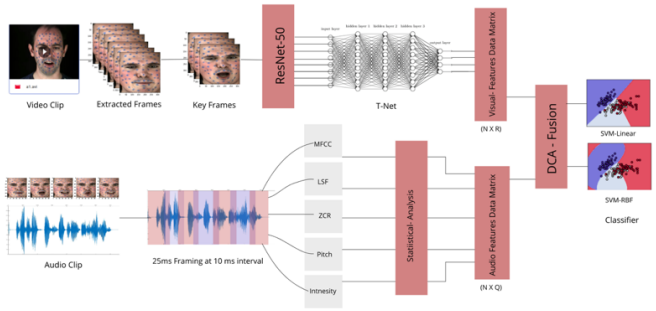


Fig.4. Proposed framework

fusion can be performed by simply concatenating or taking the summation of the transformed feature vectors as shown:

$$Z_1 = \begin{pmatrix} U^* \\ V^* \end{pmatrix} = \begin{pmatrix} W_u{}^T U \\ W_v{}^T V \end{pmatrix} = \begin{pmatrix} W_v & 0 \\ 0 & W_u \end{pmatrix}^T \begin{pmatrix} U \\ V \end{pmatrix} \quad (21)$$

Or

$$Z_2 = U^* + V^* = W_u{}^T U + W_v{}^T V = \begin{pmatrix} W_u \\ W_v \end{pmatrix} \begin{pmatrix} U \\ V \end{pmatrix} \quad (22)$$

However, the concatenation method have been used to fuse audio (U) and visual (V) feature vectors i.e., Z_1.

III. Experiment and Results

In this section, qualitative and quantitative analysis is done in the terms of accuracy with other existing techniques on three datasets i.e. SAVEE, CREMA-D, and eNTERFACE'05 and results are compared for audio, visual and fused audio-visual features.

A. SAVEE

The SAVEE database contains recordings from 4 different actors expressing 7 different emotions. The database has a total of 480 audio-visual clips recorded at 60fps. We merged classes 'Sadness' and 'Disgust' due to their high co-relation and to maintain a constant 6-class classification on each dataset.

For key-frame extraction, the value of H is set to H = 19 (to maintain 0.3 sec). The default value of 0.5 is used for τ. For LBP, radius r is set to r = 3 and neighboring points to $8 \times r = 24$. λ is set to 20. The output of the third dense layer of T-Net on the average feature vector is used as the final visual-feature set for fusion. For audio features, a frame size of 25ms at the rate of 10ms is used. 12 MFCC coefficients, 8 line spectral frequencies along with other features with their respective delta regression coefficients are extracted from each frame. Hence in total $24 \times 2 = 48$ different types of features are calculated and 17 statistical functions are applied on each of these features, therefore in total $48 \times 17 = 816$ types of audio features are calculated per audio clip.

After getting feature data matrices from both the modalities i.e., $X_{(480\times816)}$ from audio and $Y_{(480\times512)}$ from visual, DCA fusion is applied which gives us the fused feature vector of dimension (48×10) for classification. SVM [14] with linear and RBF kernel is used for classification at each stage.

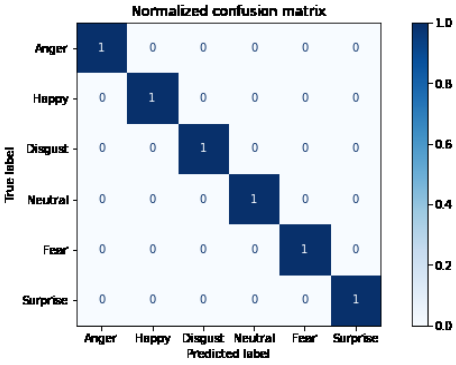


Fig. 5. Confusion matrix for 6 classes evaluated using audio-visual features with SVM classifier on the SAVEE dataset.

TABLE 1
Accuracy on SAVEE dataset

Feature type	Classifier	Accuracy
Visual	SVM (RBF)	99.00%
Audio	SVM (RBF)	90.70%
Audio-Visual	SVM (linear)	100%

This table represents the accuracy of our classification model when only the audio feature vector is used i.e. $X_{(480\times816)}$ which is 90.70%, when only visual feature vector is used i.e. $Y_{(480\times512)}$ which is 99.00% and when a DCA based fused vector is used which is 100%.

TABLE 2
Comparison of SAVEE

Method	Accuracy
ISLA [15]	86.01%
F. Noroozi, SVM [16]	98.10%
F. Cid [17]	96.20%
S. Haq [18]	99.00%
D. Gharavian [19]	96.88%
Ours	**100%**

This table compares the emotion recognition rates from our proposed framework with some other pre-existing frameworks for emotion detection.

B. CREMA-D

CREMA-D contains 7,442 clips from 91 actors. The sentences are presented using 6 different emotions at four different levels. For our application, all levels of emotions are considered as one.

For key-frame extraction, the value of H is set to H = 11 (to maintain 0.3 sec). The default value of 0.5 is used for τ. For LBP, radius r is set to r = 3 and neighboring points to $8 \times r = 24$. λ is set to 20. The output of the third dense layer of T-Net on the average feature vector is used as the final visual-feature set for fusion. For audio features, a frame size

of 25ms at the rate of 10ms is used. 14 MFCC coefficients, 10 line spectral frequencies along with other features with their respective delta regression coefficients are extracted from each frame. Hence in total $28 \times 2 = 56$ different types of features are calculated and 17 statistical functions are applied on each of these features, therefore in total $56 \times 17 = 952$ types of audio features are calculated per audio clip.

After getting feature data matrices from both the modalities i.e., $X_{(7442\times952)}$ from audio and $Y_{(7442\times512)}$ from visual, DCA fusion is applied which gives us the fused feature vector of dimension (7442×10) for classification. SVM with linear and RBF kernel is used for classification at each stage.

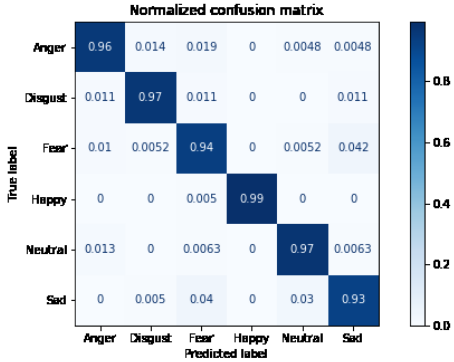


Fig. 6. Confusion matrix for 6 classes evaluated using audio-visual features with SVM classifier on CREMA-D dataset

TABLE 3
Accuracy on CREMA-D dataset

Feature type	Classifier	Accuracy
Visual	SVM (RBF)	83.80%
Audio	SVM (RBF)	62.10%
Audio-Visual	SVM (linear)	95.90%

This table represents the accuracy of our classification model when only the audio feature vector is used i.e. $X_{(7442\times952)}$ which is 62.10%, when only visual feature vector is used i.e. $Y_{(7442\times512)}$ which is 83.80% and when a DCA based fused vector is used which is 95.90%.

TABLE 4
Comparison on the CREMA-D dataset

Method	Accuracy
E. Ghaleb [20]	74.00%
R. Beard [21]	65.00%
E. Ghaleb [22]	66.50%
Ours	**95.90%**

This table compares the emotion recognition rates from our proposed framework with some other pre-existing frameworks for emotion detection.

C. eNTERFACE'05

The eNTERFACE'05 contains audio-visual clips from 46 subjects reacting to six different situations and expressing 6 different emotions. The videos are recorded at a 25fps frame rate.

For key-frame extraction, the value of H is set to $H = 7$ (to maintain 0.3 sec). The default value of 0.5 is used for τ. For LBP, radius r is set to $r = 2$ and neighboring points to $8 \times r = 16$. λ is set to 20. The output of the third dense layer of T-Net network on the average feature vector is used as the final visual-feature set for fusion. For audio features, a frame size of 25ms at the rate of 10ms is used. 12 MFCC coefficients, 10 line spectral frequencies along with other features with their respective delta regression coefficients are extracted from each frame. Hence in total $26 \times 2 = 52$ different types of features are calculated and 17 statistical functions are applied on each of these features, therefore in total $52 \times 17 = 884$ types of audio features are calculated per audio clip.

After getting feature data matrices from both the modalities i.e., $X_{(1166\times884)}$ from audio and $Y_{(1166\times512)}$ from visual, DCA fusion is applied which gives us the fused feature vector of dimension (1166×10) for classification. SVM with linear and RBF kernel is used for classification at each stage.

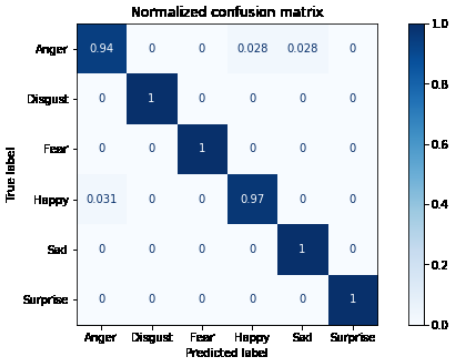


Fig. 7. Confusion matrix for 6 classes evaluated using audio-visual features with SVM classifier on eNTERFACE'05 dataset

TABLE 5
Accuracy on eNTERFACE'05 dataset

Feature type	Classifier	Accuracy
Visual	SVM (RBF)	84.10%
Audio	SVM (linear)	67.30%
Audio-Visual	SVM (RBF)	98.50

This table represents the accuracy of our classification model when only the audio feature vector is used i.e. $X_{(1166\times884)}$ which is 67.3%, when only visual feature vector is used i.e. $Y_{(1166\times512)}$ which is 84.10% and when a DCA based fused vector is used which is 98.50%.

View publication stats

TABLE 6
Comparison on the eNTERFACE'05 dataset

Method	Accuracy
M. ShamimHossain [23]	86.40%
F. Noroozi [16], SVM	94.92%
Subhasmita Sahoo [3]	81.00%
AVEF [24]	86.40%
Mansoorizadeh [25]	84.02%
Ours	**98.50%**

This table compares the emotion recognition rates from our proposed framework with some other pre-existing frameworks for emotion detection.

IV. CONCLUSION

Our proposed framework based on Discriminant Correlation Analysis on three different standard datasets have been successfully applied: SAVEE, CRAMA-D, and eNTERFACE'05. Our results on CREMA-D outperform other previously proposed frameworks in [20, 21, 22] and achieve state-of-the-art results with a recognition rate of 95.90%. On SAVEE and eNTERFACE'05, our results are equivalent to state-of-the-art results which are 100% and 98.5% respectively as compared to the results in [16]. In future, neural networks will be used for audio feature extraction.

V. REFERENCES

[1] X. Li et al., "Towards Reading Hidden Emotions: A Comparative Study of Spontaneous Micro-Expression Spotting and Recognition Methods," in IEEE Transactions on Affective Computing, vol. 9, no. 4, pp. 563-577, 1 Oct.-Dec. 2018.

[2] J. Chen, Z. Chen, Z. Chi and H. Fu, "Facial Expression Recognition in Video with Multiple Feature Fusion," in IEEE Transactions on Affective Computing, vol. 9, no. 1, pp. 38-50, 1 Jan.-March 2018.

[3] S. Sahoo and A. Routray, "Emotion recognition from audio-visual data using rule based decision level fusion," 2016 IEEE Students' Technology Symposium (TechSym), Kharagpur, 2016, pp. 7-12.

[4] P. Jackson and S. Haq, "Surrey audio-visual expressed emotion (savee) database," 2015, http://kahlan.eps.surrey.ac.uk/savee/

[5] H. Cao, D. G. Cooper, M. K. Keutmann, R. C. Gur, A. Nenkova and R. Verma, "CREMA-D: Crowd-Sourced Emotional Multimodal Actors Dataset," in IEEE Transactions on Affective Computing, vol. 5, no. 4, pp. 377-390, 1 Oct.-Dec. 2014.

[6] O. Martin, I. Kotsia, B. Macq and I. Pitas, "The eNTERFACE' 05 Audio-Visual Emotion Database," 22nd International Conference on Data Engineering Workshops (ICDEW'06), Atlanta, GA, USA, 2006, pp. 8-8.

[7] K. Zhang, Z. Zhang, Z. Li and Y. Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," in IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499-1503, Oct. 2016.

[8] A. Moilanen, G. Zhao, and M. Pietikäinen, "Spotting rapid facial movements from videos using appearance-based feature difference analysis," in ICPR. IEEE, 2014, pp. 1722–1727.

[9] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 770-778.

[10] J. Deng, W. Dong, R. Socher, L. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, 2009, pp. 248-255.

[11] G. Zhao and M. Pietikäinen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," PAMI, vol. 29, no. 6, pp. 915–928, 2007.

[12] Koduru, A., Valiveti, H.B. & Budati, A.K. Feature extraction algorithms to improve the speech emotion recognition rate. *Int J Speech Technol* **23,** 45–55 (2020). https://doi.org/10.1007/s10772-020-09672-4

[13] M. Turk and A. Pentland, "Eigenfaces for recognition," Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71–86, 1991.

[14] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt and B. Scholkopf, "Support vector machines," in IEEE Intelligent Systems and their Applications, vol. 13, no. 4, pp. 18-28, July-Aug. 1998.

[15] Y. Kim and E. M. Provost, "ISLA: Temporal Segmentation and Labeling for Audio-Visual Emotion Recognition," in IEEE Transactions on Affective Computing, vol. 10, no. 2, pp. 196-208, 1 April-June 2019.

[16] F. Noroozi, M. Marjanovic, A. Njegus, S. Escalera and G. Anbarjafari, "Audio-Visual Emotion Recognition in Video Clips," in *IEEE Transactions on Affective Computing*, vol. 10, no. 1, pp. 60-75, 1 Jan.-March 2019.

[17] F. Cid, L. J. Manso, and P. Nunez, "A novel multimodal emotion recognition approach for affective human robot interaction," Proc. FinE, pp. 1–9, 2015.

[18] S. Haq, T. Jan, A. Jehangir, M. Asif, A. Ali, and N. Ahmad, "Bimodal human emotion classification in the speaker-dependent scenario," Pakistan Academy Sci., Art. no. 27, vol. 52, no. 1, pp. 27– 38, 2015.

[19] D. Gharavian, M. Bejani, and M. Sheikhan, "Audio-visual emotion recognition using FCBF feature selection method and particle swarm optimization for fuzzy ARTMAP neural networks," Multi- media Tools Appl., vol. 76, pp. 1–22, 2016.

[20] E. Ghaleb, M. Popa and S. Asteriadis, "Multimodal and Temporal Perception of Audio-visual Cues for Emotion Recognition," 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII), Cambridge, United Kingdom, 2019, pp. 552-558.

[21] R. Beard, R. Das, R. W. Ng, P. K. Gopalakrishnan, L. Eerens, P. Swietojanski, and O. Miksik, "Multi-modal sequence fusion via recursive attention for emotion recognition," in Proceedings of the 22nd Conference on Computational Natural Language Learning, 2018, pp. 251–259.

[22] E. Ghaleb, M. Popa and S. Asteriadis, "Metric Learning Based Multimodal Audio-visual Emotion Recognition," in *IEEE MultiMedia*.

[23] Emotion recognition using deep learning approach from audio–visual emotional big data M.Shamim Hossain , Ghulam Muhammad , Emotion Recognition Using Deep Learning Approach from Audio-Visual Emotional Big Data, Information Fusion (2018), doi: https://doi.org/10.1016/j.inffus.2018.09.008.

[24] Y. Ma, Y. Hao, M. Chen, J. Chen, P. Lu, A. Košir Audio-visual emotion fusion (AVEF): a deep efficient weighted approach. Inf. Fusion, 46 (2019), pp. 184-192

[25] Mansoorizadeh, M., Moghaddam Charkari, N. Multimodal information fusion application to human emotion recognition from face and speech. Multimed Tools Appl 49, 277–297 (2010). https://doi.org/10.1007/s11042-009-0344-2

[26] M. Haghighat, M. Abdel-Mottaleb and W. Alhalabi, "Discriminant Correlation Analysis: Real-Time Feature Level Fusion for Multimodal Biometric Recognition," in IEEE Transactions on Information Forensics and Security, vol. 11, no. 9, pp. 1984-1996, Sept. 2016.

Iranian Journal of Science and Technology, Transactions of Mechanical Engineering
https://doi.org/10.1007/s40997-020-00366-5

RESEARCH PAPER

Effect of Distributed Damping on the Dynamic Behavior of Flexible Chassis of Heavy Road Vehicle Under Standardized Random Road Responses

Ashish Gupta[1] · **Vikas Rastogi**[2]

Received: 22 April 2019 / Accepted: 23 May 2020
© Shiraz University 2020

Abstract
This paper evaluates the dynamics of a heavy road truck on different random road conditions, whereas the issue of beam flexibility is also being accompanied. Heavy vehicles like trucks and buses having a large structure to carry their load ride in all kinds of terrains, for thousands of miles without failing structure. Structure (chassis or frame) of the truck should be able to sustain all types of vibration (i.e., longitudinal, transverse and flexural). This study starts with developing an analytical model with consideration of the Rayleigh beam approach, where the rotary inertia of the beam has also been admitted. This work also considered internal damping of the chassis, which has been neglected in past studies specifically for flexural vibration. Further bond graph model of the flexible structure of a heavy road vehicle is developed, where this model is derived through modal expansion approach. The dynamic response of the structure is analyzed under various random road conditions at different vehicle speeds. These random models of the road are developed according to ISO 8608 standard, whereas four (H1–H4) kinds of road categories are considered. Results show the dynamic response of the chassis under a real road response so that parameters of the various parts of the structure can be evaluated, which further raises the ride comfort and road holding capability of the structure. Further, this work also examines the effect of structural damping or internal damping on the dynamics of the system.

Keywords Lagrangian approach · Bond graph · Frame flexibility of the chassis · Structural damping · Random road conditions

1 Introduction

Ride vibration is a principal concern for vehicle dynamics, where frame flexibility plays a very significant role. Numerous studies have been conducted on problems of ride vibration. However, very few studies have been carried out to study the effect of vibration, which includes a flexible frame. It has been generally noticed that the flexural vibration can be neglected for smaller vehicles, relatively for stiffer automobiles. However, trucks and the long heavy vehicle experience significant beaming mode vibration (Margolis and Edeal 1989). Ibrahim (1996) has created a model of truck frame through FEM where he calculated the modal properties of the frame. The results were presented in terms of PSD and RMS of the vehicle vertical response. Most of the analyses on the effects of chassis flexibility in vehicle dynamics were based on the use of particular model developed, whereas the FEM (finite element method) and the modal superposition theory have been extensively used to calculate the modal properties of the frame structure. Generally, the problem occurs due to the beam flexibility of the vehicle, which is termed as 'beaming response of the vehicle' (Rideout 2012). Rideout and Khan (2010) presented the model of the frame flexibility issue for truck model. They created a pitch plane model which involved longitudinal dynamics and transverse frame vibrations. Huang and Zeng (2010) studied about flexural vibration problems for high-speed passenger rail vehicle, where car body was modeled through Euler beam approach. This research underlined

✉ Ashish Gupta
ashish.gupta@glbitm.ac.in; ashish.j001@gmail.com

[1] Mechanical Engineering Department, G.L Bajaj Institute of Technology and Management, Greater Noida 201306, India

[2] Design Centre, Department of Mechanical Engineering, Delhi Technological University, Delhi 110042, India

Published online: 23 June 2020

the effect of internal damping in ride comfort of rail vehicle. Kumar et al. (2017) analyzed the whole-body vibration for railway system. They have considered biodynamic model of human for the assessment of ride comfort, where whole-body model was integrated with flexible bogie model system. The flexible bogie system was modeled through Euler–Bernoulli beam approach. Similarly in other studies, Euler–Bernoulli approach was attempted to introduce beam flexibility in heavy truck chassis modeling (Zhou et al. 2009; Kim and Yim 1994). Heavy long vehicles have a ladder-type chassis structure across the whole length of the vehicle. These structures are stiffer in bending, which can respond to bending frequencies more than 20 Hz. When the engine and other elements are mounted on the structure, these bending frequencies are reduced up to 8 Hz, for highly stiffer frames (Cai-hong and Zeng 2010; Tomioka et al. 2017). Tomioko and Takigami (2015) have conducted the experimental study on flexural beam to evaluate the effect of frame flexibility on passenger comfort. Some other studies apparently show that the beaming frequency is higher than the rigid frequency of the vehicle (Tomioka et al. 2006; Agostinacchio et al. 2014). However, in large vehicles, beaming frequency is influenced by road inputs.

A lot of research has been conducted considering the issue of frame flexibility in the analysis of railway vehicle. But, limited studies were reported on dynamic analysis of heavy road vehicles considering random road disturbances. A country like India, where multifarious road conditions are present, which creates an infirm condition before all leading automobile industries to maintain comfort level as per standard on every road conditions. Some studies have mentioned that structural damping can participate in ride quality improvement. So, in the present work frame flexibility has been incorporated by considering frame as Raleigh beam, which also includes the effect of the rotary inertia. Lagrangian approach has been applied to obtain the energy equations for the vehicle system, which resulted from the well-known rigid body modes and the modal parameters of the frame flexibility. The work involves the importance of structural or internal damping of beam-type chassis structure on flexural vibration and ride quality of the heavy vehicle with random road disturbance prescribed by ISO 8608 (Meirovitch 1980) as input. The bond graph approach has been utilized to model the vehicle with flexible frame. Besides the beam response, the effect of different road conditions as per ISO 8608 and vehicle speeds on the dynamics of vehicle has also been presented. The work begins with deriving Lagrange equation of frame considered as Rayleigh beam for truck flexural response. Then, this formulation is being extended for bond graph modeling with or without consideration of structural damping effect. Simulations were carried out to visualize the beaming mode response with respect to random road conditions as prescribed by ISO 8608.

2 Analytical Framework for Heavy Vehicle System Through Classical Mechanics

2.1 Generalized Lagrange's Equation for a Continuous System

It is a well-known fact that the Lagrange formulation provides one of the most convenient ways of writing down the equation of motion for a wide range of mechanical systems. The principal advantage of Lagrange's equation is that it is easier to apply to dynamical systems other than the simplest. This formulation is created from scalar quantities of kinetic energy, potential energy and work expressed in terms of generalized coordinates and developed for universal handling of dynamical systems (Mukherjee and Karmakar 2000). The equation is formulated equations of motion from the principle of least action. This density function is a function comprised of time, transverse displacements and transverse velocities, and up to second derivatives of displacements to space coordinates x, which may be expressed as

$$\widehat{L}\big(w(t,x),\dot{w}(t,x),w'(t,x),w''(t,x),\dot{w}'(t,x)\big)=\hat{T}(t,x)-\hat{V}(t,x) \tag{1}$$

where $\widehat{L}(\cdot)$ is the Lagrangian density function and $\hat{T}(\cdot)$ is the kinetic energy density, which may be expressed as,

$$\widehat{T}(x,t)=\widehat{T}(\dot{w}(x,t),\dot{w}'(x,t)) \tag{2}$$

Moreover, $\hat{V}(t,x)$ is the potential energy density, which has the following functional dependence,

$$\widehat{V}(x,t)=\widehat{V}(w'(x,t),w''(x,t)) \tag{3}$$

The next step is used for finding the equation of motion and the boundary conditions for the beam through the variational principle (see Appendix 2). This formulation yields

$$\delta\int_{t1}^{t2}\widehat{L}\big(w(t,x),\dot{w}(t,x),w'(t,x),w''(t,x),\dot{w}'(t,x)\big)\mathrm{d}t=0 \tag{4}$$

where $\delta(\cdot)$ is infinitesimal variation operator and it is very similar to the total derivative operator; the only difference is that $\delta(\cdot)$ does not vary with time. Another significant property of infinitesimal operator is that it commutes with the differential operator. Using the extremization condition, one may obtain

$$\int_0^L\int_{t1}^{t2}\delta\widehat{L}\big(w(t,x),\dot{w}(t,x),w'(t,x),w''(t,x),\dot{w}'(t,x)\big)\,\mathrm{d}x\mathrm{d}t=0 \tag{5}$$

or

$$\int_0^L \int_{t1}^{t2} \left[\frac{\partial \widehat{L}}{\partial w}\delta w + \frac{\partial \widehat{L}}{\partial \dot{w}}\delta \dot{w} + \frac{\partial \widehat{L}}{\partial w'}\delta w' + \frac{\partial \widehat{L}}{\partial \dot{w}'}\delta \dot{w}' + \frac{\partial \widehat{L}}{\partial w''}\delta w'' \right] \mathrm{d}x\,\mathrm{d}t = 0 \tag{6}$$

The detailed derivation of the variational formulation is presented in "Appendix 1."

2.2 Analytical Formulation for Flexural Behavior of the Truck Chassis Structure

An analytical formulation for flexural behavior of the truck structure is being attempted in this section. The vehicle base is modeled as a rigid body with a local coordinate reference frame (w, x) attached to the center of mass (M) and aligned with the inertia principal axes as shown in Fig. 1. It has density (ρ) and $p$itch inertia (J) with respect to the body $W_1(t) - W_2(t)$ displacement axis, EI flexural rigidity and $\theta(t)$ pitch velocity of the base. It is connected with the suspension system to different elements of the vehicles. The spring stiffness characterizes the spring and damper in the suspension system k_s and damping coefficient R_s, respectively. The various displacements of the vehicle are described $w(x, t)$ with respect to the equilibrium positions. As the vehicle is assumed to be rigid, its motion can be described by the vertical displacement (bounce) as $w(x, t)$ and the rotation about the transverse horizontal axis (pitch) as $\dot{\theta}(t)$.

The model of the vehicle is being created with the following assumptions:

- The components of the vehicle body act as a rigid body.
- The springs and dampers of the suspension system have linear characteristics.
- The spring damper system is assumed to be massless.
- The tire is assumed to provide stiffness and thus modeled as a spring element.
- The positions of the two ends of the spring connecting two rigid bodies or connecting one rigid body and one contact point are required as input data.
- The spring damper system connects each rigid body.
- The traveling road is assumed to be straight.
- The road inputs are assumed to be random.

The analytical framework of truck chassis can be assumed as a free–free beam structure, where Rayleigh beam approach is being adopted. Let us consider a transverse deflection of the beam $w(x,t)$, where x varies from 0 to beam length L. The formalism of beam flexibility is easily obtained through the well-known Lagrangian equation. Lagrangian density variation may be obtained from the following expression as

$$L = \left[\frac{1}{2}\int_0^L \left(\rho A \dot{w}(x,t)^2 + \rho I \dot{w}'(x,t)^2\right)\mathrm{d}x - \frac{1}{2}\int_0^L \left(EIw''(x,t)^2\right)\mathrm{d}x \right] \tag{7}$$

where ρ is the mass density, A is the area of cross section, EI is the rigidity and I is the second moment of area of the cross section of the beam about the neutral axis. In Eq. (7), the first term represents kinetic energy of the beam, the second term presents as inertial energy and the final term is strain energy. The variational formulation presents a refined and classical method of deriving the equation of motion of a dynamical system, where all the boundary conditions of a system can be revealed. However, fundamentals of variational formulation for the continuous system are detailed in "Appendix 2."

The equation of motion is obtained after applying a variational formulation and boundary conditions in free vibration condition, which may be expressed as

$$\rho A \ddot{w}(x,t) - \left(\rho I \dot{w}''(x,t)\right)' + \left(EIw''(x,t)\right)'' = 0 \tag{8}$$

This equation of motion is based on Rayleigh beam model. The term $\left(EIw''(x,t)\right)''$ is usually referred to as the

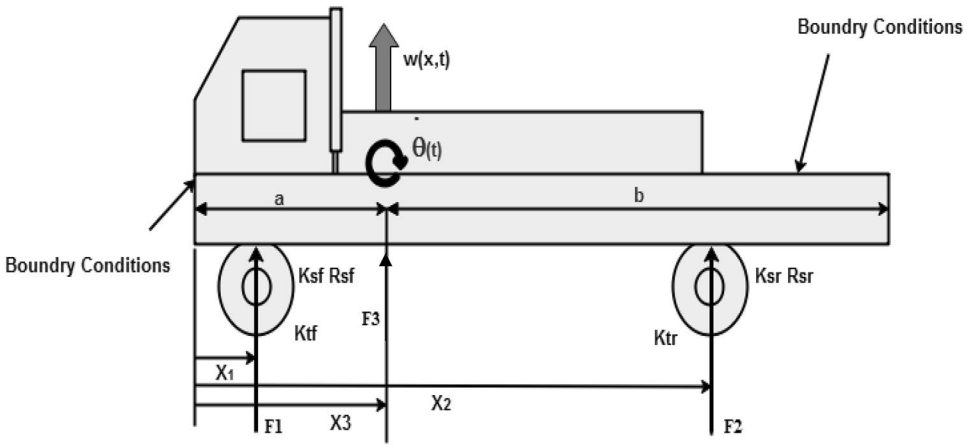


Fig. 1 Schematic diagram of a continuous frame of vehicle

flexural term, where EI is called the flexural stiffness and $\left(\rho I\dot{w}''(x,t)\right)'$ is known as the rotary inertia term. If this term is zero, then Eq. (8) will become a Euler beam model. Thus, the term $\left(\rho I\dot{w}''(x,t)\right)'$ can also be named as a Rayleigh beam term.

Assume a modal solution for Eq. (8) in the form of

$$w(x,t) = W(x)e^{i\omega t} \tag{9}$$

where ω is the eigenfrequency and $W(x)$ is the eigenfunction. The actual solution of the equation may be obtained by taking the real part of the solution. Substituting the modal solution in the field Eq. (8) yields after rearrangement as

$$-\omega^2\left[\rho AW - \left(\rho IW'\right)'\right] + \left(EIW''\right)'' = 0 \tag{10}$$

which along with the corresponding boundary condition represents the eigenvalue problem for a Rayleigh beam.

2.3 Modal Analysis for Uniform Rayleigh Beam

Considering Eq. (10) as a uniform beam and substituting in Eq. (9) a solution of the form

$$W(x) = Ce^{\lambda x} \tag{11}$$

where C and λ are constants, Eq. (10) can be rewritten as

$$EI\lambda^4 - \omega^2\rho I\lambda^2 - \omega^2\rho A = 0 \tag{12}$$

yielding

$$\lambda^2 = \frac{1}{2EI}\left[\omega^2\rho I \pm \sqrt{\omega^4\rho^2I^2 + 4\omega^2EI\rho A}\right]$$

Thus, the spatial equation (real form) of Eq. (12) has the general solution, which may be expressed as

$$W(x) = A\cosh\lambda x + B\sinh\lambda x + C\cos\lambda x + D\sin\lambda x \tag{13}$$

Here, one may consider a uniform Rayleigh beam for which the free–free boundary conditions are given by

$$W(0) = 0, \quad W'(0) = 0, \quad W''(l) = 0, \quad W'''(l) = 0, \tag{14}$$

Using, Eq. (14) in Eq. (13) yields the frequency equation as

$$\cosh\lambda_n l\cos\lambda_n l - 1 = 0 \tag{15}$$

The solution of Eq. (15) yields

$$\lambda_n = \left(\frac{2n+1}{2}\pi\right)\frac{1}{l}$$

Moreover, the mode shape function may be expressed as

$$\begin{aligned} W_n(x) = {} & (\cos\lambda_n l - \cosh\lambda_n l)(\sin\lambda_n x + \sinh\lambda_n x) \\ & - (\sin\lambda_n l - \sinh\lambda_n l)(\cos\lambda_n x + \cosh\lambda_n x) \end{aligned} \tag{16}$$

Thus, one may obtain the natural frequency of Eq. (6), which may be written as

$$\omega_n = \lambda_n^2\frac{1}{\left[1+\lambda_n^2\frac{I}{A}\right]^{1/2}}\sqrt{\frac{EI}{\rho A}} \tag{17}$$

It is recognized that conditions are force-free at this point, which permits $\omega_n = 0$ to be a mode frequency. Thus, Eq. (10) will be written as

$$\frac{\mathrm{d}^4W}{\mathrm{d}x^4} = 0 \tag{18}$$

or

$$W = c_1x^3 + c_2x^2 + c_1x + c_4 \tag{19}$$

There are two possible solutions for Eq. (13), which satisfied all boundary conditions. They are $W = \text{const}$ and $W = ax + b$. These are called rigid body modes, and it is convenient to assume them as rigid body rotation of the beam about the centrally located center of mass. Thus,

$$W_{00}(x) = 1 \tag{20}$$

and

$$W_0(x) = x - \frac{L}{2} \tag{21}$$

Let us consider an analysis for truck chassis, which is presented in Fig. 1. In this figure, the half model of the road truck is supported by front and rear suspension, which exerted an external force on the beam over the various road responses. From equation of motion, Eq. (8) will be rewritten as

$$\begin{aligned} & \rho A\ddot{w}(x,t) - \left(\rho I\dot{w}''(x,t)\right)' + \left(EIw''(x,t)\right)'' \\ & = F_1\delta(x-x_1) + F_2\delta(x-x_2) + F_3\delta(x-x_3) \end{aligned} \tag{22}$$

where F_1 and F_2 are the force exerted by front and rear suspension systems, respectively, and F_3 is dead weight on the beam, $\delta(\cdot)$ is the Dirac delta function, x_1 and x_2 are the respective distance from left end to the front and rear suspension of the vehicle, respectively, and x_3 is the distance of dead weight from the left end. One may assume that the solution of Eq. (22) is separable in space and time, and the forced solution has the following form,

$$w(x,t) = \sum_{n=0}^{\infty} W_n(x)U_n(t) \tag{23}$$

where $W_n(x)$ depends on the spatial position and $U_n(t)$ depends on time. Introducing Eq. (23) into Eq. (22), and multiplying each term by $W_m(x)$ and integrating with respect

to $x=0$, $x=l$, one may then use the orthogonal property to the modes and obtain

$$\int_0^l \rho A W_n^2 \mathrm{d}x\ddot{U}(t) + \omega_n^2(\rho I\lambda_n^2 - \rho A)\int_0^l W_n^2 \mathrm{d}xU(t) = F_1 W_n(x_1) + F_2 W_n(x_2) + F_3 W_n(x_3) \tag{24}$$

$$\rho A\left(1 - \frac{I}{A}\lambda_n^2\right)\int_0^l W_n^2 \mathrm{d}x\ddot{U}(t) + \rho A\left(1 - \frac{I}{A}\lambda_n^2\right)\omega_n^2\int_0^l W_n^2 \mathrm{d}xU(t) = F_1 W_n(x_1) + F_2 W_n(x_2) + F_3 W_n(x_3) \tag{25}$$

Equation (25) can be represented in general form as

$$m\ddot{U}(t) + kU(t) = F_1 W_n(x_1) + F_2 W_n(x_2) + F_3 W_n(x_3) \tag{26}$$

where

$$m = \rho A\left(1 - \frac{I}{A}\lambda_n^2\right)\int_0^l W_n^2 \mathrm{d}x$$

$$k = m\omega_n^2\int_0^l W_n^2 \mathrm{d}x$$

Now one may write the first frequency modes (Karnopp et al. 2012)

$$\left[\rho A\int_0^l (1)_n^2 \mathrm{d}x\right]\ddot{U}_{00} = F_1 + F_2 + F_3 \quad \text{or} \quad m\ddot{U}_{00} = F_1 + F_2 + F_3 \tag{27}$$

which states that the external force accelerates the center of mass of the beam. The other zero frequency mode yields

$$\left[\rho A\int_0^l \left(x - \frac{l}{2}\right)^2 \mathrm{d}x\right]\ddot{U}_{00} = F_1 + F_2 + F_3 \tag{28}$$

or

$$J\ddot{U}_0 = F_1 + F_2 + F_3$$

This equation states that the moment of the external forces about the center of mass produces angular acceleration $\ddot{U}_{00}$, where J is the centroid moment of inertia of beam. The next subsection will present the formulation, which includes structural damping.

2.4 Incorporation of Structural Damping

Now one may consider internal damping (the detailed derivation of internal damping is presented in "Appendix 2" for ready reference) of the beam and Eq. (22) may be rewritten as

$$\rho A\ddot{w}(x,t) - \left(\rho I\dot{w}''(x,t)\right)' + \left(EIw''(x,t)\right)'' - \mu_I I\left(\dot{w}'''\right)' = F_1\delta(x - x_1) + F_2\delta(x - x_2) + F_3\delta(x - x_3) \tag{29}$$

$$\rho A\ddot{w}(x,t) - \left(\rho I\dot{w}''(x,t)\right)' + \omega^2(\rho I\lambda^2 - \rho A) - \mu_I I\left(\dot{w}'''\right)' = F_1\delta(x - x_1) + F_2\delta(x - x_2) + F_3\delta(x - x_3) \tag{30}$$

where μ_I is the structural damping coefficient, δ is the Dirac delta function, x_1 and x_2 are the respective distance from left corner to the front and rear suspension of the vehicle and x_3 is the distance of dead weight from the extreme left corner of the vehicle. F_1 and F_2 are the suspension forces of the front and rear suspension system, respectively, and F_3 is the dead weight. The front and rear suspension forces of the vehicle may be expressed as

$$F_1 = C_{\mathrm{s}}^{\mathrm{f}}(V_{\mathrm{f}} - \dot{w}_1) + K_{\mathrm{s}}^{\mathrm{f}}\left(\int_0^t V_{\mathrm{f}}\mathrm{d}t - w_1\right) \tag{31}$$

$$F_2 = C_{\mathrm{s}}^{\mathrm{r}}(V_{\mathrm{r}} - \dot{w}_2) + K_{\mathrm{s}}^{\mathrm{r}}\left(\int_0^t V_{\mathrm{r}}\mathrm{d}t - w_2\right) \tag{32}$$

where $C_{\mathrm{s}}^{\mathrm{f}}$ and $C_{\mathrm{s}}^{\mathrm{r}}$ are, respectively, damping coefficient of the front and rear damper; V_{f} and V_{r} are the road velocity input at front and rear wheel, respectively; $K_{\mathrm{s}}^{\mathrm{f}}$ and $K_{\mathrm{s}}^{\mathrm{r}}$ are the suspension stiffness of front and rear suspension system, respectively; $\dot{w}_1$ and w_1 are the velocity and displacement of front sprung mass; and $\dot{w}_2$ and w_2 are the velocity and displacement of rear sprung mass.

One may assume solution similar to Eq. (24); applying Eqs. (23)–(30), one may obtain a solution as

$$\int_0^l \rho A W_n^2 \mathrm{d}x\ddot{U}(t) + \omega_n^2(\rho I\lambda_n^2 - \rho A)\int_0^l W_n^2 \mathrm{d}xU(t) - \rho I\lambda_n^2\int_0^l W_n^2 \mathrm{d}x\ddot{U}(t) - \mu_I I\lambda_n^4\int_0^l W_n^2 \mathrm{d}x\dot{U}(t) = F_1 W_n(x_1) + F_2 W_n(x_2) + F_3 W_n(x_3) \tag{33}$$

Rearranging Eq. (33), one may obtain

$$\rho A\left(1 - \frac{I}{A}\lambda_n^2\right)\int_0^l W_n^2 \mathrm{d}x\ddot{U}(t) + \rho A\left(1 - \frac{I}{A}\lambda_n^2\right)\omega_n^2\int_0^l W_n^2 \mathrm{d}xU(t) - \mu_I I\lambda_n^4\int_0^l W_n^2 \mathrm{d}x\dot{U}(t) = F_1 W_n(x_1) + F_2 W_n(x_2) + F_3 W_n(x_3) \tag{34}$$

Finally, Eq. (34) may be represented as

$$m_n\ddot{U}(t) + R_n\dot{U}(t) + k_nU(t) = F_1W_n(x_1) + F_2W_n(x_2) + F_3W_n(x_3) \tag{35}$$

where

$$m_n = \rho A\left(1 - \frac{I}{A}\lambda_n^2\right)\int_0^l W_n^2\mathrm{d}x, \tag{36}$$

$$k_n = m\omega_n^2\int_0^l W_n^2\mathrm{d}x, \tag{37}$$

$$R_n = \mu_I I\lambda_n^4\int_0^l W_n^2\mathrm{d}x \tag{38}$$

Equations (36–38) show that the rotary inertia affects the dynamics of the beam, which is not being taken in case of Euler beam. However, in this case, the value of $I/A \cdot \lambda_n^2$ is always less than 1. Generally, distributed elements that one may obtain to involve in almost all the system models are lightly damped. The location of structure damping cannot be precisely identified from where it arises. The mechanism of energy dissipation is very typical and complicated; some dissipation takes place because of the plastic deformation and some of the energy radiation from material surfaces. It is not assured in the modeling that individual conventional damper will attach to model for representing the damping of distributed elements. It can happen if the damping includes functionally by incorporating it into individual modes. This is accompanied by merely appending R-elements to the mode oscillators of the modal bond graph. R-element is represented in Eq. (30), where μ_I is varied from 0.01 to 0.1. The next subsection will present the bond graph modeling to visualize the simulation for different parameters.

3 Bond Graph Modeling of the Truck Chassis Structure

The physical system of a truck chassis (pitch plane model) with front and rear suspension system driven over the random road is presented in Fig. 1. The boundary condition of the structure is taken as free–free. The dynamic disturbance in vertical direction of the vehicle due to random road conditions is the main attraction of this modeling. The longitudinal and lateral dynamics are not considered in the modeling. The bond graph technique is being used as a modeling tool, and the bond graph of the system is created through formulation, which was developed in the previous section. Figure 2 presents the bond graph model of a flexible infinitesimal beam, whereas Fig. 3 shows the bond graph model of a flexible beam with internal damping along with suspension system. The rigid body modes appear simply as inertia elements with associated modal stiffness and damping. The first inertia parameter is the beam mass (M), and second is the beam moment of inertia (Jg). The transformer (-TF-) elements are connected to the rigid body modes, which correctly apply the forces and moments to these elements. The modal masses are taken from Eq. (36), whereas modal stiffness and modal damping can be taken from Eqs. (37) and (38), respectively.

The system is entirely flexible in transverse and bending, but torsionally rigid. Each 0-junction is attached to each 1-junction with modal components through transformer (TF) elements with moduli equal to appropriate mode function. The transformer function is moduli with shape function (W_n), and the values of this function can be determined from Eq. (16). The attached systems are then appropriately appended to the external 0-junction bonds to form a complete, low-order, very accurate model. Figure 3 shows a general finite mode representation with two input (front and rear) location, where Eq. (26) represents a governing contribution of *i*th mode.

In the real world, distributed elements have some damping properties, which resist vibration level up to a certain level. No mechanism can quickly identify the value of damping on specific. However, only the quantity of dissipative energy can be determined. It is customary not to attempt to perform detailed modeling of the damping mechanism, but instead to include damping functionally by incorporating its individual modes. This is accomplished by simply appending R-elements to the mode oscillator of the bond graph. The bond graph model of the distributed system including damping modes is presented in Fig. 3. The governing equation of this configuration is mentioned in Eq. (23).

4 Numerical Simulation

Computer simulation can compress the performance of a system over the years into a few minutes of the computer running time. The parameters used for simulation for truck chassis system are presented in Table 1. Simulation models are relatively flexible and can be modified to accommodate the changing environment to the real situation. At present, most of the simulation models are made using differential equations. In this research, a half car flexible model with suspension systems is investigated at

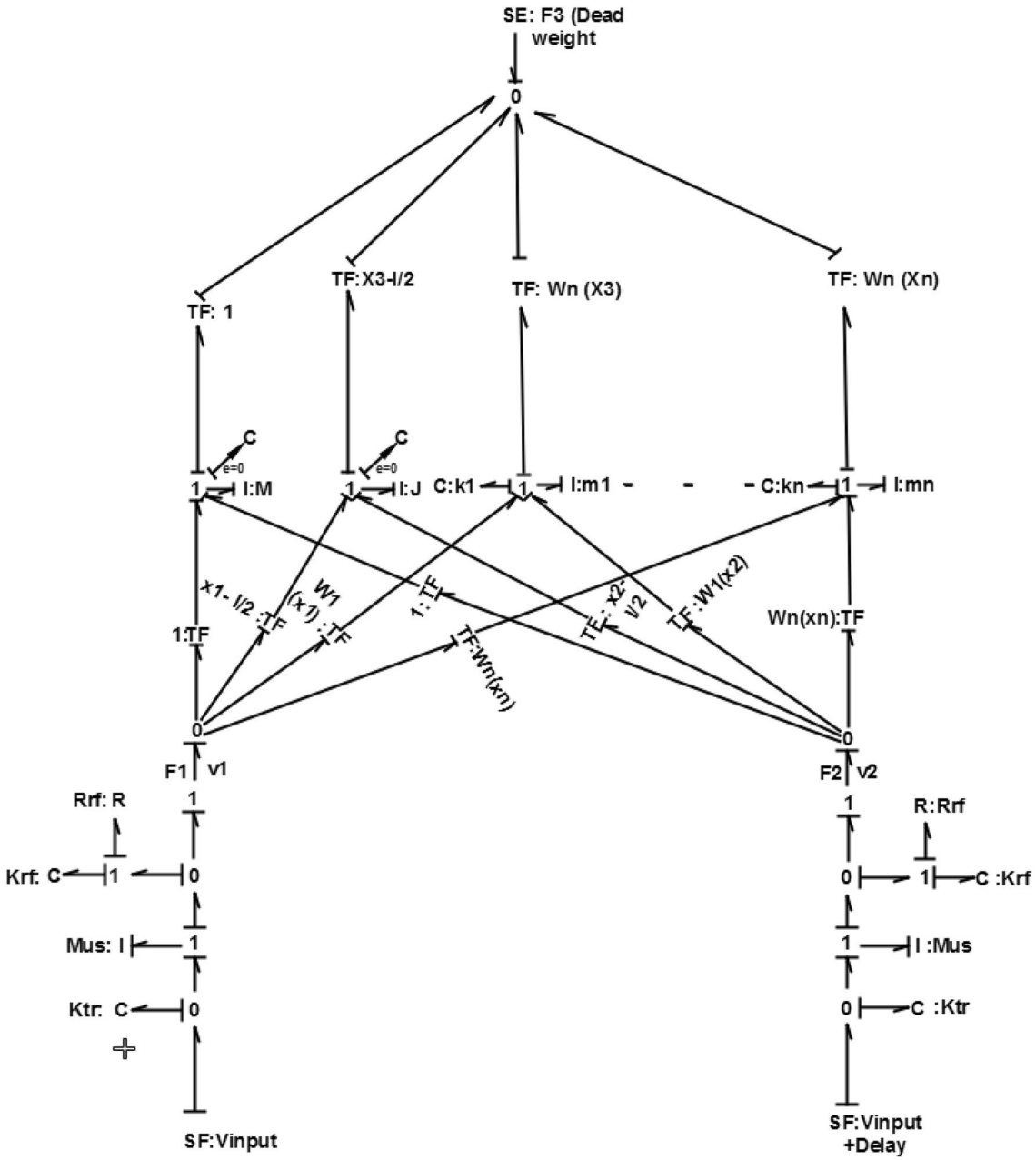


Fig. 2 Bond graph model of truck chassis considering internal damping

different operating speeds on different road profile inputs using bond graph technique and simulators of SYMBOLS Sonata® and MATLAB/Simulink® software are used (Gupta and Rastogi 2016; Hassani 2013; Mukherjee and Samantaray 2006). The bond graph model of the vehicle is simulated for 10 s to obtain different output responses. In total, 1024 records are used in the simulation and error is kept in the order of 5.0×10^{-4}. Fifth-order Runge–Kutta Gill method is used in this present work to solve the various differential equations generated through bond graph model.

4.1 Road Inputs

Two types of road profile inputs of the simulation of this system are used, viz. steady-state input and random road input.

4.1.1 Sine Wave Input

The steady-state input has been taken in a sinusoidal form, and the following expression is used for simulation (Hagedorn and DasGupta 2007):

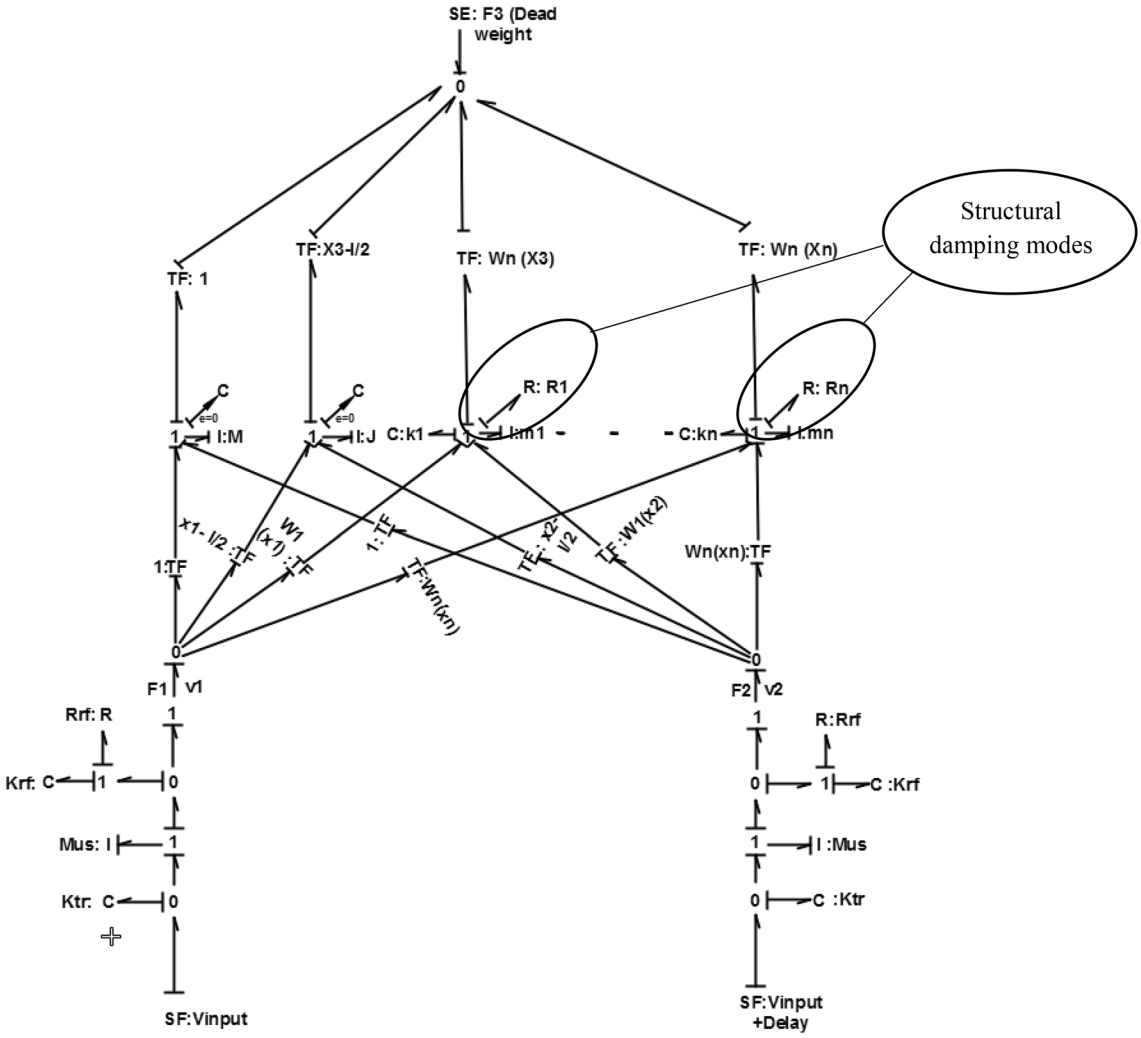


Fig. 3 Bond graph model of truck chassis considering internal damping

Table 1 Simulation parameters for flexible chassis system (Rideout and Khan 2010)

Parameter	Value	Unit
Vehicle length	10	m
Frame mass (Mg)	4350	kg/m
Frame inertia (J)	24,000	kg/m^2
Suspension stiffness (K_{sf}, K_{sr})	375, 870	kN/m
Suspension damping (R_{sf}, R_{sr})	31,895, 33,384	Ns/m
Flexural rigidity (EI)	1.05e+08	N m^2
Unsprung mass (Mus)	50	kg
x_1, x_2, x_3, x_4	2.95, 5.75, 7, 3	m

$$V_{\text{input}} = A_{\text{jump}} * \omega * \sin(\omega * t), \quad \text{for front wheel} \tag{39}$$

$$V_{\text{input}} = A_{\text{jump}} * \omega * \sin(\omega * (t + d/v)), \quad \text{for rear wheel} \tag{40}$$

where A_{jump} is the amplitude of vibration and ω is the excitation frequency.

4.1.2 Random Road Condition

A random road profile is generated according to the International Organization for Standardization (ISO 8608). It gives a depiction of the road profile through estimation of the PSD of the vertical displacements G_d, as a function of spatial frequency $n(n = \Omega/2\pi(\text{cycles})/\text{m})$ and also of angular spatial frequency Ω. ISO 8608 introduces a classification, which is evaluated in accordance with conventional values of spatial frequency $n_0 = 0.1$ cycles/m and angular spatial frequency $\Omega_0 = 1$ cycles/m. There are eight classes of roads mainly

identified from classes H1 to H8 according to the values of $G_d(n)$ and $G_d(\Omega)$ established in ISO 8608, which are shown in Table 2 (Meirovitch 1980).

In the simulation, ISO 8608 provides the roughness of the road surface profile, which may be stated using the following equations:

$$G_d(n) = G_d(n_0)\left(\frac{n}{n_0}\right)^{-w} \tag{41}$$

$$G_d(\Omega) = G_d(\Omega_0)\left(\frac{n}{n_0}\right)^{-w} \tag{42}$$

where w denotes the waviness and its value is taken to be 2 in this case. In this research, random road inputs have been developed by taking into consideration the PSD of vertical displacement G_d as a function of spatial frequency n.

Beginning with a distributed road profile to specify the value of spatial frequency n, looped within a frequency band Δn, the PSD function value is represented through the following expression as

$$G_d(n) = \lim_{\Delta n \to 0}\left(\frac{\psi_x^2}{\Delta n}\right) \tag{43}$$

where the mean square value is denoted by ψ_x^2 which signifies the component of the signal for spatial frequency n contained by the frequency band Δn. Accordingly, the signal of road profile is discretized, and thus, it is characterized by a series of elevation points along with evenly spaced ones. Here, road profile length is denoted by L, whereas the sampling interval is denoted by B. Further, the maximum value of sampling spatial frequency c is $1/B$, the maximum effective sampling spatial frequency (n_{eff}) is $n_{\max}/2$ and the discretized spatial frequency values n_i are equally spaced under the frequency domain, with an interval of (Δn) as $1/L$.

So herein, generic spatial frequency value n_i can be regarded as $i{\cdot}\Delta n$ and Eq. (43) may be written in the discrete form as

$$G_d(n) = \frac{\psi_x^2(n_i, \Delta n)}{\Delta n} = \frac{\psi_x^2(i \cdot \Delta n, \Delta n)}{\Delta n} \tag{44}$$

where i varying from 0 to N as $\frac{n\max}{\Delta n}$, one may obtain the road profile by a simple harmonic function

$$h(x) = A_i \cos(2\pi \cdot n_i \cdot x + \phi) = A_i \cos(2\pi \cdot i \cdot x \cdot \Delta n + \phi) \tag{45}$$

where the amplitude is denoted by A_i, spatial frequency is denoted by n_i and phase angle is φ. One may generate a harmonic signal through mean square value

$$\psi_x^2 = \frac{A_i^2}{2} \tag{46}$$

Therefore,

$$G_d(n_i) = \frac{\psi_x^2(n_i)}{\Delta n} = \frac{A_i^2}{2 \cdot \Delta n} \tag{47}$$

It has been shown in several works that the development of an artificial road profile using Eq. (45) is only possible if the PSD function of vertical displacements is well known, presuming a random phase angle φ_i following a uniform probabilistic distribution within range 0 to 2π. The artificial profile can be given as

$$\begin{aligned} h(x) &= \sum_{i=0}^{N} A_i \cos(2\pi \cdot n_i \cdot x + \phi_i) \\ &= \sum_{i=0}^{N} \sqrt{2 \cdot \Delta n \cdot G_d(i \cdot \Delta n)} \cdot A_i \cos(2\pi \cdot i \cdot x \cdot \Delta n + \phi) \end{aligned} \tag{48}$$

Using Eq. (47) in Eq. (45), a random road profile can be generated according to ISO classification by the following equation,

Table 2 ISO 8608 values of $G_d(n_0)$ and $G_d(\Omega_0)$

Road class	$G_d(n_0)$ (10^{-6} m^3)		$G_d(\Omega_0)$ (10^{-6} m^3)		Category
	Lower limit	Upper limit	Lower limit	Upper limit	
H1	–	32	–	2	New roadways with asphalt layers
H2	32	128	2	8	New roadways with concrete layers
H3	128	512	8	32	Old roadways layers which are not maintained
H4	512	2048	32	128	Roadways layer consisting if cobble stone layers or similar materials
H5	2048	8192	128	512	Very poor road condition
H6	8192	32,768	512	2048	Off-road condition
H7	32,768	131,072	2048	8192	–
H8	131,072	–	8192	–	–
	$n_0 = 0.1$ cycles/m		$\Omega_0 = 1$ cycles/m		

$$h(x) = \sum_{i=0}^{N} \sqrt{\Delta n.} \cdot 2^{k} \cdot 10^{-3} \cdot \left(\frac{n_0}{i \cdot \Delta n}\right) \cos(2\pi \cdot i \cdot x \cdot \Delta n + \phi) \tag{49}$$

where x is the abscissa variable from 0 to L; Δn is $1/L$; $n_{\max}$ is taken as $1/B$; N is $\frac{^{n}\max}{\Delta n}$ or L/B; where $L = 250$, $N = 100$; and constant is denoted as k, depending on ISO road profile classification. It has also assumed integers augmenting from 3 to 9, which corresponds to the profiles from classes H1 to H8 (as shown in Table 3); here, k is assumed to take the value 3 corresponding to the class H1 road profile and $G_d(n_0)$ is taken to be 32, where n_0 is 0.1 cycles/m and φ_i is the random phase angle following an uniform probabilistic distribution within range 0 to 2π. In this work, four road conditions (H1-H4) are considered, which is shown in Figs. 4, 5, 6 and 7.

The simulation of different road categories is obtained by MATLAB/Simulink of the previously discussed ISO 8608 random input. This road profiles are simulated for five different speeds, viz. 40 kmph, 60 kmph, 80 kmph, 90 kmph and 100 kmph. Figure 4 shows H1 road condition, where unevenness of the random-type road with maximum magnitude is 0.03 m in an upward direction at 40 km/h and − .035 m in a downward direction at 100 km/h. Similarly, H2, H3 and H4 road categories have a maximum magnitude 0.06 m, 0.12 m and 0.24 m, respectively, in an upward direction at 40 km/h and − 0.65 m, − 0.12 m and − 0.24 m, in a downward direction at 100 km/h.

Table 3 k values for ISO road roughness classification

Road class		k
Lower limit	Upper limit	
H1	H2	3
H2	H3	4
H3	H4	5
H4	H5	6
H5	H6	7
H6	H7	8
H7	H8	9

5 Results and Discussion

The model is exhibited first on sine wave input from start to end and runs for 20 s with zero initial conditions. The amplitude of sine wave is 0.01 m, and the input frequency is taken 50 Hz. In Fig. 8, the magnitude of the vertical acceleration is 0.8 m/s^2 in un-damped condition, whereas acceleration amplitude has been reduced to 0.7 m/s^2 in damped condition. Thus, the magnitude of vibration has been reduced from 12 to 15% due to the impact of structural damping of the vehicle chassis. Similarly, in Fig. 9, the amplitude of pitching acceleration has been reduced up to 20% due to the influence of structural damping. However, frequencies in both conditions have almost similar values. The dead load on the center of chassis has been taken around 500 N, which lies in the downward direction.

The RMS acceleration responses of truck chassis at the first three modes at sine wave condition are shown in Fig. 10a–c. In these figures, the un-damped condition of the beam is superimposed over damped condition. The frequencies of the first three modes are 34.4 Hz, 86.89 Hz and 174.8 Hz, respectively. Figure 10a–c shows that the acceleration magnitude is continuously reduced, whereas maximum acceleration is found at the first mode, and minimum

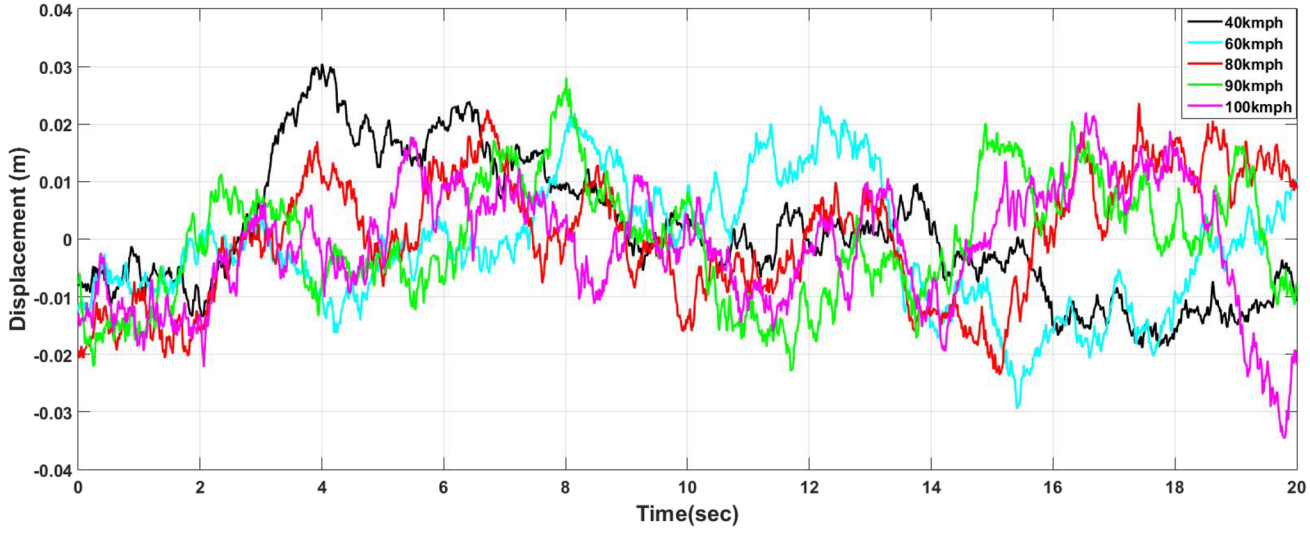


Fig. 4 Random road profile of H1 road category at 40 kmph, 60 kmph, 80 kmph, 90 kmph and 100 kmph

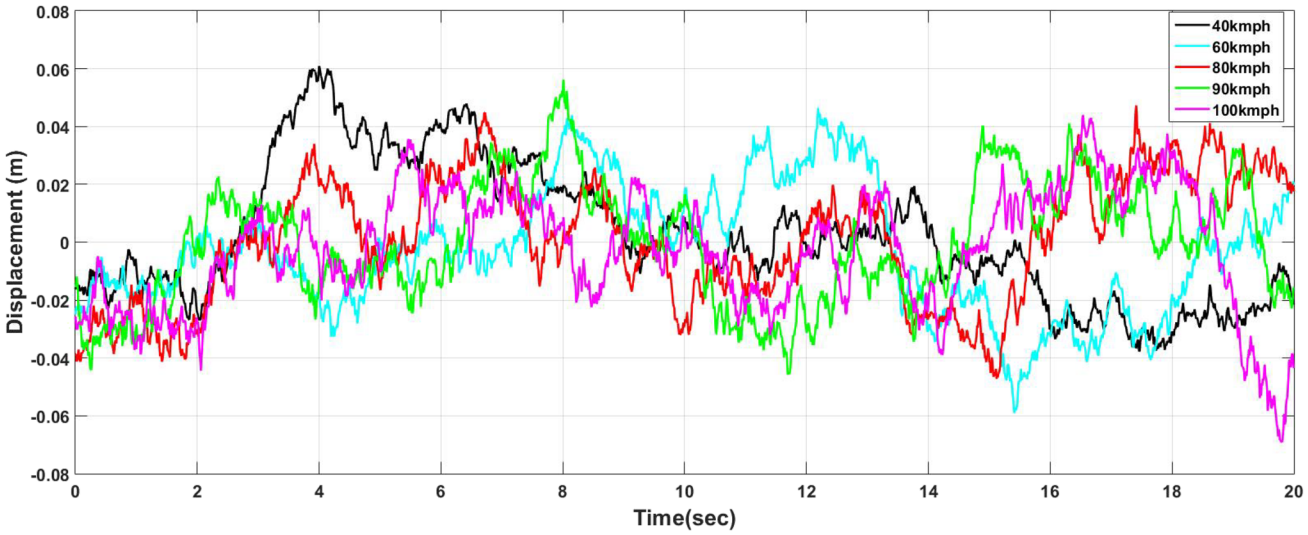


Fig. 5 Random road profile of H2 road category at 40 kmph, 60 kmph, 80 kmph, 90 kmph and 100 kmph

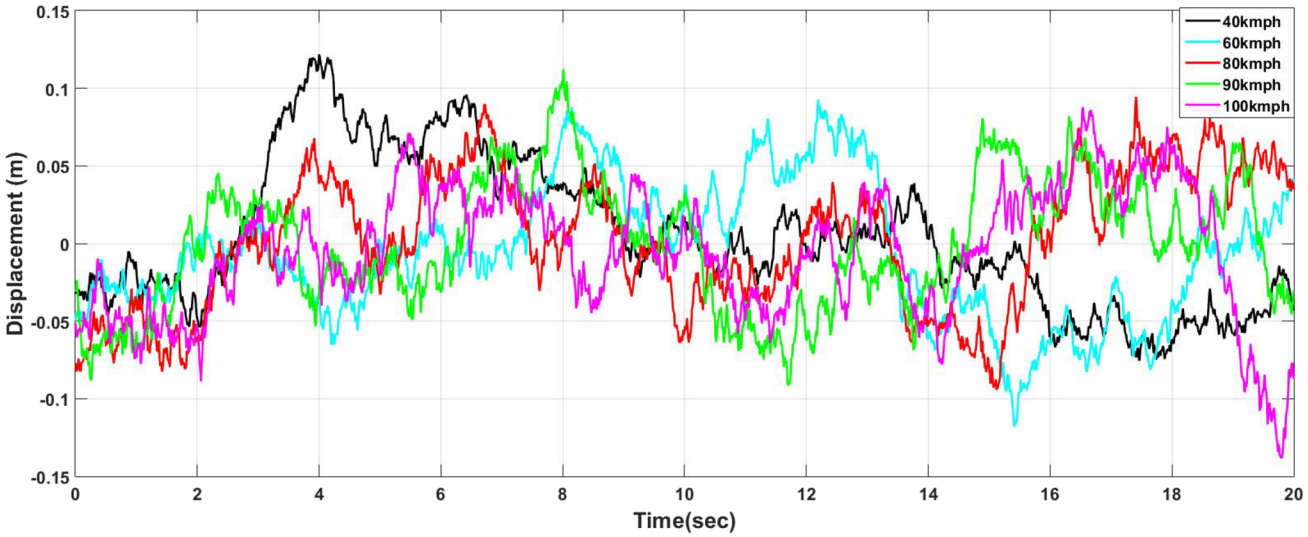


Fig. 6 Random road profile of H3 road category at 40 kmph, 60 kmph, 80 kmph, 90 kmph and 100 kmph

acceleration is in the third mode. The influence of internal damping is presented in terms of RMS acceleration response, whereas the amplitude has reduced due to the effect of internal damping at the same frequency.

The ride comfort analysis may be evaluated through RMS acceleration of the vehicle chassis. Figures 11, 12, 13 and 14 present the RMS acceleration response of the truck CG for four different classes of roads (H1–H4) as prescribed by ISO 8608. The H1 is the best class, and H4 represents the worst road class in these four classes. The model is simulated for each road class at different speeds ranging from 40 to 100 kmph with and without consideration of the internal damping of the chassis. Figures 11, 12, 13 and 14 show that the magnitude of vertical acceleration increases with vehicle speed and with road class. Figure 11a demonstrates the truck response neglecting the chassis damping, where maximum RMS acceleration magnitude is observed as 0.65 m/s^2 at 100 km/h and 0.25 m/s^2 at 40 km/h on H1 class road. This peak acceleration magnitude gets reduced up to 0.58 m/s^2 at 100 km/h and 0.22 m/s^2 at 40 km/h with due consideration to chassis structural damping, as shown in Fig. 11b. A similar kind of result has been noticed in other plots shown in Figs. 12, 13 and 14. Thus, it is evident from the above results that internal structural damping of the base structure of heavy vehicle plays an important role

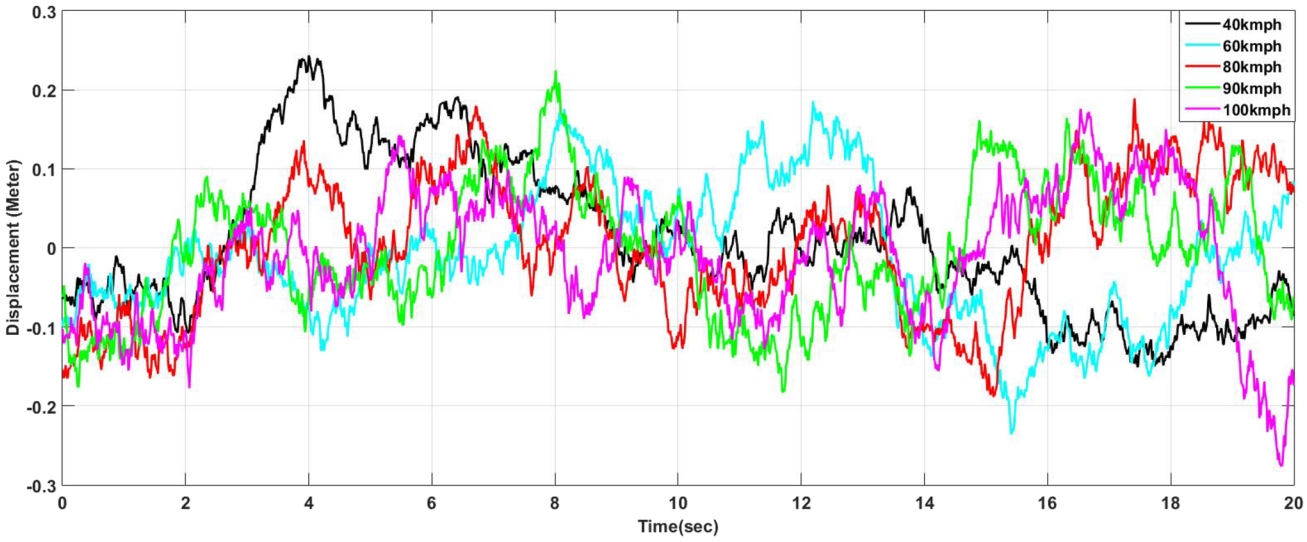


Fig. 7 Random road profile of H4 road category at 40 kmph, 60 kmph, 80 kmph, 90 kmph and 100 kmph

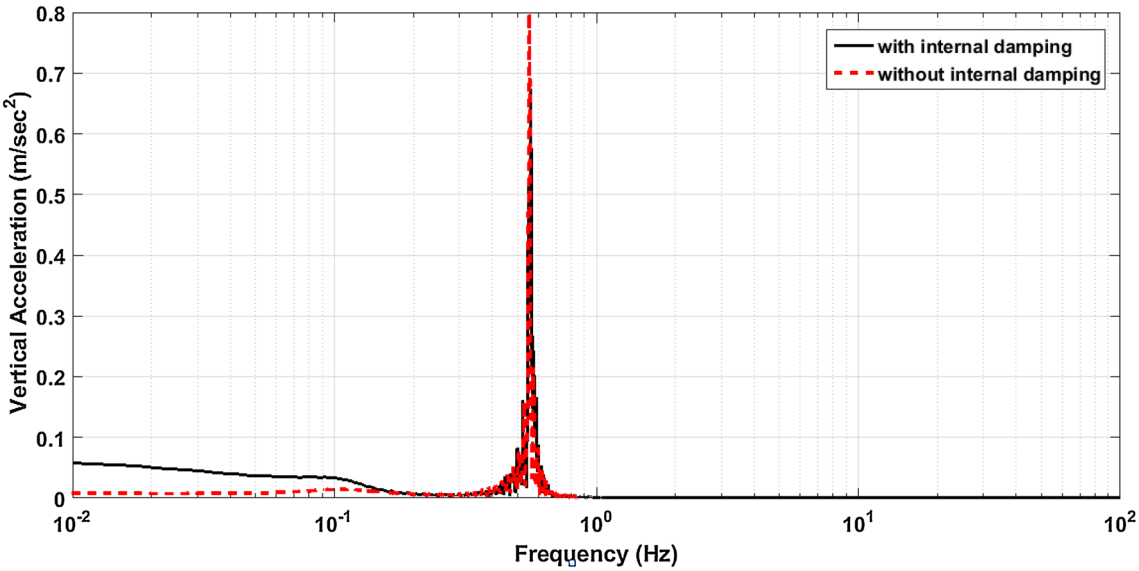


Fig. 8 Vertical response of vertical acceleration of chassis at steady-state input

and it needs to be included in the dynamical model of the vehicle to get the true dynamic behavior of the vehicle.

It has been observed that from the RMS curves that the H1 road condition is perfect for driving the vehicle for long duration, whereas it is not advisable to drive vehicles for long duration on H4 road condition. Indian roads lie between H2 and H3 conditions. So, a driver can comfortably drive the given vehicle at 60–80 km/h in single stretch without a break.

6 Conclusions

Numerous models have been constructed for the heavy vehicle system, but most of the models are developed with the lumping process. In this work, an analytical framework for truck chassis has been developed through classical mechanics and through extended Lagrangian approach. The following conclusion has been drawn from present research work:

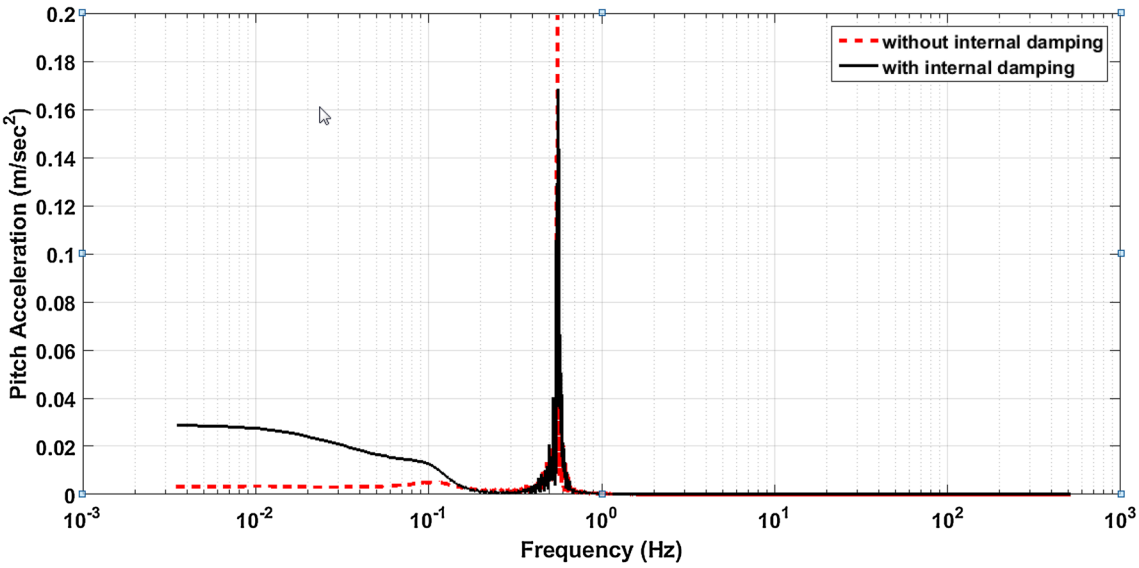


Fig. 9 Response of pitch acceleration of chassis at steady-state input

- In the classical mechanics approach, the idea of lumped–distributed modeling has been explored within the framework of bond graphs.
- The Lagrange variation of truck chassis, including flexural behavior has been constructed, where it was integrated with lumped suspension system. However, the structural damping effects were also involved in terms of fifth-order partial differential term in generalized equation of motion. The detailed derivation for fifth-order differential term is presented in "Appendix 1."
- The integrated bond graph model of flexural beam was the extended form of the analytical model. Results have been presented under the two types of road conditions, i.e., sine wave input and random road input. In sine wave road conditions, the vertical response of flexural beam is reduced up to 12–15% and pitch response of beam is reduced up to 20% due to the structural damping effect. In random road conditions, four types of road category (H1, H2, H3, H4) have been used at four different speeds (40 kmph, 60 kmph, 80 kmph and 100 kmph).
- It is depicted from the result that maximum acceleration at maximum speed (100 kmph) of chassis has been increased up to 0.5 m/s^2 in H1–H2 conditions, 1 m/s^2 in H2–H3 conditions and 2.25 m/s^2 in H3–H4 conditions. Similarly, the solution has been extended for structural or internal damping, whereas maximum acceleration at maximum speed has been reduced up to 10%, 5%, 7% and 5% (approx.) in H1, H2, H3 and H4 road conditions, respectively.

Appendix 1

This appendix gives a brief introduction to the variational formulation of dynamics of continuous systems. Consider for simplicity a one-dimensional continuous system with the field variable $w(x, t)$ which uniquely represents the configuration of the system at any time t. In the course of temporal evolution of the system, let the configurations at two time instants $t = t_1$ and $t = t_2$ be recorded as, respectively, $w(x, t_1)$ and $w(x, t_2)$. The actual path is that any infinitesimal variation over that path should leave the value of I unchanged (Hagedorn and DasGupta 2007). Now the action integral may be expressed as

$$I = \int_{t_0}^{t_1} \int_{x_0}^{x_1} \delta L(\ldots)\, \mathrm{d}t\mathrm{d}x \tag{50}$$

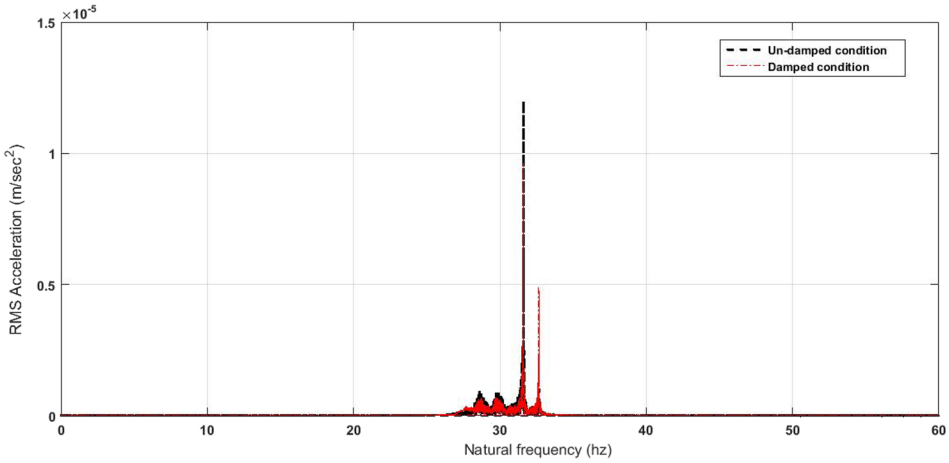
Using the extremization condition of Eq. (50), one may obtain

$$\int_{t_0}^{t_1} \int_{x_0}^{x_1} \delta \hat{L}\left(w, w', \dot{w}, w'', t\right) \mathrm{d}t\mathrm{d}x = 0 \tag{51}$$

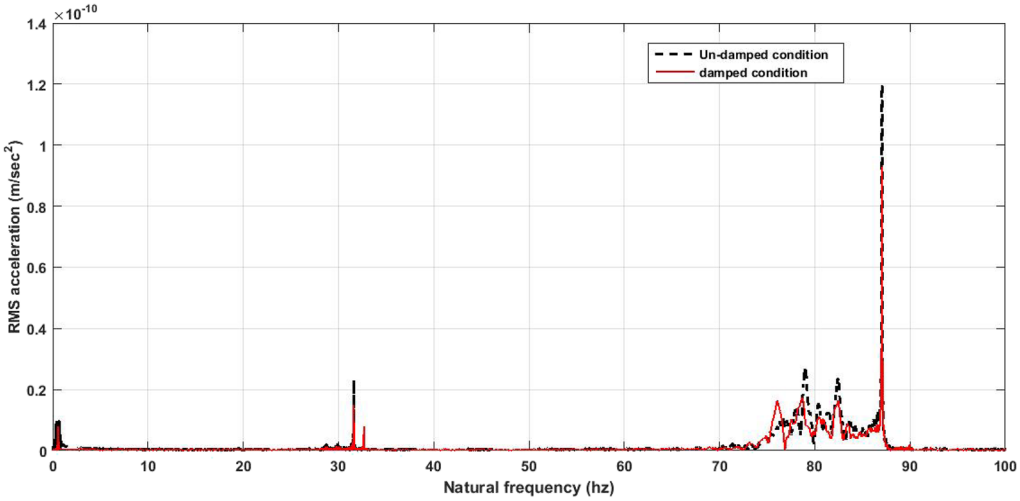
One may write Lagrangian density variation,

$$\delta \hat{L} = \frac{\partial \hat{L}}{\partial w}\delta w + \frac{\partial \hat{L}}{\partial \dot{w}}\delta \dot{w} + \frac{\partial \hat{L}}{\partial w'}\delta w' + \frac{\partial \hat{L}}{\partial w''}\delta w'' \tag{52}$$

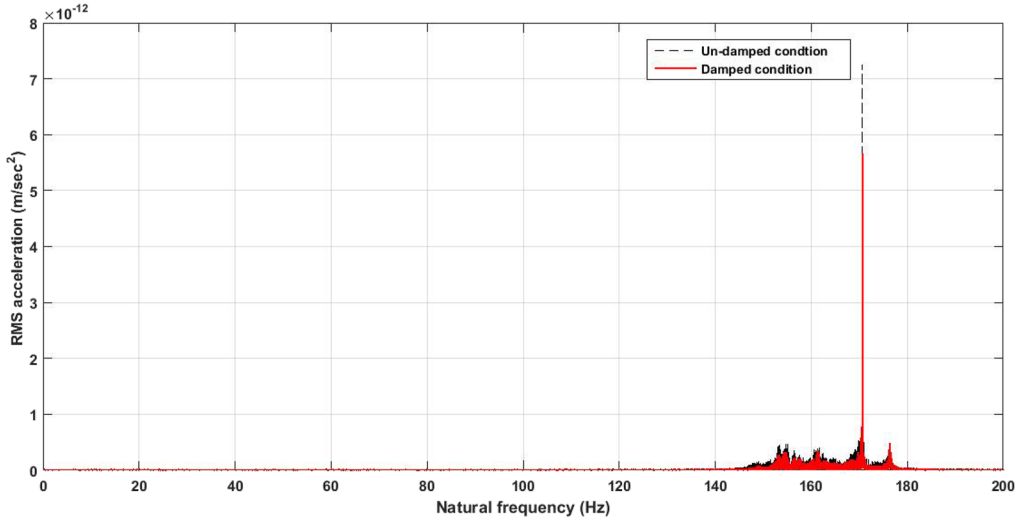
Introducing Eqs. (51) and (52), one may have


×10^-5
Un-damped condition
Damped condition
RMS Acceleration (m/sec^2)
Natural frequency (hz)

(a)


×10^-10
Un-damped condition
damped condition
RMS acceleration (m/sec^2)
Natural frequency (hz)

(b)


×10^-12
Un-damped condtion
Damped condition
RMS acceleration (m/sec^2)
Natural frequency (Hz)

(c)

◂**Fig. 10** RMS acceleration response of different modes of chassis under sine wave condition: **a** first mode = 34.4 Hz; **b** second mode = 86.89 Hz; **c** third mode frequency = 174.8 Hz

$$\int_{t_0}^{t_1}\int_{x_0}^{x_1}\left[\frac{\partial\hat{L}}{\partial w}\delta w+\frac{\partial\hat{L}}{\partial\dot{w}}\delta\dot{w}+\frac{\partial\hat{L}}{\partial w'}\delta w'+\frac{\partial\hat{L}}{\partial w''}\delta w''\right]\mathrm{d}t\mathrm{d}x=0 \tag{53}$$

The next step is to transform the integrand in Eq. (52) into one containing only δw, i.e., one that is free of $\delta\dot{w},\delta w',\delta w'',\delta\dot{w}'$. This can be accomplished by integration by parts, with respect to both space and time.

$$\begin{aligned}&\int_0^l\frac{\partial\hat{L}}{\partial w}\delta w\Big|_{t_1}^{t_2}\mathrm{d}x+\int_{t_0}^{t_1}\left[\frac{\partial\hat{L}}{\partial w''}\delta w''+\left\{\frac{\partial\hat{L}}{\partial w'}-\frac{\partial}{\partial x}\left(\frac{\partial\hat{L}}{\partial w''}\right)\right\}\delta w\right]\Bigg|_0^l\mathrm{d}t\\&+\int_{t_0}^{t_1}\int_{x_0}^{x_1}\left[\frac{\partial\hat{L}}{\partial w}-\frac{\partial}{\partial t}\left(\frac{\partial\hat{L}}{\partial\dot{w}}\right)-\frac{\partial}{\partial x}\left(\frac{\partial\hat{L}}{\partial w'}\right)+\frac{\partial^2}{\partial x^2}\left(\frac{\partial\hat{L}}{\partial w''}\right)+\frac{\partial^2}{\partial x\partial t}\left(\frac{\partial\hat{L}}{\partial\dot{w}'}\right)\right]\delta w\,\mathrm{d}t\mathrm{d}x=0\end{aligned} \tag{54}$$

But the vertical displacement δw is arbitrary by definition, which implies that it can be assigned values at will provide these values are compatible with the system constraint, such as geometric conditions at the end points. Let $\delta w=\delta w'=0$ at $x=0$ and $x=l$. One may conclude that Eq. (54) can be satisfied for all values of δw in the open domain $0<x<l$.

Thus,

$$\frac{\partial\hat{L}}{\partial w}-\frac{\partial}{\partial t}\left(\frac{\partial\hat{L}}{\partial\dot{w}}\right)-\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial\hat{L}}{\partial w'}\right)+\frac{\partial^2}{\partial x^2}\left(\frac{\partial\hat{L}}{\partial w''}\right)+\frac{\partial^2}{\partial x\partial t}\left(\frac{\partial\hat{L}}{\partial\dot{w}'}\right)=0 \tag{55}$$

Equation (55) represents the *Lagrange differential equation of motion* for this continuous system and must be satisfied at every point of open domain $0<x<l$.

$$\frac{\partial\hat{L}}{\partial w''}\equiv 0\quad\text{or}\quad\delta w|_{x=0}\equiv 0 \tag{56a}$$

$$\left.\frac{\partial\hat{L}}{\partial w''}\right|_{x=l}\equiv 0\quad\text{or}\quad\delta w|_{x=l}\equiv 0 \tag{56b}$$

$$\left.\frac{\partial\hat{L}}{\partial w'}-\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial\hat{L}}{\partial w''}\right)\right|_{x=0}=0\quad\text{or}\quad\delta w|_{x=0}\equiv 0 \tag{57a}$$

$$\left.\frac{\partial\hat{L}}{\partial w'}-\frac{d}{dx}\left(\frac{\partial\hat{L}}{\partial w''}\right)\right|_{x=l}=0\quad\text{or}\quad\delta w|_{x=l}\equiv 0 \tag{57b}$$

Equations (56–57) are known as the boundary conditions and only two boundary conditions must be satisfied at either end, one from condition Eq. (56) and other from Eq. (57), which shows that clearly the selection is not arbitrary, but it must reflect physical conditions at two ends.

Appendix 2

In this formulation, the vertical translation of a uniform free–free beam is considered. If the internal damping of the material is involved in the system, then the strain displacement relationship or Hook's law equation can be written as Karnopp et al. (2012) and Hagedorn and DasGupta (2007).

$$\sigma(x,t)=E\varepsilon(x,t)+\mu_I\varepsilon(x,t) \tag{58}$$

The strain displacement relation at any height from the plane of the neutral fibers can be written as

$$\varepsilon(x,t)=\frac{zw''(x,t)}{\left[1+w'^2(x,t)\right]^{\frac{3}{2}}}\approx -zw''(x,t)\quad(\text{Assuming } w'<<1) \tag{59}$$

The bending moment at any section (from Fig. 15) can be written as

$$M(x,t)=-\int_{h/2}^{-h/2}z\sigma(x,t)\mathrm{d}A=-\left(E\varepsilon(x,t)+\mu_I\varepsilon(x,t)\right)\int_{h/2}^{-h/2}z^2\mathrm{d}A \tag{60}$$

$$\Rightarrow\quad -\left(EIw''(x,t)+\mu_I I\dot{w}''(x,t)\right) \tag{61}$$

where I is the moment of inertia cross section of beam about the neutral axis. The equation of the vertical force of an infinitesimal beam is

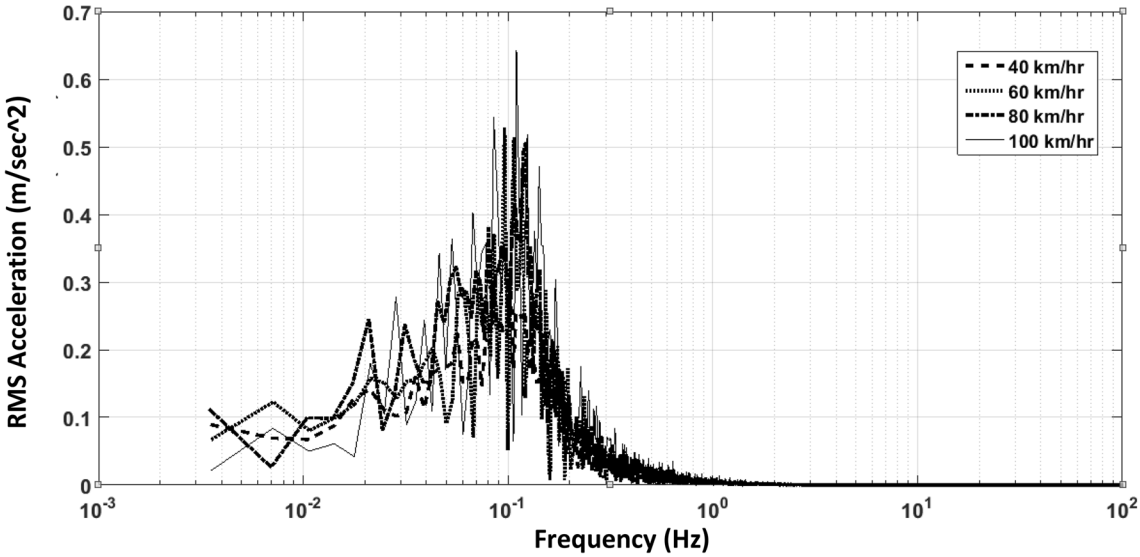
$$(\rho A\mathrm{d}x)\ddot{w}(x,t)=p(x,t)\mathrm{d}x+(V+\mathrm{d}V)\cos(\psi+\mathrm{d}\psi)-V\cos\psi$$

or

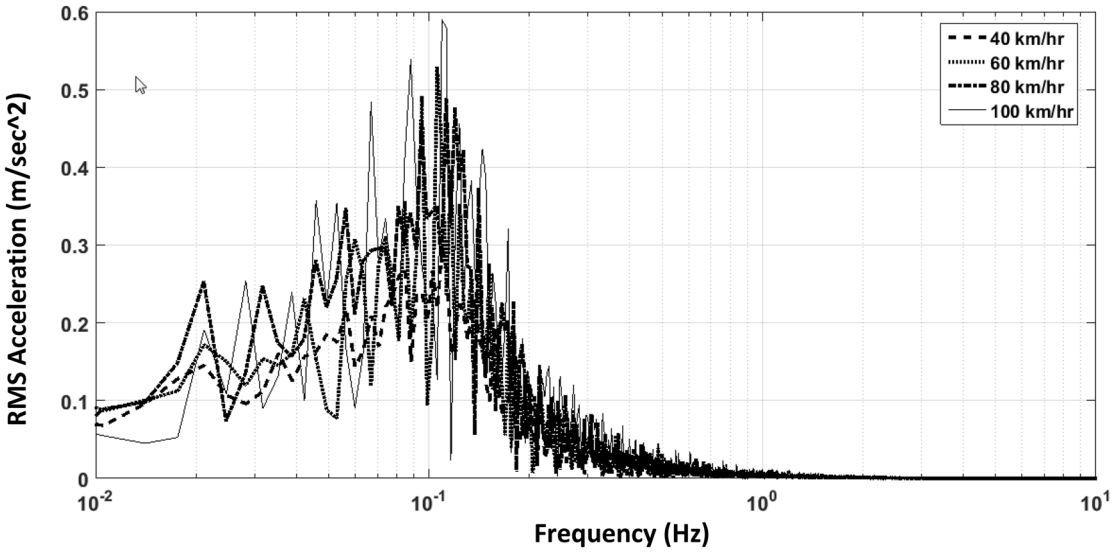
$$pA\ddot{w}(x,t)=p(x,t)+V' \tag{62}$$

The rotational dynamics of infinitesimal beam may be written as

$$(\rho I\mathrm{d}x)\ddot{\psi}=(M+\mathrm{d}M)-M(V+\mathrm{d}V)\frac{\mathrm{d}x}{2}+V\frac{\mathrm{d}x}{2}$$



(a)



(b)

Fig. 11 **a** Response of chassis acceleration at various random road inputs in H1 road as per ISO 8608. **b** Response of chassis acceleration consideration of internal damping at different random inputs in H1 road condition as per ISO 8608

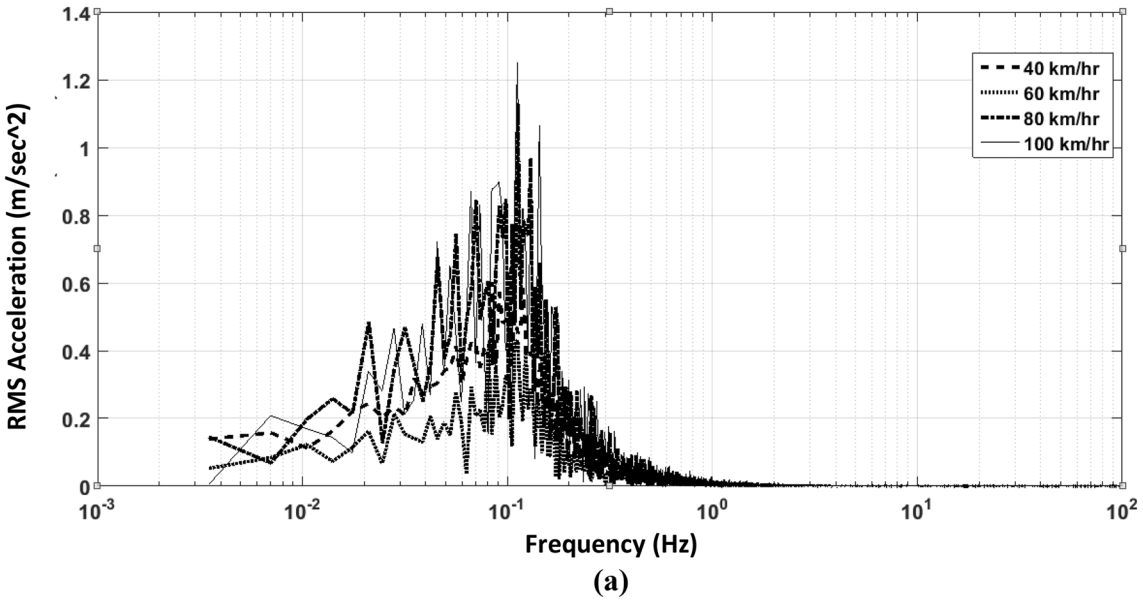
or

$$\rho I\ddot{\psi} = M' + V \tag{63}$$

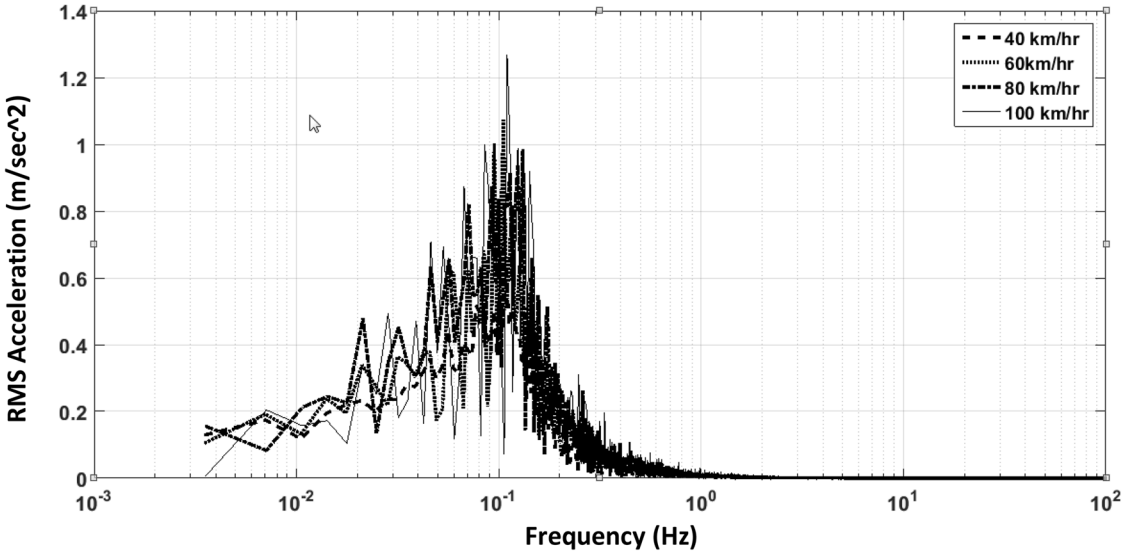
One may write $\psi = \frac{w'(x,t)}{1+w'^2(x,t)} \approx w'(x,t)$Thus, from Eqs. (62) and (63), one may have

$$\rho A\ddot{w}(x,t) - \rho I\ddot{w}''(x,t) + \mu_I I\dot{w}''''(x,t) + EIw''''(x,t) - p(x,t) = 0 \tag{64}$$

The third term of Eq. (64) represents the structural damping or internal damping of the beam.



(a)



(b)

Fig. 12 **a** Response of chassis acceleration at various random road inputs in H2 road as per ISO 8608. **b** Response of chassis acceleration consideration of internal damping at different random inputs in H2 road condition as per ISO 8608

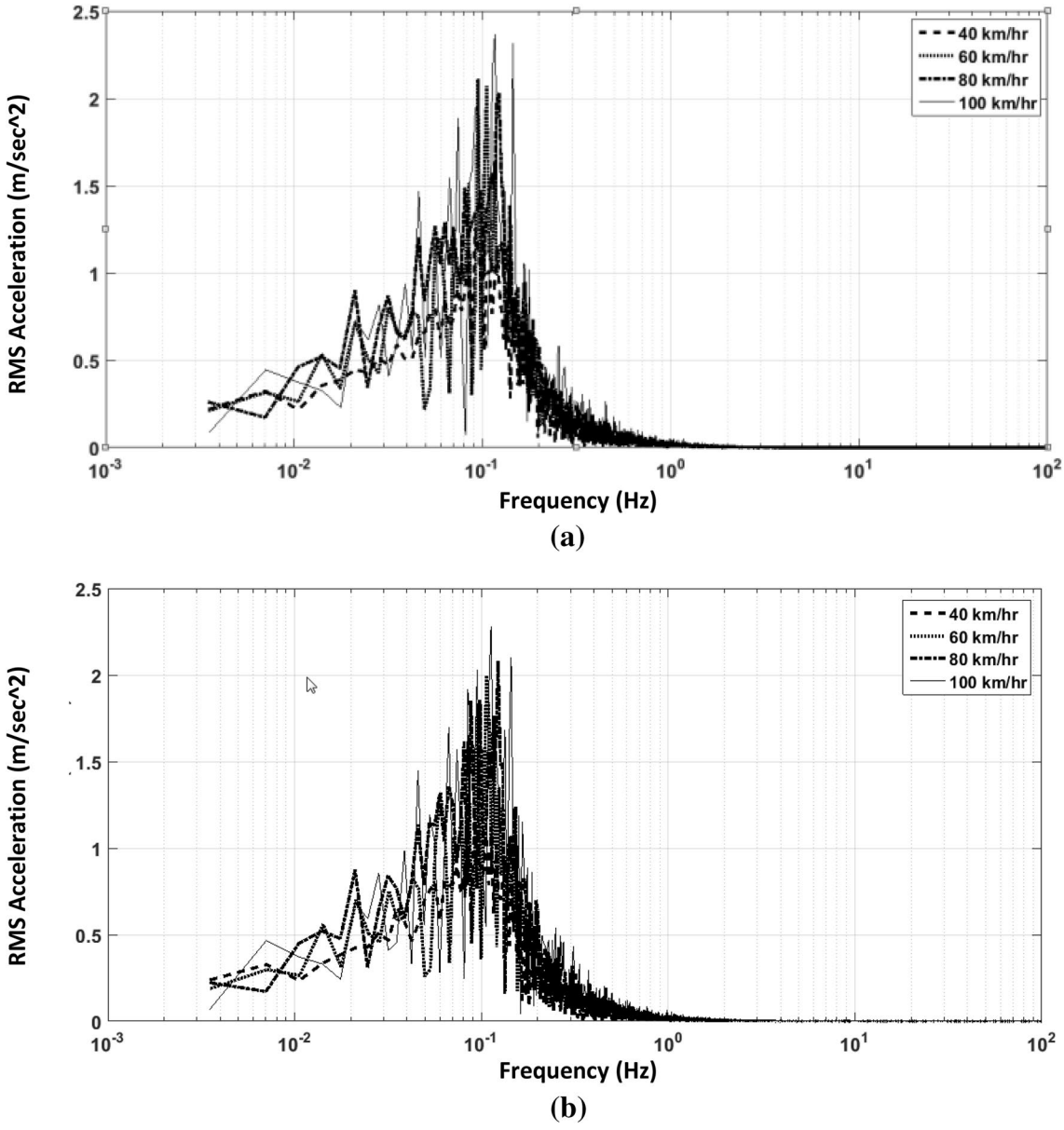


Fig. 13 **a** Response of chassis acceleration at various random road inputs in H3 road as per ISO 8608. **b** Response of chassis acceleration consideration of internal damping at different random inputs in H3 road condition as per ISO 8608

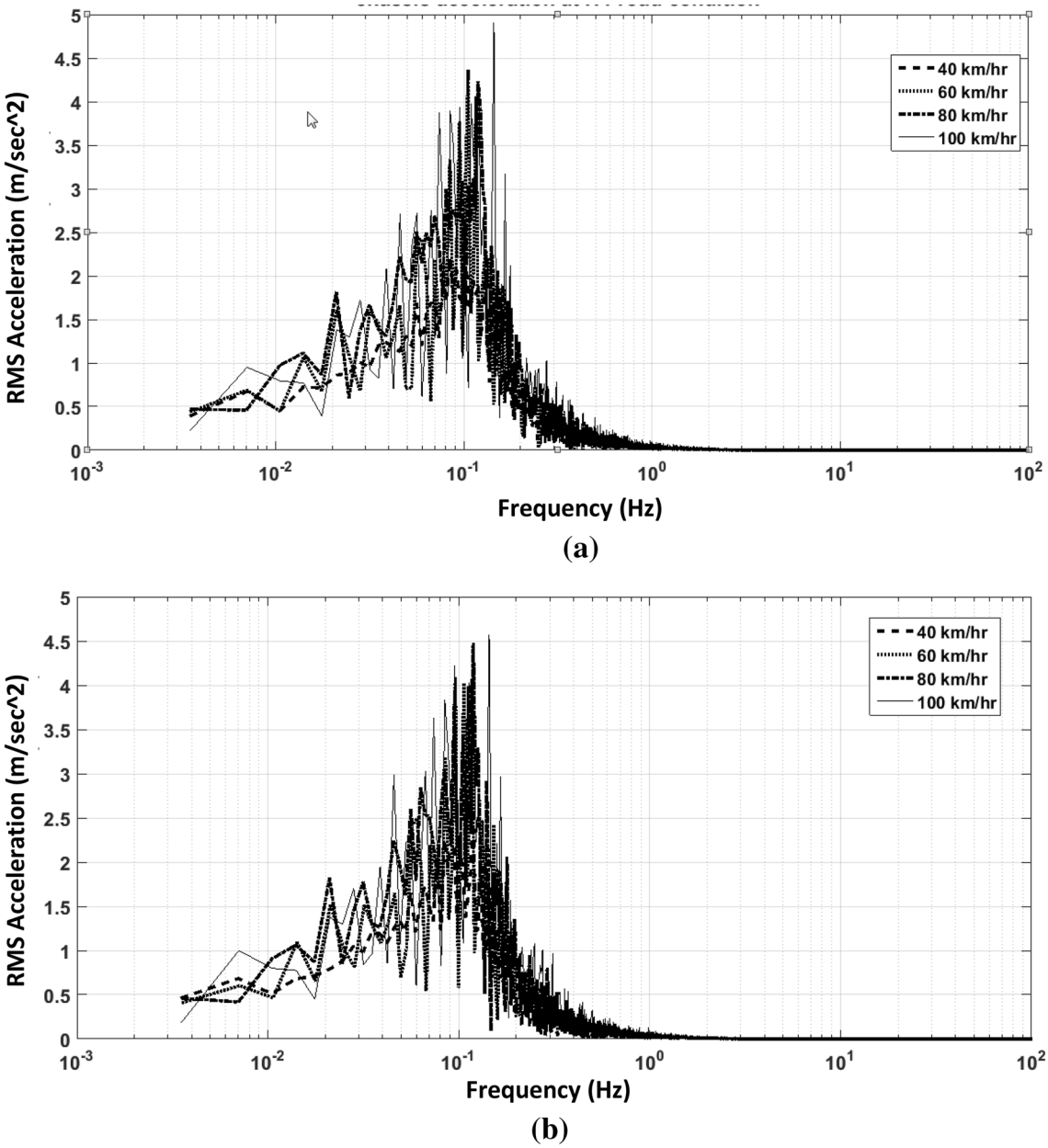


Fig. 14 **a** Response of chassis acceleration at various random road inputs in H4 road as per ISO 8608. **b** Response of chassis acceleration consideration of internal damping at different random inputs in H4 road condition as per ISO 8608

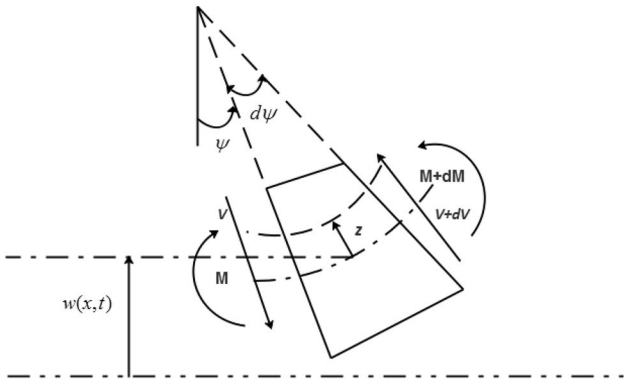


Fig. 15 Infinitesimal element of a deflected beam

References

Agostinacchio M, Ciampa D, Olita S (2014) The vibrations induced by surface irregularities in road pavements—a Matlab® approach. Eur Transp Res Rev 6(3):267–275

Gupta A, Rastogi V (2016) Effects of various road conditions on dynamic behavior of heavy road vehicle. Procedia Eng 144:1129–1137

Hagedorn P, DasGupta A (2007) Vibrations and waves in continuous mechanical systems. Wiley, New York

Hassani S (2013) Mathematical physics: a modern introduction to its foundations. Springer, Berlin

Huang CH, Zeng J (2010) Flexural vibration suppression of car body for high-speed passenger car based on constrained damping layers. J Traffic Transp Eng 10(1):36–41

Ibrahim IM (1996) Effect of frame flexibility on the ride. Vib Heavy Trucks 58(4):709–713

Karnopp DC, Margolis DL, Rosenberg RC (2012) System dynamics: modeling, simulation, and control of mechatronic systems. Wiley, New York

Kim JH, Yim HJ (1994) Influence of chassis flexibility on dynamic behavior of engine mount systems. SAE Paper, 942269

Kumar V, Rastogi V, Pathak P (2017) Simulation for whole-body vibration to assess ride comfort of a low–medium speed railway vehicle. Simulation 93(3):225–236

Margolis D, Edeal D (1989) Modeling and control of large flexible frame vehicles using bond graphs SAE Paper 892488. Society of Automotive Engineers, Warrendale

Meirovitch L (1980) Computational methods in structural dynamics. Springer, Netherlands

Mukherjee A, Karmakar R (2000) Modelling and simulation of engineering system through bond graph, Narosa Publishing House, New Delhi, reprinted by CRC Press for North America and by Alpha Science for Europe

Mukherjee A, Samantaray AK (2006) SYMBOLS-Shakti user's manual. High-Tech Consultants, STEP Indian Institute of Technology, Kharagpur

Rideout DG (2012) Simulating coupled longitudinal, pitch and bounce dynamics of trucks with flexible frames. Mod Mech Eng 2(4):176–189

Rideout D, Khan TM (2010) Flexible truck modelling and investigation of coupling between rigid and flexible dynamics. In: Proceeding of the ICBGM 2010, Orlando, USA

Tomioka T, Takigami T, Suzuki Y (2006) Numerical analysis of three-dimensional flexural vibration of railway vehicle car body. Veh Syst Dyn 44(Supp-1):272–285

Tomioka T, Takigami T, Aida K (2017) Experimental investigations on the damping effect due to passengers on flexural vibrations of railway vehicle car body and basic studies on the mimicry of the effect with simple substitution. Veh Syst Dyn 55(7):995–1011

Tomioko T, Takigami T (2015) Experimental and numerical study on the effect due to passengers on flexural vibrations in railway vehicle car bodies. J Sound Vib 343:1–19

Zhou J, Goodall A, Ren L, Zhang H (2009) Influences of car body vertical flexibility on ride quality of passenger railway vehicles. F J Rail and Rapid Transit 2009:223

Defence Technology xxx (xxxx) xxx

Contents lists available at ScienceDirect

Defence Technology

journal homepage: www.keaipublishing.com/en/journals/defence-technology

Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075

Husain Mehdi*, R.S. Mishra

Department of Mechanical Engineering, Delhi Technological University, Delhi, India

ARTICLE INFO

Article history:
Received 7 January 2020
Received in revised form
4 March 2020
Accepted 22 April 2020
Available online xxx

Keywords:
Residual stress
TIG+FSW
Heat transfer
Micro-hardness
Tensile strength

ABSTRACT

Tungsten inert gas (TIG) welding is the most commonly used joining process for aluminum alloy for AA6061 and AA7075 which are highly demanded in the aerospace engineering and the automobile sector, but there are some defects occur during TIG welding like micro-crack, coarse grain structure, and porosity. To improve these defects, the TIG welded joint is processed using friction stir processing (FSP).

This paper presents the effect of friction stir processing on TIG welding with filler ER4043 and ER 5356 for dissimilar aluminum alloy AA6061 and AA7075. The mechanical characterization, finite element formulation and mathematical equations of heat transfer of TIG + FSP welded joints are investigated using ANSYS Fluent software by adjusting process parameters of FSP. The results show that the maximum compressive residual stress 73 MPa was obtained at the fusion zone (FZ) of the TIG weldment with filler ER4043, whereas minimum compressive residual stress 37 MPa was obtained at stir zone (SZ) of the TIG + FSP with filler 5356. The maximum heat flux 5.33 × 106 W/m2 and temperature 515 °C have observed at tool rotation 1600 rpm with a feed rate of 63 mm/min. These results give a satisfactory measure of confidence in the fidelity of the simulation

© 2020 China Ordnance Society. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Friction stir welding (FSW) is a solid-state joining process that uses a non-consumable tool to join similar or dissimilar materials. The FSW does not involve the melting of parent material, which have many advantages on conventional fusion welding processes [1]. The FSW process involves the heat generation, which produced by pressure and friction between the tool and the workpiece. As a result of the generated heat, the weldment surfaces in contact soften become plasticized [2]. The modern development in the automobile sector and defense sector have been transformed from conventional materials to light material such as aluminum alloy. Due to excellent physical and mechanical properties of Al-alloy such as high corrosion resistance, low density, high strength to weight ratio and high thermal conductivity, it is mostly used for making various components such as military aircraft, rocket, and rocket launcher, axle shafts, rims, bumpers, and car bodies. FSW is used in the fabrication of automotive and shipbuilding components, aircraft structural components and reducing structural weight [3–7]. To study material flow and temperature distribution in friction stir welding of similar or dissimilar aluminum alloys many authors developed the CFD model to analyze the plastic flow, heat transfer and heat generation in the welding process [8–10], and the effect of tool profile on the material flow [11–13]. Due to low heat generation in the friction stir welding, the residual stresses are also low in the weldment. The transverse force of friction stir welding tools plays an important role in stress measurement [14–16]. The residual stresses in the weldment have a big impact on the performance of the welded structure. So far, information on the residual stress distribution of FSW has been limited [17]. The effects of tool feed rate on residual stresses of FSW of Al-alloy joints were studied with synchrotron X-ray measurement and analyze the residual stresses in longitudinal and transverse directions [16]. Many researchers have used finite element methods for the residual stress distribution in FSW. The computational fluid dynamic supports the heat transfer model and heat source of friction stir welding. The residual stress depends on the empathetic of heat transfer during FSW, although the correct prediction of temperature distribution depends on the heat source [18–20]. FSW can affect the performance of the residual stresses in

* Corresponding author.
E-mail address: husainmehdi4u@gmail.com (H. Mehdi).
Peer review under responsibility of China Ordnance Society

https://doi.org/10.1016/j.dt.2020.04.014
2214-9147/© 2020 China Ordnance Society. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Please cite this article as: Mehdi H, Mishra RS, Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, Defence Technology, https://doi.org/10.1016/j.dt.2020.04.014

Nomenclature

u	Plastic flow velocity	τ	Max. Shear stress
μ	Non-Newtonian viscosity	η	Mechanical efficiency
p	Pressure	ω	Angular velocity
ρ	Density	P_N	Normal pressure
U_1	Welding velocity	q_1	Total heat generation rate
z	Zener Hollomon parameter	T_a	Ambient temperature
a and n	Material constant	σ	5.67×10^{-16} J/(K^4 m^2 s)
Q	Temperature activation energy	ε	Convection heat transfer coefficient
Z	Temperature effective strain rate	h_b	Heat transfer coefficient
K	Thermal conductivity	σ_2	Friction stress
C_p	Specific heat	σ_1	Tensile stress
S_i	Heat generation rate	d	Main grain diameter
h_t	Emissivity at top surface	k	Constant
r	Radial distance from tool axis	R_S	Shoulder radius
A_r	Small area on tool pin	R_P	Tool pin radius
V	Control volume under A_r	h_{b0}	Heat transfer parameter of bottom.surface
θ	Angle of negative x-axis	μ_f	Spatially variable fraction slip.between the work piece and tool
		S_b	Heat generation rate due to plastic.deformation

the weldment and the processing parameters such as rotating tool may affect the development of residual stresses [21,22]. The mechanical properties, grain size are dependent on tool geometry, processing parameters and chemical composition of alloying elements. A new grain structure could be formed in friction stir welding by controlling the process parameter and heat input [23–27]. Comparison of residual stresses of TIG and FSW has been done for AA6061 and AA7075, and found that the longitudinal residual stress is larger than the traverse residual stress in the welded joint [28]. There are no successful investigations have been done to find out residual stress and heat transfer in TIG welded joint with friction stir processing. The small plate having the dimension of 150 mm × 25 mm × 0.8 mm, have been investigated, as a result, the properties at the beginning and end of the welded joints can be significantly different. The tensile strength first increases and then decreases with the growth of rotational speed. Three layers i.e. top, middle, and bottom, have considered measuring the microhardness of the welded joint. The top layer of the weld has a higher recrystallization degree and dynamic recovery and the bottom layer is strongly stirred under mechanical action of the tool pin such that each layer of welded joint shows the different mechanical properties [29]. A new welding approach of TIG+FSP was successfully applied to the AA2024 to enhance the mechanical properties of the TIG welded joints. They conclude that the defects and porosities in the TIG welded joints are completely reduced by the FSP process and modified the microstructure and mechanical properties of the TIG+FSP welded joint [30]. The application of FSP on the TIG welded joint improved the ductility and tensile strength of the FSW and TIG welded joint of AA5083-H111 and results revealed best-reduced dimple size [31].

In this work, the effect of FSP on TIG welding for dissimilar Al-alloy AA6061 and AA7075 with filler wire ER4043 and ER5356 was carried out and analyze the mechanical properties, residual stress distribution and heat transfer in (TIG+FSP) welded joint. The present study was focus on the improvement of mechanical properties and the residual stress distribution of the TIG-welded joint with different filler wire (ER4043 and ER5356) by using a friction stir processing approach. This topic is carefully selected after noticing that there is little literature review available on the TIG+FSP welding approach. The outcomes of this paper will give the new approach for enhancing the mechanical properties and heat transfer of TIG+FSP welded joints which play a huge role in welded structures.

2. Materials and methods

Two AA6061 and AA7075 aluminum alloy plates having dimensions (150 mm × 40 mm × 6.2 mm) were welded together using the TIG+FSP welding approach. V butt joints of 45° were prepared for TIG welding with the help of milling cutter and power hacksaw. There are two types of samples were prepared for TIG welded joint. In the first sample, ER4043 filler wire was used and for the second sample, ER 5356 filler wire was used. The chemical composition of filler rod and base materials are shown in Table 1. The processing parameters and filler wire are strappingly affected by the mechanical properties of the welded joint. In this work, the filler wire of ER4043 and ER5356 of diameter 2.4 mm were used to join a single V groove plate at voltage 22 V and current 50 amp for all welding processes. Argon (25 L/min) is used as a shielding gas and travel speed 3.6 mm/s to fabricate the joint.

After TIG welding, the FSP technique was used on TIG weldment to improve the welding quality and mechanical properties. The non-consumable H13 steel tool with pin diameter, shoulder diameter and a pin length of 3 mm, 19.5 mm, and 5.5 mm respectively was used. 0° tilt angle was used during friction stir processing on the AA6061 and AA7075 plate at a rotational speed of 800 rpm−1600 rpm and a traverse speed of 63 mm/min is shown in Fig. 1(a). The tool nomenclature is shown in Fig. 1(b). The processing parameters of TIG welding and friction stir processing are shown in Table 2 and Table 3.

Fig. 2 demonstrate the dimension of the tensile sub test specimen as per ASTM E8. The tensile test specimens were sliced and machined from the welded joint using a milling cutter and shaper machine. Single-pass FSP is used to fabricate TIG+FSP welded joint. Three sub tensile specimens were tested on a UTM machine at

Table 1
Chemical composition of filler wire and parent material.

Material	Si	Cu	Fe	Zn	Mg	Mn	Cr	Ti	Al
ER 4043	5.300	0.025	0.80	0.10	0.05	–	–	0.02	Bal.
ER 5356	0.025	0.015	0.15	0.01	4.90	0.14	0.11	0.09	Bal.
AA7075	0.580	1.200	0.35	5.10	2.10	0.12	0.20	0.05	Bal.
AA6061	0.400–0.800	0.200	0.70	0.25	0.80	0.10	0.20	0.15	Bal.

Please cite this article as: Mehdi H, Mishra RS, Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, Defence Technology, https://doi.org/10.1016/j.dt.2020.04.014

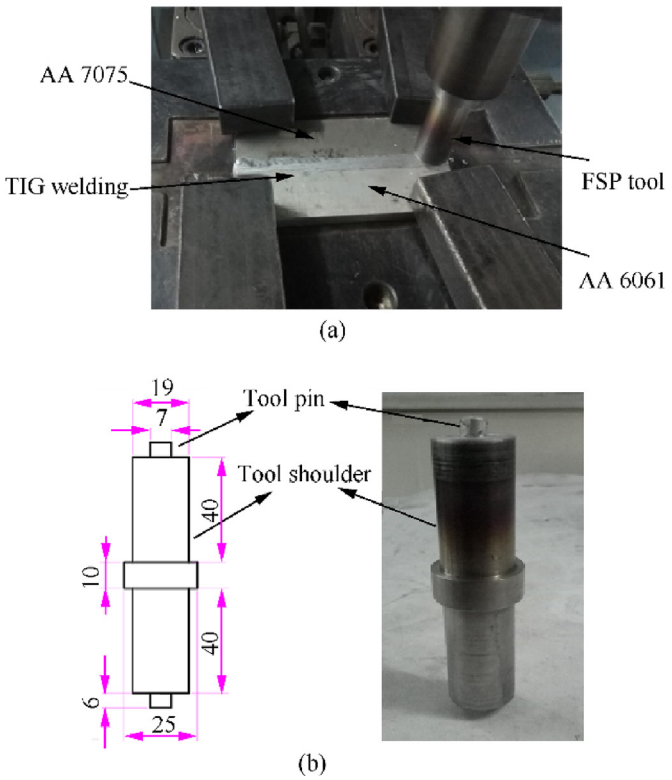


Fig. 1. (a) FSP process after TIG welding, (b) Friction stir processing tool of H13 tool steel.

room temperature and the average of these three results is taken. Vicker microhardness machine was used for measuring the hardness across the welded joints with a load of 100 g and dwell time 30 s. Microstructure observations were carried out using a scanning electron microscope (SEM) for TIG and TIG+FSP welded joints.

3. Numerical modeling [10]

For incompressible flow, the continuity equation in the direction of x, y and z is given by

$$\frac{\partial u_i}{\partial x_i} = 0 \tag{1}$$

The momentum conservation equation of the heat source is given as

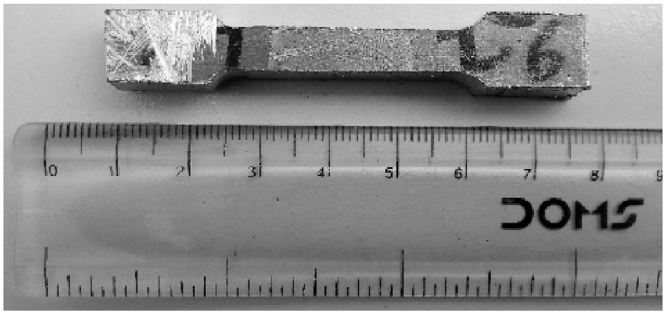


Fig. 2. Dimension of tensile test specimen as per ASTM E8.

$$\rho \frac{\partial u_I\, u_j}{\partial x_i} = -\frac{\partial P}{\partial x_j} + \frac{\partial}{\partial x_I}\left(\mu \frac{\partial U_j}{\partial x_i} + \mu \frac{\partial U_i}{\partial x_j}\right) - \rho\, U_1 \left(\frac{\partial U_j}{\partial x_I}\right) \tag{2}$$

The flow stress can be calculated as

$$\sigma_e = \frac{1}{\alpha} \sinh^{-1}\left[\left(\frac{z}{A}\right)^{1/n}\right] \tag{3}$$

$$A = 1.80 \times 106 + 1.74 \times 10^8 (\%C) - 6.5 \times 10^7 (\%C)^2 \tag{4}$$

$$\alpha = 1.07 + 1.70 \times 10^{-4} T - 2.8 \times 10^{-7} T^2 \tag{5}$$

$$n = 0.2 + 3.96 \times 10^{-4} T \tag{6}$$

$$z = \varepsilon \exp\left(\frac{Q}{RT}\right) \tag{7}$$

The effective strain may be written as

$$\varepsilon = \left(\frac{2}{3}\, \varepsilon_{ij} \varepsilon_{ij}\right)^{1/2} \tag{8}$$

Where ε_{ij} is given by

$$\varepsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \tag{9}$$

So, the viscosity can be calculated as

$$\mu = \frac{\sigma_e}{3\varepsilon} \tag{10}$$

The thermal energy conservation equation at steady state is given by

Table 2
Processing parameters of TIG welded joint [32].

Working Parameters	Shielding Gas/(L·min^{-1})	Current/A	Electrode Diameter/mm	Voltage/V	Welding Speed/(mm·s^{-1})
Description	25 (Argon)	50	3.2	22	3.6

Table 3
Processing parameters of friction stir processing [33].

Sample No	Current	Welding Speed/(mm·min^{-1})	Tool rotational speed/rpm	Frequency	Tilt angle/(°)
1	6.9	63	800	28.52	0
2			1000	33.36	
3			1250	40.00	
4			1600	52.61	

Please cite this article as: Mehdi H, Mishra RS, Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, Defence Technology, https://doi.org/10.1016/j.dt.2020.04.014

$$\rho C_P \frac{\partial (u_i T)}{\partial x_i} = -\rho C_P U_1 \frac{\partial T}{\partial x_i} + \frac{\partial}{\partial x_i}\left(k \frac{\partial T}{\partial x_i}\right) + S_i + S_b \tag{11}$$

Then S_i may be written as

$$S_i = \left[(1-\delta)\eta\tau + \delta\mu_f P_N\right]\ (\omega_r - U_1 \sin\theta)\ \frac{A_r}{V} \tag{12}$$

During friction stir welding, mixing is not atomic, only grains structure is deformed not mixing of atoms.

$$\phi = 2\left[\left(\frac{\partial u_1}{\partial x_1}\right)^2 + \left(\frac{\partial u_2}{\partial x_2}\right)^2 + \left(\frac{\partial u_3}{\partial x_3}\right)^2\right] + \left(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1}\right)^2 + \left(\frac{\partial u_1}{\partial x_3} + \frac{\partial u_3}{\partial x_1}\right)^2 + \left(\frac{\partial u_3}{\partial x_2} + \frac{\partial u_2}{\partial x_3}\right)^2 \tag{13}$$

3.1. Boundary conditions

When the work-piece top surface away from the tool shoulder edge, the boundary condition for heat exchange involves both convection and radiation heat transfer. However, at the sides and the bottom surface of the work-piece, the boundary condition for heat exchange involves only convection heat transfer. The heat generation rate at the interface between tool and the work-piece can be given be given by

$$f = \frac{J_w}{J_T} = \frac{\sqrt{(k_\rho C_P)_W}}{\sqrt{(k_\rho C_P)_T}} \tag{14}$$

Where T and W denote the tool and work piece respectively. The heat flux continuity on the shoulder interface yields

$$k \left.\frac{\partial T}{\partial z}\right|_{top} = \frac{J_w}{J_w + J_T}\ q1 \quad R_P \le r \le R_s \tag{15}$$

Where J_w and J_T are the heat conducted to the work-piece and tool respectively

$$J = \sqrt{k\rho\ C_p} \tag{16}$$

Where C_p is heat capacity, ρ and K are density and thermal conductivity respectively.

The total heat generation rate may be written as

$$q_1 = [\eta(1-\delta)\tau + \delta\mu f.P_H]\ (\omega r - U_1 \sin\theta) \tag{17}$$

The heat transfer coefficient may be calculated as.

$$k \left.\frac{\partial T}{\partial z}\right|_{bottom} = h_b (T - T_a) \tag{18}$$

Value of h_b can be calculated as

$$h_b = h_{b0}(T - T_a)^{0.25} \tag{19}$$

The heat transfer due to radiation and convection is written as

$$k \left.\frac{\partial T}{\partial z}\right|_{top} = h_t (T - T_a) + \sigma\varepsilon^1 \left(T^4 - T_a^4\right) \tag{20}$$

In this model, the computational region is considered as a single-phase visco-plastic non-Newtonian fluid and the FSW tool is considered as rotating in a fixed position as shown in Fig. 3. The material flows into the computational domain from the inlet velocity and out from outlet velocity at the welding speed (traverse speed). Top, bottom and side surfaces of the work-pieces are considered equivalent to the wall surface, having the same velocity as welding speed but opposite in direction.

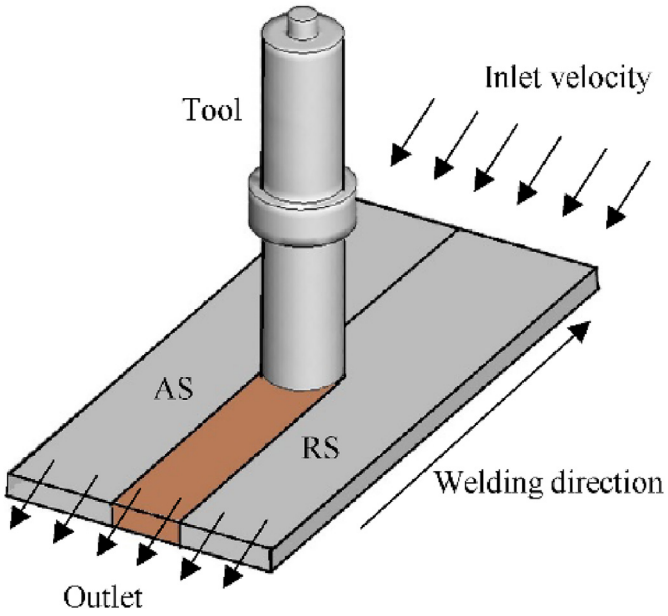


Fig. 3. Boundary condition.

Velocity at the tool pin periphery have been defined in terms of tool translation velocity and the tool pin angular velocity

$$\left.\begin{aligned} u &= (1-\delta)(\omega R_P \sin\theta - U_1) \\ v &= (1-\delta)\omega R_P \cos\theta \\ w &= k\ \frac{\omega}{2\pi} R_P \end{aligned}\right\},\ R_P \le r \le R_S \tag{21}$$

Similarly, at the shoulder contact, the velocity boundary condition may be written as

$$\left.\begin{aligned} u &= (1-\delta)(\omega r \sin\theta - U_1) \\ v &= (1-\delta)\omega r \cos\theta \end{aligned}\right\},\ R_P \le r \le R_S \tag{22}$$

3.2. Material properties

The density of AA7075 and AA6061 is taken as constant, equals to 2810 kg/m^3 and 2700 kg/m^3 respectively, while the specific heat and thermal conductivity are considered as temperature dependent properties [34] as given below

$$k = 25.2 + 3.98 \times 10^{-1} T + 7.36 \times 10 - 6T^2 - 2.25 \times 10^{-7} T^3$$
$$C_P = 929 - 6.27 \times 10^{-1} T + 1.48 \times 10 - 3T^2 - 4.33 \times 10^{-8} T^3$$

3.3. Residual stress measurement by cosα method

A mini portable X-ray diffraction apparatus (pulstec μ-X360) used for analyzing the residual stresses in welded joint of AA6061 and AA7075 samples. The X-ray incident angle was set 35° and ±5° oscillation was applied. The X-ray incident time was 4–5 min throughout this process for each sample. Under these conditions, a diffracted beam from the workpiece surface has captured the

Please cite this article as: Mehdi H, Mishra RS, Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, Defence Technology, https://doi.org/10.1016/j.dt.2020.04.014

images of the welded plates at 50 μm resolution, the size of the beam spot is approximately 2.5 mm for 1.2 mm pinhole collimator.

For residual stress determination, the cosα method was described [35]. The translation from the diffractometer space to the sample inherently more complex due to the 2D planar geometry of the measurement and can be expressed as

$$q_s = \begin{bmatrix} \cos\eta \sin\psi_0 + \sin\eta \cos\psi_0 \cos\alpha \\ \cos\eta \sin\psi_0 \sin\varphi_0 + \sin\eta \cos\psi_0 \sin\varphi_0 cos\alpha + \sin\eta \cos\varphi_0 \sin\alpha \\ \cos\eta \cos\psi_0 + \sin\eta \sin\psi_0 cos\alpha \end{bmatrix} \tag{23}$$

The strain projection along (η, α) coordinates can be written as in terms of scattering vector and strain component as

$$\begin{aligned} \varepsilon_\alpha &= q_i \cdot q_j \cdot \varepsilon_{ij} \\ \varepsilon_{ij} &= \frac{1+\nu}{E}\sigma_{ij} - \delta_{ij}\frac{\nu}{E}\sigma_{kk} \end{aligned} \tag{24}$$

So, the strain projection may be written as

$$\varepsilon_\alpha = \frac{1+\nu}{E} q_i q_j \sigma_{ij} - \frac{\nu}{E}\sigma_{kk} \tag{25}$$

Now, defining two parameters a_1 and a_2 for linear determination of σ_{11} and σ_{22}

$$a_1 = \frac{1}{2}[(\varepsilon_\alpha - \varepsilon_{\pi+\alpha}) + (\varepsilon_{-\alpha} - \varepsilon_{\pi-\alpha})] \tag{26}$$

$$a_2 = \frac{1}{2}[(\varepsilon_\alpha - \varepsilon_{\pi+\alpha}) - (\varepsilon_{-\alpha} - \varepsilon_{\pi-\alpha})] \tag{27}$$

After re-expressing of Eq. (26) and Eq. (27) to lead the final relationship for this method.

$$\begin{aligned} a_1 = \frac{1+\nu}{E}\sin\psi_0 \sin 2\eta \cos\alpha [\sigma_{11}(1+\cos 2\varphi_0) \\ + \sigma_{22}(1-\cos 2\varphi_0) + 2\sigma_{12}(\sin 2\varphi_0)] \end{aligned} \tag{28}$$

$$\begin{aligned} a_2 = \frac{1+\nu}{E}\sin\psi_0 \sin 2\eta \sin\alpha [\sigma_{22} \sin 2\varphi_0 - \sigma_{11} \sin 2\varphi_0 \\ + 2\sigma_{12} \cos 2\varphi_0] \end{aligned} \tag{29}$$

at $\varphi_0 = 0$, the above equations will be

$$a_1 = \sigma_{11}\frac{1+\nu}{E}\sin 2\psi_0 \sin 2\eta \cos\alpha \tag{30}$$

$$a_2 = 2\sigma_{12}\frac{1+\nu}{E}\sin\psi_0 \sin 2\eta \sin\alpha \tag{31}$$

Thus the term cosα in Eq. (30) is the origin of the name for this method.

The value of stresses after re-expression maybe written as

$$\sigma_{11} = -\frac{E}{(1+\nu)} \frac{1}{\sin 2\psi_0 \sin 2\eta} \tag{32}$$

$$\sigma_{12} = -\frac{E}{2(1+\nu)} \frac{1}{\sin 2\psi_0 \sin^2\eta} \tag{33}$$

4. Results and discussion

4.1. Tensile strength

In order to analyze the effect of friction stir processing on the TIG-welded joint, the tensile load applied on TIG and TIG+FSP welded joint by universal testing machine (UTM) at room temperature and fractured surfaces obtained from the tensile test were compared with one another. Three test specimens were tested at each condition and the average of these three results are presented in Tables 4–10. The tensile strength of the weldment made by TIG and TIG+FSP techniques are compared. Due to the absence of porosity, small grain size and presence of extra material in the welded region with filler wire, the tensile strength of TIG+FSP obtained higher value than the conventional TIG joining process. Filler ER5356 present in the welded joint makes more compact pressure leading to increase bond strength instead of filler ER4043. The grain size in the fusion zone for ER5356 is smaller than the ER4043. The tensile strength with filler ER5356 was obtained higher than ER4043 [36]. According to Hall Petch equation $\sigma_1 = \sigma_i + kd^{(-1/2)}$, the tensile strength is inversely proportional to the grain size [37]. The tensile strength of the TIG-welded joint using filler 4043 and 5356 was calculated 158.6 MPa and 172.2 MPa respectively which is less than the TIG+FSP as shown in Table 4. The tensile strength of the TIG+FSP welded joint increased with increasing tool rotational speed. The highest tensile strength (233.17 MPa) was observed in TIG+FSP with filler ER5356 at tool rotational 1600 rpm with feed rate 63 mm/min. Three statistical parameters are investigated i.e., standard deviation (SD), standard error (SE) and 95% confidence interval for TIG+FSP welded joint with filler ER4043 and ER5356 as shown in Tables 5–10. The standard deviation (SD) provides the deviance of the experimental

Table 4
Mechanical properties of TIG welded joint.

Processing Parameters	Tensile Strength/MPa	Hardness/HV	Residual stress/MPa
TIG with filler ER4043	158.6	66	71
TIG with filler ER 5356	172.2	74	67

Table 5
Tensile strength of TIG+FSP welded joint with filler ER4043.

Processing Parameters	Tensile Strength/MPa			Mean Tensile Strength/MPa	Standard Deviation	Standard error	95% confidence Interval	
	Specimen-1	Specimen-2	Specimen-3				Minimum	Maximum
TIG+FSP_800	178.4	178.0	178.8	178.40	0.40	0.23	177.95	178.85
TIG+FSP_1000	189.0	189.4	190.4	189.60	0.72	0.42	188.78	190.42
TIG+FSP_1250	204.5	204.8	205.0	204.77	0.25	0.15	204.48	205.05
TIG+FSP_1600	219.0	219.6	220.0	219.53	0.50	0.29	218.96	220.10

Please cite this article as: Mehdi H, Mishra RS, Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, Defence Technology, https://doi.org/10.1016/j.dt.2020.04.014

Table 6
Hardness of TIG+FSP welded joint with filler ER4043.

Processing Parameters	Hardness/HV			Mean Hardness /HV	Standard Deviation	Standard error	95% confidence Interval	
	Specimen-1	Specimen-2	Specimen-3				Minimum	Maximum
TIG+FSP_800	82.5	82.0	83.5	82.6	0.76	0.44	81.80	83.53
TIG+FSP_1000	89.0	90.0	89.0	89.3	0.58	0.33	88.68	89.99
TIG+FSP_1250	92.6	92.0	92.0	92.2	0.35	0.20	91.81	92.59
TIG+FSP_1600	96.0	97.0	96.5	96.5	0.50	0.29	95.93	97.07

Table 7
Residual Stress of TIG+FSP welded joint with filler ER4043.

Processing Parameters	Residual Stress/MPa			Mean Residual Stress/MPa	Standard Deviation	Standard error	95% confidence Interval	
	Specimen-1	Specimen-2	Specimen-3				Minimum	Maximum
TIG+FSP_800	63	63	64	58.67	0.58	0.33	58.01	59.32
TIG+FSP_1000	59	58	57	56.33	1.00	0.58	55.20	57.46
TIG+FSP_1250	49	50	49	51.00	0.58	0.33	50.35	51.65
TIG+FSP_1600	43	43	45	44.67	1.15	0.67	43.36	45.97

Table 8
Tensile strength of TIG+FSP welded joint with filler ER5356.

Processing Parameters	Tensile Strength/MPa			Mean Tensile Strength/MPa	Standard Deviation	Standard error	95% confidence Interval	
	Specimen-1	Specimen-2	Specimen-3				Minimum	Maximum
TIG+FSP_800	187.5	187.0	187.6	187.37	0.32	0.19	187.00	187.73
TIG+FSP_1000	194.2	194.8	195.4	194.80	0.60	0.35	194.12	195.48
TIG+FSP_1250	218.4	219.4	218.8	218.87	0.50	0.29	218.30	219.44
TIG+FSP_1600	233.0	233.5	233.0	233.17	0.29	0.17	232.84	233.49

Table 9
Hardness of TIG+FSP welded joint with filler ER5356.

Processing Parameters	Hardness/HV			Mean Hardness /HV	Standard Deviation	Standard error	95% confidence Interval	
	Specimen-1	Specimen-2	Specimen-3				Minimum	Maximum
TIG+FSP_800	89.5	89.0	89.0	89.2	0.29	0.17	88.84	89.49
TIG+FSP_1000	93.0	94.5	94.0	93.8	0.76	0.44	92.97	94.70
TIG+FSP_1250	99.5	99.0	100.0	99.5	0.50	0.29	98.93	100.07
TIG+FSP_1600	102.0	102.0	102.5	102.2	0.29	0.17	101.84	102.49

Table 10
Residual Stress of TIG+FSP welded joint with filler ER5356.

Processing Parameters	Residual Stress/MPa			Mean Residual Stress/MPa	Standard Deviation	Standard error	95% confidence Interval	
	Specimen-1	Specimen-2	Specimen-3				Minimum	Maximum
TIG+FSP_800	58	58	60	58.7	1.15	0.67	57.36	59.97
TIG+FSP_1000	52	53	52	52.3	0.58	0.33	51.68	52.99
TIG+FSP_1250	39	38	40	39.0	1.00	0.58	37.87	40.13
TIG+FSP_1600	32	31	32	31.7	0.58	0.33	31.01	32.32

values from the mean it may be calculated as $SD = [\Sigma(X_i - M)^2/(N-1)]^{1/2}$ and the standard error is a quantity used to measure how to close the prediction values to the experimental values. *SE* is calculated as $SE = SD/N^{1/2}$, where N= No of observation and M = mean.

The confidence interval has shown that tensile strength and hardness increased with increasing tool rotation, whereas residual stress decreased with increasing tool rotation.

4.2. Micro-hardness

The micro-hardness distribution of the TIG-welded joint of AA6061 and AA7075 with different filler wire and TIG+FSP weldments with different processing parameters were analyzed by Vickers hardness testing results are shown in Fig. 4. Vicker microhardness machine was used for measuring the hardness across the welded joints with a load of 100 g and dwell time 30 s. The hardness distributions are asymmetrical in the weld center due to the microstructure of the advancing side and retreating side introduced by unsteady plastic flow from the base metal to the weld center [38–40]. The hardness slopes downward from the base metal to the welded region. The hardness fluctuates largely on the advancing side while the hardness holds steady in the retreating side [41].

The alloying elements such as Si and Mg existing in the weld center make precipitation reaction and form a strong precipitate of Mg_2Si to give a higher strength as shown in Fig. 5(a). Fig. 4 shows the micro-hardness profile at the heat-affected zone of aluminum alloy 6061 and 7075. It showed a significant difference, where the

Please cite this article as: Mehdi H, Mishra RS, Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, Defence Technology, https://doi.org/10.1016/j.dt.2020.04.014

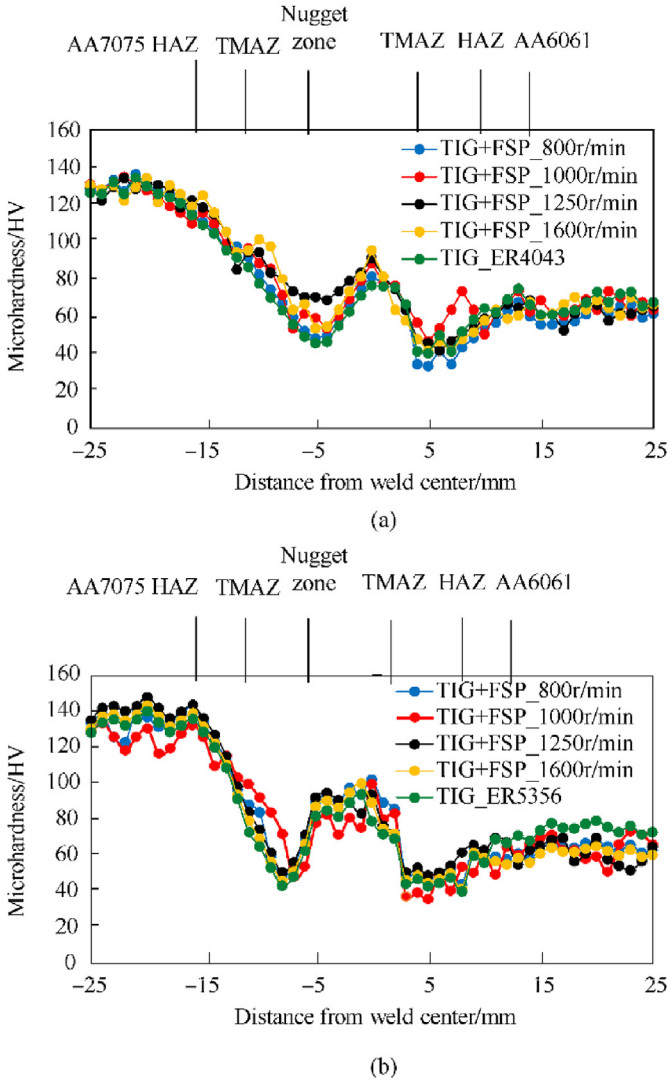


Fig. 4. Comparison of micro-hardness of TIG and TIG+FSP, (a) filler wire ER4043, (b) filler wire ER5356.

welded joint using filler ER4043 showed a lower average hardness value compared to filler ER5356. The hardness value at the center of weldment for filler ER4043 and ER5356 are 96 HV and 102 HV respectively, during welding, the filler wire ER4043 shows the columnar grains while fine equiaxed grains are found in ER5356. It may be analyzed that the fine equiaxed grains improve mechanical properties than the columnar grains [42].

4.3. XRD and EDX analysis

X-ray diffraction (XRD) analysis has been taken on transverse cross-sections of dissimilar TIG and TIG+FSP welded joint of AA6061 and AA7075 with filler ER4043 and ER5356. The corresponding pattern for the welded joints are shown in Fig. 5 and found four major phases i.e. Al, Al_2CuMg, $MgZn_2$, and Mg_2Si. Magnesium (Mg) and Silicon (Si) elements were found in the weldment besides the aluminum (Al), it is found that Mg and Si created the phase after the precipitation reaction in the weldment. The very high intensity was found from aluminum, because of the fragmentation of precipitates the intensity of $MgZn_2$ was decreases after friction stir processing on TIG welded joint with filler ER5356. The alloying elements such as Si and Mg existing in the weld center make precipitation reaction and form a strong precipitate of Mg_2Si to give a higher strength. The same phase of Al_2CuMg was detected in both the cases (filler ER4043 and ER 5356).

The hardness of the FSP joints is based on boundary energy, brittle intermetallic formation, strain hardening and precipitates formation in the joint. Fine recrystallized grains and increase of grain boundaries in the (SZ) of AA6061 and AA7075 joints predict higher hardness. Because of fine precipitates and fine grains structure, the stir zone is associated with plastic deformation and high temperature, due to precipitates formation at high temperature along the grain boundaries the hardness value of TIG+FSP at 1600 rpm with filler ER5356 recorded higher values than the TIG welded joints. The energy-dispersive X-ray spectroscopy (EDX/EDS) of the TIG-welded joint with filler ER4043 and ER 5356 welded joint has been analyzed. It was found that the atomic percentage of Si and Mg in the welded joint is higher than the other elements. Fig. 5(b) and (c) illustrates the EDX image and percentage of element concentration at fusion zone (FZ) in the TIG welded joint with filler ER4043 and ER5356. Zinc (Zn), Magnesium (Mg) and Silicon (Si) elements were found in the weldment besides the aluminum (Al), it is found that Mg, Zn, and Si created the phases after the precipitation reaction in the weldment.

4.4. Residual stresses measurement

A mini portable X-ray diffraction apparatus (Pulstec μ-X360) at Delhi technological university, Delhi, India was used to determine the residual stresses in weldment of AA6061 and AA7075 by the cosα method. The cosα method was introduced in japan for residual stress analysis in 1978 [43]. This method is allowed the stress analysis by capturing the results by a single incident X-ray beam via a 2D detector. It shows the peak center is very stable throughout the welding. Debye ring and distortion ring is obtained by using cosα method. An investigation has been done to analyze the residual stress. The 3D Debye ring and distortion ring of welded samples are shown in Fig. 7. The X-ray incident angle was set 35° and ±5° oscillation was applied. The variation of residual stresses in the transverse direction of weldment with filler ER4043 and ER5356 are shown in Fig. 6(a) and (b). Residual stresses (compressive or tensile) will influence the mechanical behavior of the welded joint. It can reduce brittle fracture strength, buckling strength and cracking in the weldment. Residual stress is also influenced by the prediction of brittle failure and affect the lifetime prediction of the component [44–47]. Residual stress contributes both negative and positive effects to the weldment, generally, the tensile residual stress leads to a negative effect on the weldment [48]. When the fixtures are unconstrained and the welded plate temperature is reduced to room temperature, the material in the nugget zone (NZ) tends to recover, but the material in the heat-affected zone (HAZ) has smaller deformation and will prevent the recovery process in NZ, so the maximum residual stress occurs in the boundaries of the HAZ with a minimum in the NZ. The deformation on the advancing side and retreating side differ, causing the recovery processes to differ too. So the residual stress cannot be symmetry to the welding line [49]. The base material AA7075 on LHS of the weldment shows a minimum compressive residual stress, however, the residual stress gradually increases from the base material to towards the weldment and then decreases till second base material AA6061. The maximum residual stress occurs in the location where the equivalent of the plastic strain is decreased suddenly. The maximum compressive residual stress 71 MPa were located at the fusion zone (FZ) of the TIG weldment with filler ER4043, whereas minimum compressive residual stress 37 MPa was obtained at stir zone (SZ) of the TIG+FSP with filler 5356, the residual stresses profile of TIG+FSP were not symmetrical about the centerline of the

Please cite this article as: Mehdi H, Mishra RS, Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, Defence Technology, https://doi.org/10.1016/j.dt.2020.04.014

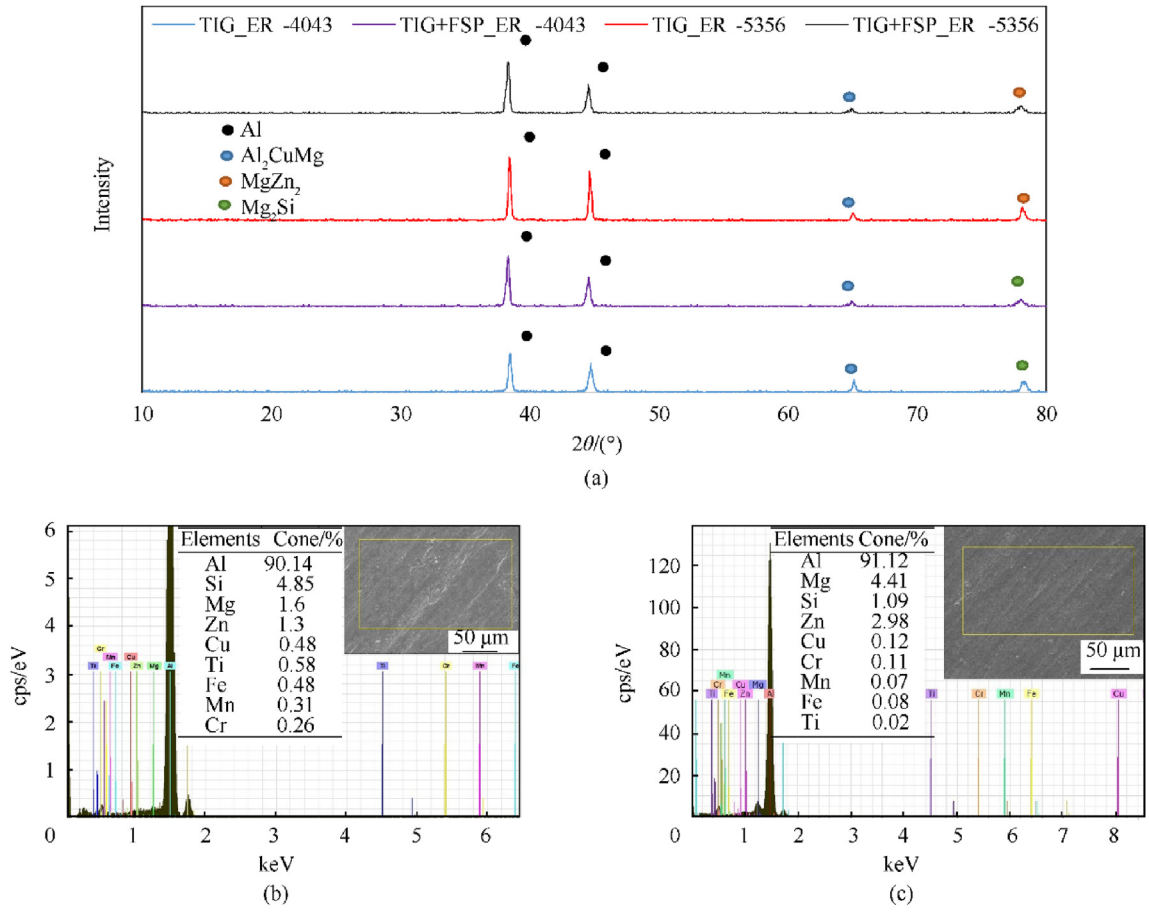


Fig. 5. (a) XRD peaks of TIG and TIG+FSP welded joint, (b&c) EDX analysis of TIG welded joint in fusion zone with filler ER4043 and filler ER 5356.

weldment for both the cases. The left side peak value of the weldment was greater than the right side because the forming processes are different for the retreating side (RS) and advancing side (AS).

4.5. Model validation

To confidently use CFD results for investigating the TIG+FSP process, the CFD model has to be correctly defined and a thorough validation has to be achieved. The model was first applied to simulate the experimental work on friction stir welded joint of Al-alloy 6061 carried out by Hwang el al [50]. The experimental temperatures were measured by thermocouple units placed equally at a distance of 5 mm, along the traverse direction of the rotating tool. The tool rotational speed and traverse speed, was 920 rpm and 20 mm/min respectively. The present simulation result was validated by this experimental results which gives the satisfactory amount of assurance in the fidelity of the simulation of welded joints as shown in Table 11.

4.6. Temperature variation in TIG+FSP process

The experimental temperature results of TIG+FSP measured by thermocouples. The advancing side temperatures in the transverse direction are A_1, A_2, A_3, and A_4 and the retreating side has R_1, R_2, R_3, and R_4. The peak temperature of advancing side is marginally higher than the retreating side [51,52]. The temperature of FSP tool is symmetric about the tool axis. The high temperature in the vicinity of the welding tool is attributed to the localized heat generation. During the TIG+FSP process, the temperature of the base plate around the welding tool is around the 765 K. It is still lower than the melting point of AA6061 and AA7075. When the position weld bead is far away from the FSP tool, the temperature drops quickly as shown in Fig. 9(a) and (b). Fig. 8 shows the temperature distribution plot for TIG+FSP welding at 160 s. During this simulation, eight points were observed to obtain the temperature-time curves which can be compared with the experimental results. The area around the FSP tool reached the maximum temperature. The maximum temperature was observed 760 K at the advancing side whereas minimum temperature was observed 307 K at the retreating side. Higher heat is generated in the SZ at high tool rotation, the prevailing thermal conditions are controlled by the distribution and availability of precipitates in the matrix. Fig. 10 shows the temperature variation profile on the mid-section of the top surface of (TIG+FSP) welded joint with different processing parameters. The region where its peak temperature is higher almost 450 °C in (TIG+FSP) of AA6061 and AA7075 aluminum alloy at tool rotational speed of 1600 rpm. The heat is transmitted to the aluminum alloy which is preheated during tool rotation. The initial heating was predicted as the monitoring thermal contour of the rotating tool. When the rotating tool moves on the monitoring location, the temperature contour leads to slow cooling. Due to the relation between tool pin rotation and material flow, the higher shearing rate was observed at (A.S). Thus the temperature of (RS) is slightly lower than the (AS), while almost the symmetric temperature was observed at the bottom of the welded joint.

The estimated maximum temperature about 515 °C was calculated in the stir zone of the TIG+FSP welded joint at tool rotation of

Please cite this article as: Mehdi H, Mishra RS, Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, Defence Technology, https://doi.org/10.1016/j.dt.2020.04.014

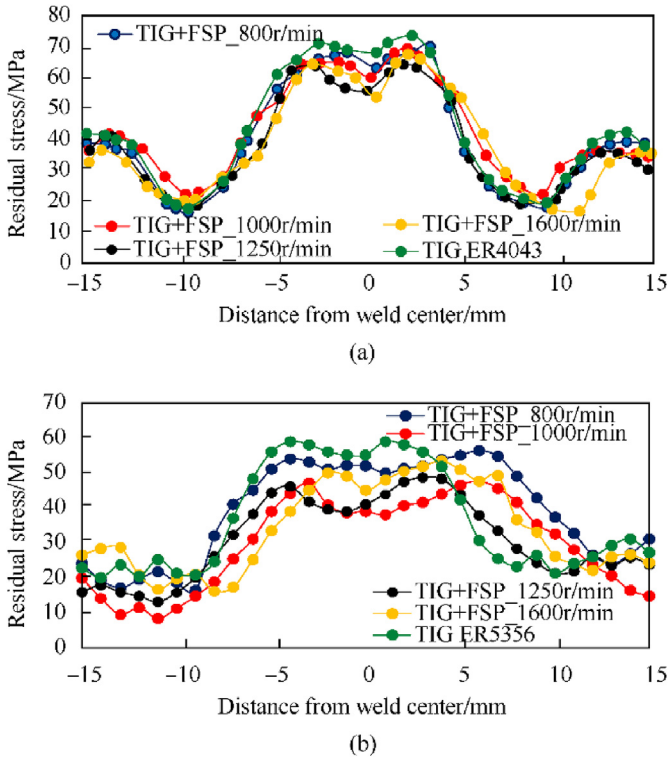


Fig. 6. Variation of transverse compressive residual stress at the welded joint, (a) filler ER4043, (b) filler ER5356.

1600 rpm whereas 408 °C was calculated in the stir zone at 800 rpm. The temperature and heat flux distribution with different processing parameters are shown in Figs. 10 and 11. In general, the temperature values at the (A.S) about 10–25 °C greater than the

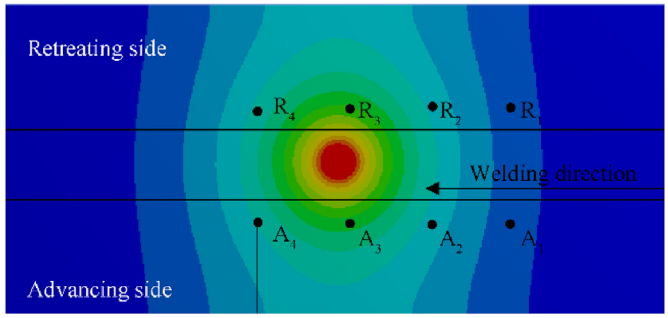


Fig. 8. Temperature contour of TIG+FSP at 160 s.

(R.S), When the tool approaches the target location then there is a rapid change in the temperature, while the slower cooling rate was observed when the tool moves away from the target location [53]. The variation of heat flux at the tool workpiece interface is shown in Fig. 11. It was observed that the heat flux is directly proportional to the tool rotational speed. The maximum heat flux about 5.33×10^6 w/m^{2}was obtained in the SZ at 1600 rpm, whereas minimum heat flux was obtained at 1000 rpm. The non-uniformity was observed in the heat flux pattern at different processing parameters of the tool, because the rapid recirculation of plasticized material, the heat flux rate does not lead to the variation of local temperature.

4.7. Microstructure analysis of TIG and TIG+FSP welded joint

For metallography observation, samples were first ground and mechanically polished and then etched in keller reagent (HNO_3, HF and HCL) and observed by the microscopic machine. The microstructure of TIG welded joint of Al-6061 and 7075 with filler ER4043 and ER 5356 were observed at fusion zone as shown in

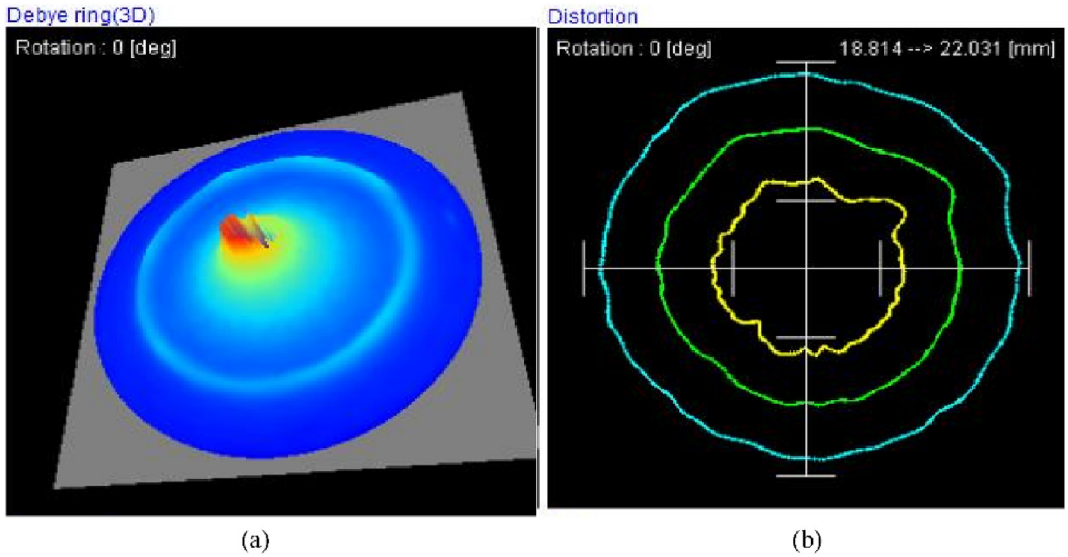


Fig. 7. (a) 3D Debye ring at the center of the TIG+FSP welded joint, (b) 2D Distortion ring at the center of the TIG+FSP welded joint.

Table 11
Validation of temperature distribution of friction stir welded joint of AA6061 [50].

Author	Rotation speed/rpm	Traverse speed/(mm·min^{-1})	Maximum temperature/°C
Present result	920	20	398
Hwang et al. [50]	920	20	386

Please cite this article as: Mehdi H, Mishra RS, Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, Defence Technology, https://doi.org/10.1016/j.dt.2020.04.014

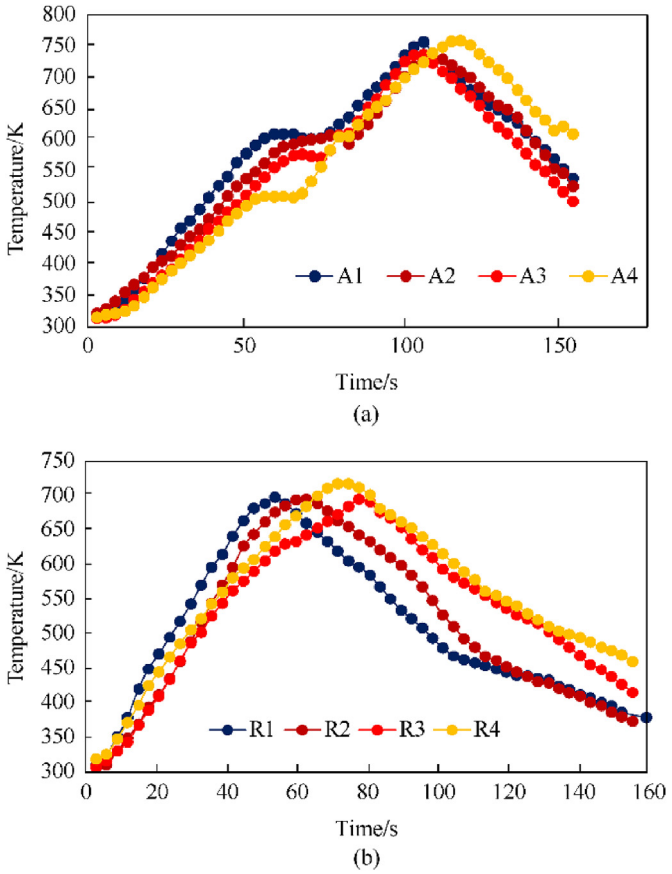


Fig. 9. Variation of temperature during TIG+FSP welding: (a) Advancing side, (b) Retreating side.

Fig. 12. The dispersed precipitates of Al_2CuMg intermetallic compound were observed in welded region and coarse equiaxed grains were also observed in fusion zone with filler ER4043. The equiaxed grains is absent due to the material close to the fusion line provides plenty of sources for the crystal nucleation of the liquid metal during the TIG welding [54]. Fine grains with small precipitates was reported in TIG welded joint with filler ER5356 as compare to ER4043.

Three different zones have been recognized in TIG+FSP weldment at low magnification due to mechanical and thermal stresses caused by the processing parameters. These zones are nugget zone (NZ), thermo-mechanically affected zone (TMAZ), and heat affected zone (HAZ). The formation of nugget zone shape in TIG+FSP welded joint is recognized to the maximum deformation and plasticization in the material which shows the fine recrystallized equiaxed grains. The formation of nugget shapes depends on thermal gradient, processing parameters and tools geometry in the work-piece [55]. Therefore, coarse grain structure of TIG welded joint is transformed into the uniform and fine grains structures in the weld nugget zone due to adequate softening of material revealed the maximum tensile strength and microhardness of the TIG+FSP welded joint as shown in Fig. 13. At high welding speed, the weld nugget zone is more homogenous than those produce low welding speed because high heat input gives the effective recrystallization and more homogenous temperature distribution in weld nugget zone. The grains size in stir zone change crucially which is depend on the heat input and processing parameters [56,57].

The grain size of TIG weldment were analyzed by the image J software and observed grain size in fusion zone with filler ER4043 and ER5356 are 20.4 μm, and 18.2 μm. Fig. 14 shows the effect of tool rotational speed on grain size of TIG+FSP welded joint at nugget zone. It can be conclude that the rotational speed has a significant effect of grain size of the welded joint. When the rotational speed increases, grain size decreases, this observation give the satisfactory amount of assurance with Yupeng Li [58]. The average minimum grain size i.e. 3.4 μm was observed in TIG+FSP with filler ER5356 at tool rotational speed 1600 rpm.

Fig. 15 shows the fractured surface after the tensile test. In this analysis two mode of failure were observed. The fractured occurs in retreating side at low rotational speed (800 rpm), on the other hand failure was observed in advancing side at high tool rotational speed (1600 rpm). This observation gives the satisfactory amount of assurance with Mishra [5]. SEM fractograph has been taken from the fractured tensile test. fracture morphology between the TIG and TIG+FSP shows that TIG welded portion shows the large voids whereas TIG+FSP welded portion shows fine dimples, this is the evidence of crack nucleation and growth 4 mm away from the weld line. The ductile fracture of welded joint occurs, an improvement in

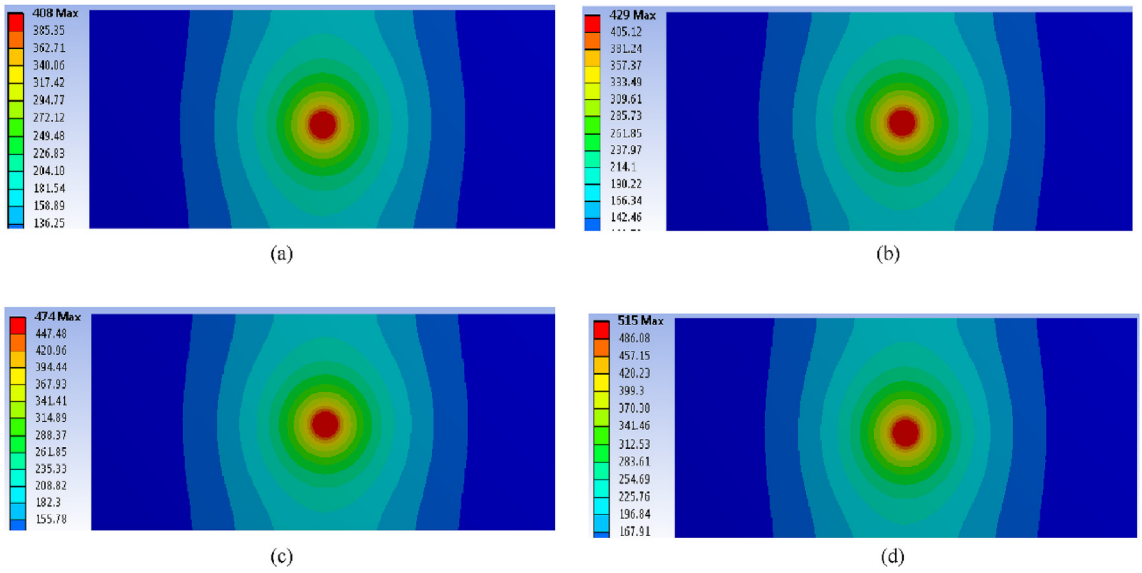


Fig. 10. Temperature distribution at the center of the TIG+FSP welded joint, (a) 800 rpm, (b) 1000 rpm, (c) 1250, (d) 1600 rpm.

Please cite this article as: Mehdi H, Mishra RS, Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, Defence Technology, https://doi.org/10.1016/j.dt.2020.04.014

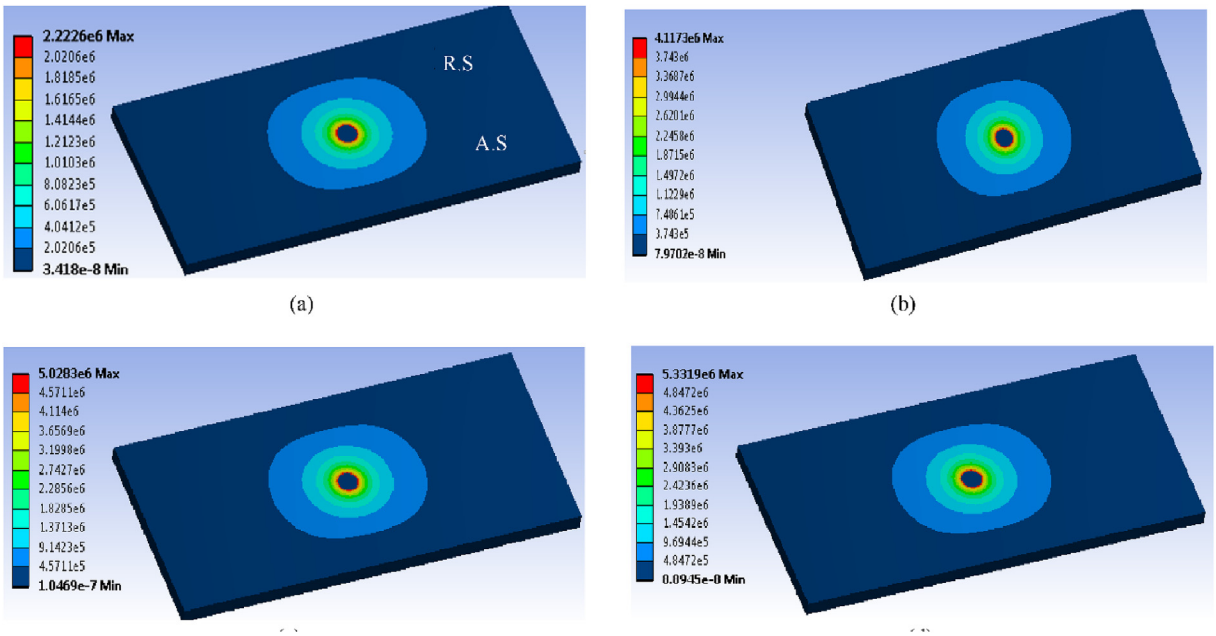


Fig. 11. The heat flux at the center of TIG+FSP welded joint, (a) 800 rpm, (b) 1000 rpm, (c) 1250 rpm, (d) 1600 rpm.

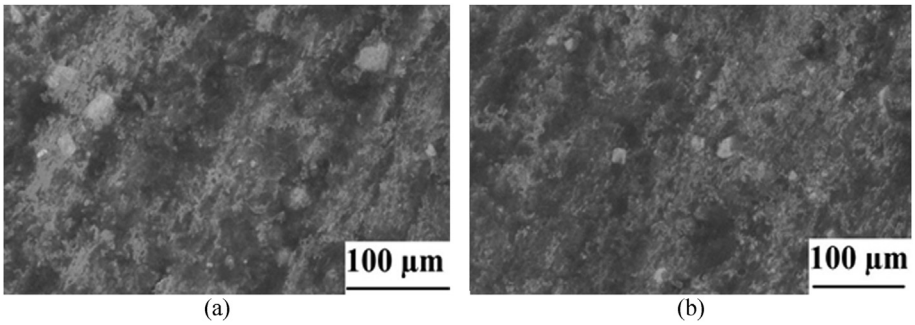


Fig. 12. SEM images of TIG welded joint at fusion zone (a) filler ER 4043, (b) filler ER 5356.

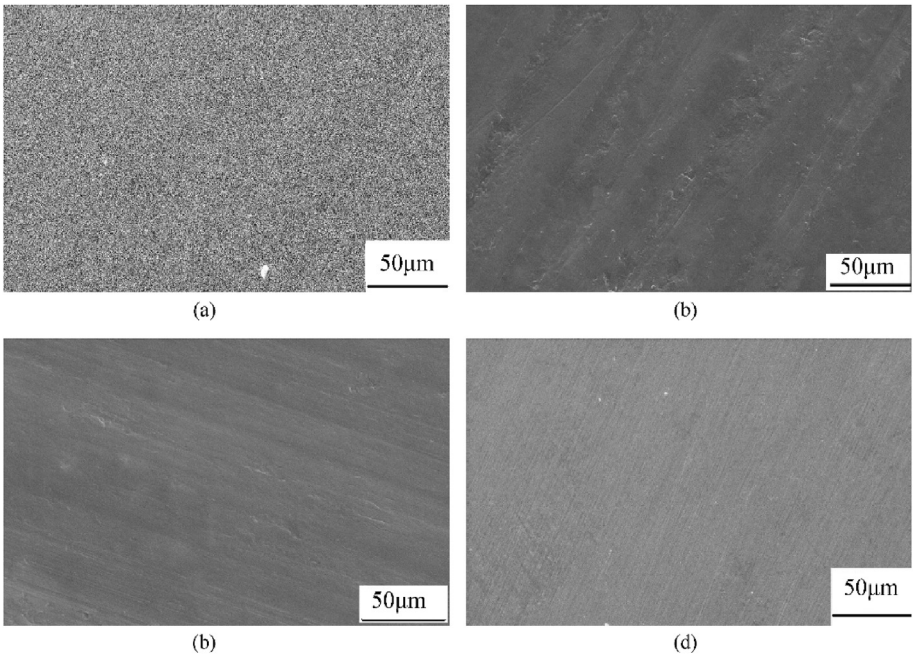


Fig. 13. SEM images of TIG+FSP welded joint at tool rotation (a) 800 rpm, (b) 1000 rpm, (c) 1250 rpm, (d) 1600 rpm.

Please cite this article as: Mehdi H, Mishra RS, Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, Defence Technology, https://doi.org/10.1016/j.dt.2020.04.014

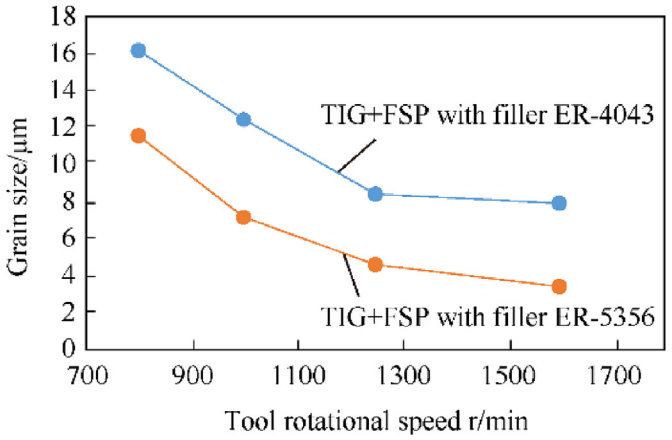


Fig. 14. Effect of tool rotational speed on grain size of TIG+FSP welded joint.

ductility may be achieved when the cavity nucleation could be suppressed [59]. The maximum inter-facial normal stress is depends upon the grain particle size and the volume fraction of the grain particles [60]. The fractured surface at room temperature were shown in Fig. 15.

The large dimples and quasi cleavage with sharp edge and various depth were found on the fractured tensile specimen surface of TIG welded joint with filler ER4043 and ER5356as shown in Fig. 15(a) and (b). Many large and equiaxed dimples were observed in TIG welded joint with filler ER4043 whereas the main fractured mode was quasi cleavage with lot of tiny dimples gathered around the large and quasi cleavage dimples found in filler ER5356 which shows the ductile fracture. The FSW processing parameters effect the temperature distribution and material flow across the TIG welded joint. Fig. 15(c) shows the SEM micrograph of factored surface of TIG+FSP welded joint, finer dimples are observed in TIG+FSP weldment as compare to TIG welded joint, which shows the better mechanical properties.

5. Conclusion

A new welding approach has been investigated to see the effect of FSP on TIG welding for dissimilar alloy AA6061 and AA7075 followed by the different filler wire ER4043 and ER5356. The mechanical characterization, finite element formulation and mathematical equations of heat transfer of TIG+FSP welded joints are investigated using ANSYS Fluent software by adjusting process parameters of FSP. The confidence interval has shown that tensile strength and hardness increased with increasing tool rotation, whereas residual stress decreased with increasing tool rotation. The maximum compressive residual stress 71 MPa were located at the fusion zone (FZ) of the TIG weldment with filler ER4043, whereas minimum compressive residual stress 37 MPa was obtained at stir zone (SZ) of the TIG+FSP with filler 5356. The large dimples and quasi cleavage with sharp edge and various depth were found on the fractured tensile specimen surface of TIG welded joint with filler ER4043 and ER5356 whereas fine dimples are observed in TIG+FSP weldment, which shows the better mechanical properties. TIG+FSP resulted in generated fine grains of 3.2−4 μm. The grain size decreases by increasing the tool rotational speed. The minimum grain size was obtained in TIG+FSP welded joint with filler ER5356 at 1600 rpm.

The asymmetry of the temperature distribution during TIG+FSP welding of AA7075 and AA6061 determined by numerical simulation corresponded with the experimental results. The peak temperatures on AS were higher than the RS of 20 K. The heat transfer is

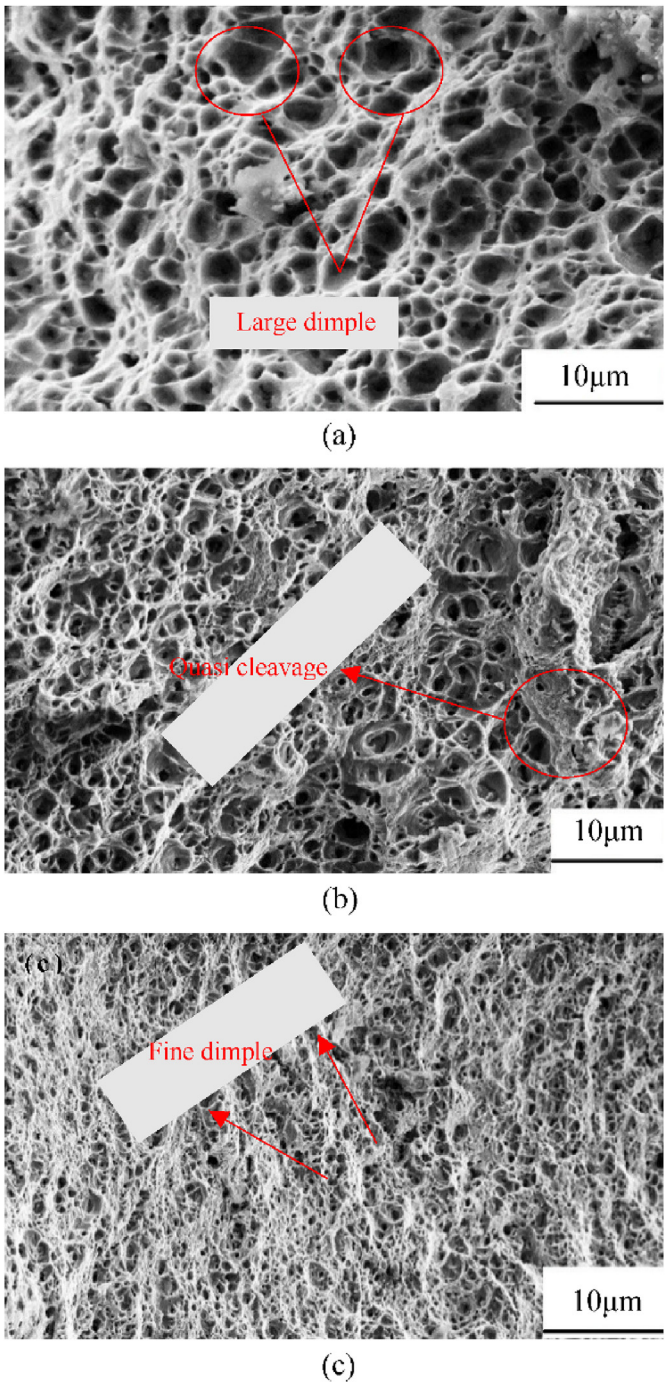


Fig. 15. SEM images of fractured tensile specimen, (a) TIG welded joint with filler ER4043, (b) TIG welded joint with filler ER4043, (c) TIG+FSP welded joint.

analyzed by the effect of different processing parameters of FSP tool. The temperature at advancing side is higher than the retreating side. The predicted peak values of temperature at the weld region was calculated by the ANSYS software and found the maximum temperature about 515 °C at tool rotation of 1600 rpm, whereas the maximum heat flux rate about 5.33×10^6 was observed at 1600 rpm.

Declaration of competing interest

The author declare that they have no conflict interest.

Please cite this article as: Mehdi H, Mishra RS, Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, Defence Technology, https://doi.org/10.1016/j.dt.2020.04.014

References

[1] Dialami N, Chiumenti M, Cervera M, de Saracibar CA. Challenges in thermo-mechanical analysis of friction stir welding processes. Arch Comput Methods Eng 2017b;24:189–225.
[2] Gemme F, Verreman Y, Dubourg L, Jahazi M. Numerical analysis of the dwell phase in friction stir welding and comparison with experimental data. Mater Sci Eng A 2010;527(16):4152–60.
[3] Mosneaga VA, Mizutani T, Kobayashi T, Toda H. Impact toughness of weldments in Al-Mg-Si alloys. Mater Trans 2002;43:1381–9.
[4] Karunakaran N. Effect of pulsed current on temperature distribution and characteristics of GTA welded Magnesium alloy, vol. 4; 2013. p. 1–8.
[5] Mishra RS, Ma ZY. Friction stir welding and processing. Mater Sci Eng R Rep 2005;50(1):1–78.
[6] Zhan M, Guo K, Yang H. Advances and trends in plastic forming technologies for welded tubes. Chin J Aeronaut 2015;29(2):305–15.
[7] Masubuchi K. Analysis of welded structures, residual stresses, distortion, and their consequences. New York: Pergamon Press; 1980.
[8] Chen G-Q, Shi Q-Y, Li Y-J, Sun Y-J, Dai Q-L, Jia J-Y, Zhu Y-C, Wu J-J. Computational fluid dynamics studies on heat generation during friction stir welding of aluminum alloy. Comput Mater Sci 2013;79:540–6.
[9] Su H, Wu CS, Pittner A, Rethmeier M. Thermal energy generation and distribution in friction stir welding of aluminum alloys. Energy 2014;77:720–31.
[10] Nandan R, Roy GG, Lienert TJ, Debroy T. Three-Dimensional heat and material flow during friction stir welding of mild steel. Acta Mater 2007;55(3):883–95.
[11] Ji S, Shi Q, Zhang L, Zou A, Gao S, Zan L. Numerical simulation of material flow behavior of friction stir welding influenced by rotational tool geometry. Comput Mater Sci 2012;63:218–26.
[12] Yu Z, Zhang W, Choo H, Feng Z. Transient heat and material flow modeling of friction stir processing of Magnesium alloy using threaded tool. Metall Mater Trans A 2012;43(2):724–37.
[13] Su H, Wu CS, Bachmann M, Rethmeier M. Numerical modeling for the effect of pin profiles on thermal and material flow characteristics in friction stir welding. Mater Des 2015;77:114–25.
[14] Dalle Donne C, Biallas G, Ghidini T, Raimbeaux G. Effect of weld imperfections and residual stresses on the fatigue crack propagation in friction stir welded joints. In: 2nd international conference on friction stir welding. Gothenburg, Sweden; 2000.
[15] Dalle Donne C, Lima E, Wegener E, Pizalla A, Buslaps T. Investigations on residual stresses in friction stir welds. In: 3rd international conference on friction stir welding, Kobe, Japan; 2001.
[16] Peel M, Steuwer A, Preuss M, Withers PJ. Withers, microstructure, mechanical properties and residual stresses as a function of welding speed in aluminum AA5083 friction stir welds. Acta Mater 2003;51(16):4791–801.
[17] Wang X-L, Feng Z, David SA, Spooner S, Hubbard CR. In: 6th int. Conf. Residual stresses, Oxford, UK, July 10-12; 2000.
[18] Cao YJ, Qi X. Thermal and thermo-mechanical modeling of friction stir welding of aluminium alloy 6001-T6. J Mater Process Manuf Sci 1998;7:215–33.
[19] Chen CM, Kovacevic R. Finite element modeling of friction stir welding-thermal and thermomechanical analysis. Int J Mach Tool Manufact 2003;43:1319–26.
[20] Zhu XK, Chao YJ. Numerical simulation of transient temperature and residual stresses in friction stir welding of 304L stainless steel. J Mater Process Technol 2004;146:263–72.
[21] Li T, Shi Q-Y, Li H-K, Wang W, Cai Z-P. In: Proc. Int. Weld. Join. Conf. Seoul, Korea: COEX Convention Center; May 2007.
[22] Sutton MA, Reynolds AP, Wang DQ, Hubbard CR. J Eng Mater Technol 2002;124(4):215–21.
[23] Mehdi Husain, Mishra RS. Mechanical properties and microstructure studies in friction stir welding (FSW) joints of dissimilar alloy-A review. J Achiev Mater Manuf Eng 2016;77(1):31–40.
[24] Mehdi Husain, Mishra RS. Influences of process parameter and microstructural studies in friction stir welding of diferent alloys: a review. Int J Adv Prod Ind Eng 2017;509:55–62.
[25] Mehdi Husain, Mishra RS. Mechanical and microstructure characterization of friction stir welding for dissimilar alloy—a review. Int J Res Eng Innov 2017;1(5):57–67.
[26] Mehdi Husain, Mishra RS. Analysis of material fow and heat transfer in reverse dual rotation friction stir welding: a review. Int J Steel Struct 2019;19:422–34.
[27] Husain Mehdi H, Mishra RS. Study of the influence of friction stir processing on tungsten inert gas welding of diferent aluminum alloy. SN Appl Sci 2018;1:712. 2019.
[28] Taheri-Behrooz Fathollah, Aliha Mohammad RM, Maroofi Mahmood, Hadizadeh Vahid. Residual stresses measurement in the butt joint welded metals using FSW and TIG methods. Steel Compos Struct 2018;28(No. 6):759–66.
[29] Zhang Changqing, Wang Weijie, Jin Xin, Rong Chen, Qin Zhuo. A study on microstructure and mechanical properties of micro friction stir welded ultra-thin al-1060 sheets by the shoulder less tool. Metals 2019;9(5):507.
[30] Devireddy Krishnaja, Devuri Venkateswarlu, Cheepu Murali Mohan, Kranthi Kumar Balina. Analysis of the influence of friction stir processing on gas tungsten arc welding of 2024 auminum alloy weld zone. Int J Mech Prod Eng Res Dev 2018;8(1):243–52.
[31] Mabuwa S, Msomi V. Effect of friction stir processing on gas tungsten arc-welded and friction stir-welded 5083-H111 aluminium alloy joints. Adv Mater Sci Eng 2019:1–14.
[32] Ishak Mahadzir, Noordin Nur Fakhriah Mohd, Shah Luqman Hakim Ahmad. Feasibility study on joining dissimilar aluminum alloys aa6061 and aa7075 by tungsten inert gas (TIG). J Teknol 2015;75(7):79–84.
[33] Mohammadi-pour M, Khodabandeh A, Mohammadi-pour S, et al. Microstructure and mechanical properties of joints welded by friction-stir welding in aluminum alloy 7075-T6 plates for aerospace application. Rare Met 2016. https://doi.org/10.1007/s12598-016-0692-9.
[34] Arora A, De A, DebRoy T. Toward optimum friction stir welding tool shoulder diameter. Scripta Mater 2011;64(1):9–12.
[35] Sasaki T, Hirose Y, Sasaki K, Yasukawa S. Influence of image processing conditions of debye Scherrer ring images in x-ray stress measurement using an imaging plate. Adv X Ray Anal 1997;40:588–94.
[36] Salleh MNM, Ishak M, Shah LH, Idris SRA. The effect of ER4043 and ER5356 filler metal on welded Al 7075 by metal inert gas welding. WIT Trans Built Environ 2016;166:213–24.
[37] Smith William F, Hashemi Javad. Foundations of materials science and engineering. fourth ed. McGraw-Hill; 2006, ISBN 978-0-07-295358-9.
[38] Xu W, Liu J, Luan G, Dong C. Temperature evolution, microstructure and mechanical properties of friction stir welded thick 2219-O aluminum alloy joints. Mater Des 2009;30:1886–93.
[39] Xu W, Liu J, Luan G, Dong C. Microstructure and mechanical properties of friction stir welded joints in 2219-T6 aluminum alloy. Mater Des 2009;30:3460–7.
[40] Luijendijk T. Welding of dissimilar aluminium alloys. J Mater Process Technol 2000;103:29–35. 2000.
[41] Jia Yonggui Qin, Ou Yiwen, Wang Kehong, Liu Jizi. The influence of microstructural heterogeneity on mechanical properties of friction stir welded joints of T6-treated Al-Zn-Mg alloy 7A52Yang. Metals 2018;8(7):527.
[42] Ishak Mahadzir, Noordin Nur Fakhriah Mohd, Shah Luqman Hakim Ahmad. Feasibility study on joining dissimilar aluminum alloys aa6061 and aa7075 by tungsten inert gas (TIG). J Teknol (Sci Eng) 2015;75(7):79–84.
[43] Taira S, Tanaka K, Yamazaki T. A method of x-ray microbeam measurement of local stress and its application to fatigue crack growth problems. J Soc Mater Sci Jpn 1978;27(294):251–6.
[44] Yeni C, Sayer S, Pakdil M. Comparison of mechanical and microstructural behaviour of TIG, MIG and friction stir welded 7075 aluminium alloy. Kovove Mater 2009;47:341–7.
[45] Rossini NS, Dassisti M, Benyounis KY, Olabi AG. Methods of measuring residual stresses in components. Mater Des 2012;35:572–88.
[46] James MN. Residual stress influences in mechanical engineering. In: XVIII congreso nacional de Ingeniería Mecánica; 2010. p. 1–13.
[47] Yang YP, Babu SS, Kikel JM, Brust FW. Investigation of weld crack mitigation techniques with advanced numerical modeling and experiment-Summary. In: Lippold J, Böllinghaus T, Cross CE, editors. Hot cracking phenomena in welds III; 2011. p. 353–65.
[48] Buchanan DJ, John R. Residual stress redistribution in shot peened samples subject to mechanical loading. Mater Sci Eng A 2014;615:70–8.
[49] Zhang Zhao, Zhang Hongwu. The simulation of residual stresses in friction stir welds. J Mech Mater Struct 2007;2(5):951–64.
[50] Hwang Yeong-Maw, Kang Zong-Wei, Chiou Yuang-Cherng, Hsu Hung-Hsiou. Experimental study on temperature distributions within the workpiece during friction stir welding of aluminum alloys. Int J Mach Tool Manufact 2008;48:778–87.
[51] Yaduwanshi DK, Bag S, Pal S. Numerical modeling and experimental investigation on plasma-assisted hybrid friction stir welding of dissimilar materials. Mater Des 2016;92:166–83.
[52] Nandan R, Roy GG, Debroy T. Numerical simulation of three-dimensional heat transfer and plastic flow during friction stir welding. Metall Mater Trans 2006;37(4):1247–59.
[53] Shi L, Wu CS, Chen MA. Numerical analysis of heat transfer and material flow in reverse dual-rotation friction stir welding. In: Proceedings of the 1st international joint symposium on joining and welding, Osaka, Japan, 6–8 November; 2013.
[54] Kou S. Welding metallurgy. second ed. New York: John Willy & Sons; 2002.
[55] Mahoney MW, Rhodes CG, Flintoff JG, et al. Properties of friction-stir-welded 7075 T651 aluminum. Metall Mater Trans A 1998;29:1955–64.
[56] Mohammadzadeh Jamalian H, Farahani M, Besharati Givi MK, et al. Study on the effects of friction stir welding process parameters on the microstructure and mechanical properties of 5086-H34 aluminum welded joints. Int J Adv Manuf Technol 2016;83:611–21.
[57] Kasman Ş, Yenier Z. Analyzing dissimilar friction stir welding of AA5754/AA7075. Int J Adv Manuf Technol 2014;70:145–56.
[58] Li Yupeng, Sun Daqian, Gong Wenbiao. Effect of tool rotational speed on the microstructure and mechanical properties of Bobbin tool friction stir welded 6082-T6 aluminum alloy. Metals 2019;9(8):894.
[59] Good SH, Brown IM. Acta Mater 1979;27:1.
[60] Tszeng TC. Interfacial stresses and void nucleation in discontinuously reinforced composites. J Eng Mater Technol 2000;122:86–92.

Please cite this article as: Mehdi H, Mishra RS, Effect of friction stir processing on mechanical properties and heat transfer of TIG welded joint of AA6061 and AA7075, Defence Technology, https://doi.org/10.1016/j.dt.2020.04.014

Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 06, June-2020

Effect of Mesh Size on Soil-Structure Interaction in Finite Element Analysis

Nasasira Derrick[1]
Department of Civil and Environmental Engineering,
Delhi Technological University,
New Delhi, India.

Amit Kumar Srivastava[2]
Department of Civil and Environmental Engineering,
Delhi Technological University,
New Delhi, India.

***Abstract*- Numerous consitutive models have been developed over the past years for modelling the stress-strain behaviour of soils, using finite element and/or finite difference calculations of soil structures and soil/structure interaction problems under axisymetric, plane strain, and/or general three dimensional conditions. Selection of the appropriate parameters and soil model can have a significant impact on the results of numerical analysis. In this study, the interaction of soil and reinforced cement concrete raft foundation in a sandy soil at a depth of 9m below the existing ground level is analysed by using with both analytical and finite element analysis (FEM). Analytical calculations were done using method available in literature and a finite element analysis was carried out using the hardening soil model available in PLAXIS 2D (2020, connect version). The results obtained from analytical calculations were compared with those obtained from the Hardening Soil (HS) Model. In the model sizes of mesh were changed and there effect on interaction behaviour is observed by studying the strength and deformation results obtained from FEM analysis. It was observed that very fine mesh size gave more accurate results compared to the other sizes.**

Keywords- Plaxis, sand, settlement, Raft

I. INTRODUCTION

Evaluation of correct strength and deformation behaviour of raft resting on the soil is important for safe and economical design. Strength and deformation behaviour of raft foundation can be analysed by different numerical methods. With the development of numerical methods such as finite element anaylsis and finite difference method, it has become feasible to analyse and predict the behaviour of complex soil structures and soil/structure interaction problems. Such anaylses depend on the representation of the relations between stresses and strains in a given material and are represented by a consitutive model, consisiting of mathematical expressions that model the behaviour of soil in a single element. Therefore the purpose of a consitutive model is to simulate the soil behaviour with sufficient accuracy under all conditions with sufficient accuracy under all loading conditions in the numerical computations.

II. LITERATURE REVIEW

The Hardening Soil model was established in 1999 [1]. In the framework of the theory of elasticity. The model involves friction hardening to model the plastic shear strain in deviatoric loading, and cap hardening to model the plastic volumetric strain in primary compression.because of these two types of hardening, the model is accurate for problems (like excavations) involving a reduction of mean effective stress and at the same time mobilization of shear strength. The verification and formulation of this model was discussed [1]. It is an advanced and latest model for the simulation and modelling of soil behaviour. The Hardening soil model requires the input of 10 parameters, i.e. three reference stifness parameters (E_{50}^{ref} for triaxial compression, E_{ur}^{ref} for triaxial unloading and E_{oed}^{ref} for oedometer loading) at a reference stress level P^{ref}, a power, *m*, for the stress dependent stifness formulation, poisson's ratio for unloading and reloading, μ_{ur}, the mohr coulomb strength parameters, Ø, and c, the dilatancy angle, ψ, the K_o value in primary one-dimensional compression (K_o^{nc}), a parameter called the failure ratio, R_f which determines the strain level at failure [2]. Fig. 1 shows the failure criterion of the harening soil model in principal stress space. The yield surface of the hardening soil model is not fixed in principal stress space, but it can expand due to plastic straining. The Plaxis hardening model can be used to accurately predict displacement and failure for general types of soils in various geotechnical applications [3]. The model does not include anisotropic strength or stiffness, nor time-dependent behaviour (creep). The hardening soil model is also known for stress-dependency of stiffness moduli [4]. This explains that all stiffnesses increases with the pressure.

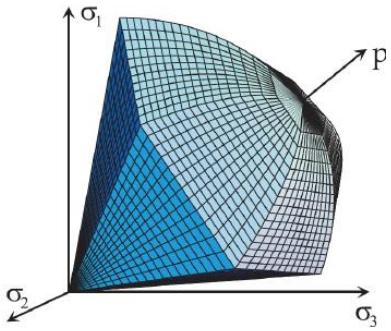


Figure 1: Failure criterion of the Hardening soil model in principal stress space [1].

The Hardening Soil model, also called as HS-Standard was explained by [1]. In order to modify the important phenomena which is exhibited by soils such as:

1. Densification, defines as decrease of the voids volume in soil due to plastic deformations, which also decreases the void ratio,

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 06, June-2020

2. Stress dependent stiffness, is basically an observed technique of increasing the stiffness modules with the increasing of confining stress,
3. Soil stress history, explains the accounting for preconsolidation process and its effects,
4. Plastic yielding, is very important which accounts for the development of irreversible strains with reaching a yield criteria,
5. Dilatancy, explains as the uses for an occurrence of negative volumetric strains during shearing.

This model is known to be one of the simplest and important in the class of latest models designed to handle the small strain stiffness as explained by [5]. It has two plastic mechanisms, shear and volumetric. Different to the other material models such as the Cap model or the Modified Cam Clay model, the magnitude of soil deformations, stresses, forces, etc. can be modelled more accurately by using three different input stiffness parameters which is related to triaxial loading stiffness (E_{50}), triaxial unloading-reloading stiffness (E_{ur}), and oedometer loading modulus (E_{oed}).
There are five parameters in the hardening soil model differ than the MC model is:

1. (E_{50ref}) secant modulus 50% strength[kN/m^2]
2. (E_{oedref}) Oedometric modulus [kN/m^2]
3. (E_{urref}) Unloading-reloading modulu[kN/m^2]
4. (ν_{ur}) Unloading-reloading Poisson's ratio [-]
5. (m) Exponent of the stress-stiffness function

Parameters previously used in MC model:

1. (Ø) Friction angle [°]
2. (c) Cohesion [kN/m^2]
3. (ψ) Dilatancy angle [°]

The hardening model can be used to predict displacement and failure for different types of soils in various geotechnical applications. This model does not include anisotrpic strength or stifness, nor time-dependent behaviour (creep). It's capabilities for dynamic applications are limited, but this model is, in principle, without any doubt the most accurate general soil model.

A. *Correlation between Angle of Friction and Standard Penetration Number*

The peak friction angle Ø', of granular soil has also been correlated with N_{60} by several investigators. A correlation between N_{60} and Ø' was given by [6] in a graphical form, which can be approximated as,

$$Ø'(deg) = 27.1 + 0.3N_{60} - 0.00054[N_{60}]^2 \quad (1)$$

B. *Correlation between Modulus of Elasticity and Standard Penetration Number*

The modulus of elasticity of granular soils (E_s) is an important parameter in estimating the
elastic settlement of foundations. A first order estimation for E_s was given by [7] as

$$\frac{E_S}{p_a} = \propto N_{60} \quad (2)$$

Where
α = 5 for sands with fines
10 for clean normally consolidated sands
15 for clean overconsolidated sands
p_a= atmospheric pressure in the same units as E_s
N_{60}= Field standard penetration number

III. PROBLEM STATEMENT

Bearing capacity requirement is a basic criteria to be satisfied in the analysis and design of shallow foundations. The criterion on bearing capacity ensures that the foundation does not undergo shear failure under loading. The use of standard penetration test in the analysis of bearing capacity has received numerous attentions [8]. Details of the field application of Standard Penetration Test are specified in IS 2131 [9]. This paper attempts to report interaction behavior of raft foundations resting on sand using Finite element software package provided by Plaxis 2D (connect editon). Steps for modeling is presented in fig. 2 from the model definition to the output results

IV. SELECTION OF PARAMETERS

Parameters used in this study were obtained through laboratory investigation of the undisturbed samples collected from site located at Noida in India, as most of the strata is having poorly graded sand, collection of undisturbed sample was difficult. Hence, the result of standard penetration test conducted as per IS 2131 [9] is used for analysis . Table 1 summarizes the selection of model paramters for analysis of the hardening soil model used in this study.

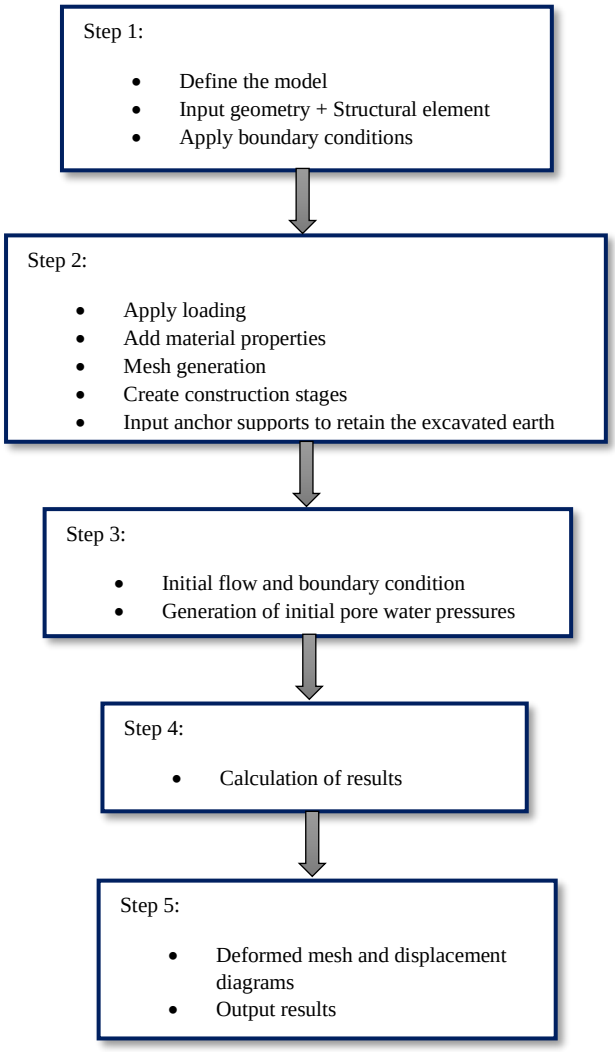


Figure 2: Steps for numerical modeling

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 06, June-2020

SOIL PROFILE AND BOREHOLE DETAILS

The samples were collected from two bore holes drilled at GH-01A/01, SECTOR-16, GREATER NOIDA WEST, U.P (India) and tested as per the relevant protocol of bureau of Indian Standard. The results of SPT are presented in Table2.

ANALYTICAL METHOD

The analytical calculations were done in accordance with Indian standards IS 6403 [10] and permissible settlement is restricted as per IS 1904 [11].

Based on the shear criterion IS code 6403 [10], net ultimate bearing capacity for local shear failure for a cohesionless sand

$$q_{nu} = q(N_q - 1)s_q d_q i_q + \frac{1}{2} B\gamma N_\gamma s_\gamma d_\gamma i_\gamma w' \quad (3)$$

Where s_q, s_γ= shape factors

d_q, d_γ= depth factors

i_q, i_γ= Inclination factors

N_q, N_γ= Bearing capacity factors

w', w''= Correction factor for location of water table.

Assuming there will be rise in water table during mansoon and water will reach to a level of bottom of raft.

$$\emptyset' = tan^{-1}(0.67 tan\emptyset) = 22.78^o$$

Nq=8.76, Nγ=8.44

$$from\ equation\ (3), q_{nu} = \gamma D_f (N_q - 1)s_q d_q i_q w' + \frac{1}{2} B\gamma N_\gamma s_\gamma d_\gamma i_\gamma w'' = 2209.88 KN/m^2$$

Assuming a factor of safety of 2.5, net safe bearing capacity(q_{ns}) $= 883.95 KN/m^2$

TABLE 1: SELECTION OF MODEL PARAMETERS FOR ANALYSIS USING THE HARDENING SOIL MODEL

General			
Material model	Model	Hardening soil model	-
Type of material behaviour	Type	Drained	-
Soil unit weight above phreatic level	γ_{unsat}	16.20 *(lab)*	KN/m^3
Soil unit weight below phreatic level	γ_{sat}	19.10 *(lab)*	KN/m^3
Parameters			
Young's modulus	E'	-	KN/m^2
Secant stiffness in standard drained triaxial test	E_{50}	$3.0x10^4$ *(Plaxis)*	KN/m^2
Tangent stiffness for primary oedometer loading	E_{oed}	$3.0x10^4$ *(Plaxis)*	KN/m^2
Unloading/reloading stiffness	E_{ur}	$1.00x10^5$ *(Plaxis)*	KN/m^2
Power for stress-level dependency of stiffness	m	0.5 *(Plaxis)*	-
Cohesion (constant)	C	0 *(lab)*	KN/m^2
Friction angle	Ø	32 *(lab)*	^{o}
Dilatancy angle	ψ	2.0 *(lab)*	^{o}
Poisson's ratio	μ	0.2 *(lab)*	-
K_0-value for normal consoilidation	K_0^{nc}	0.4701 *(plaxis)*	-
Interfaces			
Strength reduction factor Inter.	R_{inter}	0.67 *(plaxis)*	-
Initial			
K_0 determination	-	Automatic	-
Over-consolidation ratio	OCR	1.0 *(plaxis)*	-

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 06, June-2020

TABLE 2: SPT VALUES FOR DIFFERENT BOREHOLES

	B1		B2	
Depth (m)	Field N-Value	Corrected N-value	Field N-Value	Corrected N-value
1.5	10	17	7	12
3	12	17	9	13
4.5	18	21	13	15
6	20	20	17	17
7.5	24	22	19	17
9	25	21	24	20
10.5	26	20	27	21
12	34	24	29	21
13.5	37	25	30	18
15	39	20	34	19
18	44	21	39	20
21	51	22	46	21
24	56	23	52	22
27	58	23	54	22
30	59	23	59	23
33	55	21	62	23
36	60	22	66	23
39	68	23	71	24
42	70	23	76	25
45	75	24	81	25

TABLE 3: PARAMETER USED FOR CALCULATION BY ANALYTICAL METHOD

Bore hole No	1 & 2	Unit	Protocol
Depth of water table	14	m	
density above water table	16.2	KN/m^3	
density below water table	19.1	KN/m^3	
Type of foundation	raft		
width of foundation	40	m	
length of foundation	42	m	
depth of foundation	9	m	
Permissible settlement of foundation	0.075	m	
Cohesion (c)	0	kN/m^2	Direct Shear Test
Angle of internal friction (φ)	32	Degree	

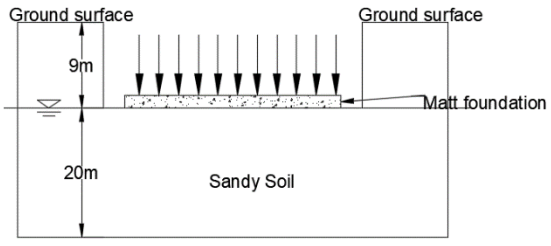


Figure 3: Prototype dimensions

Bearing capacity based on the settlement criterion

$N_{cor} = 22$
$\gamma_d = 16.2KN/m^2$
$\gamma_{sat} = 19.1KN/m^2$
$\mu = 0.2$
$C = 0$
$\emptyset = 32^o$

$$S_e = \frac{q_n B(1-\mu^2)I_F}{E_s} \quad (4)$$

Where q_n= Net safe bearing pressure

B = breadth of foundation

E_s= Young's modulus

I_F = Influence factor

s_e = allowable settlement (75mm)

at

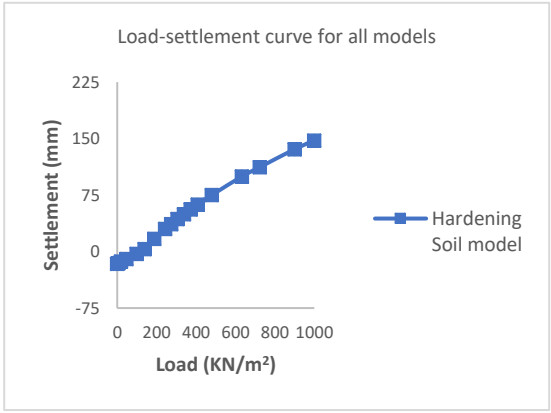


Figure 4: Load-settlement relationship for the Hardening soil model

75mm allowable settlement, from equation 4, q_n= 456.7KN/m^2.Therefore, considering the lower of the two values obtained, Allowable bearing capacity of the foundation was obtained as 456.7KN/m^2.

V. NUMERICAL ANALYSIS

Plaxis 2D was used to evaluate the performance of the Hardening Soil model model to estimate the bearing capacity of a raft foundation situated at a depth of 9m below the surface and to investigate the relationship between the results obtained. Fig 3 shows a sketch of the prototype dimensions showing the location of the water table. Simulations for different excavation stages were done. The distance between the boundary of the floor of the model and the lower boundary of the footing plays an important role in the pattern of bearing capacity value obtained.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 06, June-2020

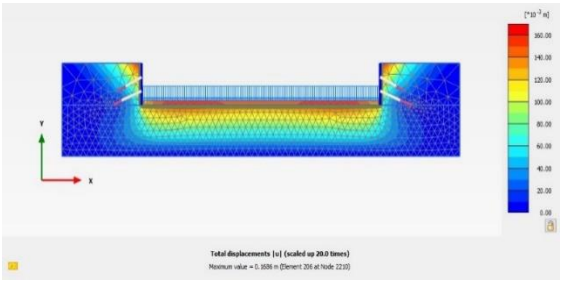


Figure 5: Output image showing total vertical displacement of foundation

To eliminate the effect of this factor, the bottom line intended at a distance greater than 2d.

It can be seen from fig 4, that Plaxis 2D software provides a linear load-settlement relationship for the hardening soil model, which is variant from the theoretical load-settlement curve for a foundation on sandy soil. The curve also indicates an initial negative settlement indicating a swelling in the soil before the footing is placed and load is applied. The reason for initial negative settlement is due to release of in situ soil pressure due to excavation.

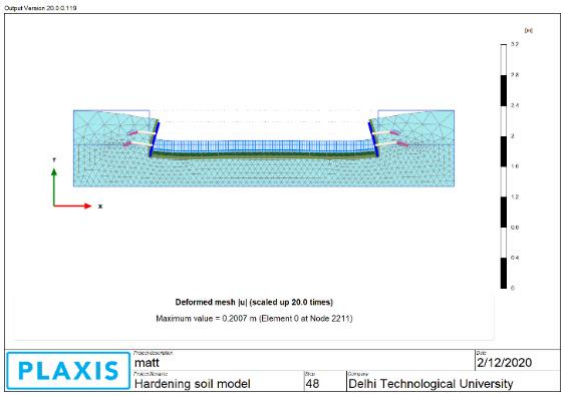


Figure 6: Deformed mesh of foundation.

It can be seen from the graph that settlement increases directly proportional to increase in stress applied on foundation. A value of 481KN/m^2 is obtained at 57mm settlement. It can be seen that the hardening soil model provides a result close to the value obtained from the IS code calculation (456.7KN/m^2) therefore making it more accurate to use in the field for design of foundations. Fig 5 shows the deformation of the raft footing after applying the design load. The contours are drawn for different displacements value below the raft.

Fig 6 shows the typical deformed mesh of the foundation. Three excavation stages were input into Plaxis software since the foundation is laid at a 9m depth. To support the dry excavation, anchors were used. The anchors were prestressed with 500KN force.

VI. EFFECT OF CHANGES OF MESH SIZE ON THE BEARING CAPACITY OF FOUNDATION

Here, the effect of mesh size changes on the model output results was analysed. Different mesh sizes were used in Plaxis 2D which are; very coarse, coarse, medium, fine and very fine. It should be noted that the mesh size of 15 nodal element, is a better distribution of nodes and more accurate results than to element with six nodes. Therefore, the analysis of 15 nodal element is used. Table 3 shows the effect of changes of mesh size on settlement pressure of foundation for 75mm settlement. Figure 7 shows the effect of change in mesh size on deformation of a raft foundation.

TABLE 3: EFFECT OF CHANGE IN MESH SIZE ON ALLOWABLE BEARING CAPACITY OF FOUNDATION

Allowable bearing capacity of foundation at 75mm settlement (KN/m^2)	Change in type of mesh size
494	Very coarse
488	Coarse
480	Medium
465	Fine
460	Very fine

From Fig 7, it can be seen that there is slightly or no difference in the pressure for 75mm of settlement. But at settlement more than 75 mm the effect of mesh size is to decrease the settlement for a given pressure if mesh size is increased from very fine to very coarse.

VII. CONCLUSION

Based on the findings, the following conclusions can be drawn;

1. The bearing pressure for a given settlement obtained by the hardening soil model was closer to the IS analytical method. However an initial settlement of 20mm should be deducted to cater for the swelling provided by plaxis FEM, because of removal of insitu stress because of excavation.
2. Mesh size with 15 nodal elements produces more accurate results as compared with element with six nodes.
3. The mesh sizes influences the bearing pressure for a given settlement. The influences are negligible at a settlement value of 75mm, but influences increase with increase in settlement.
4. Bearing capacity value increases with increase in refinement of mesh size in Finite element analysis. When a finer mesh is used, the value of bearing capacity obtained at a particular settlement is higher than that obtained when a coarse mesh is used. This therefore means researchers should ensure that a fine mesh is used when carrying out finite element analysis to obtain more accurate results.

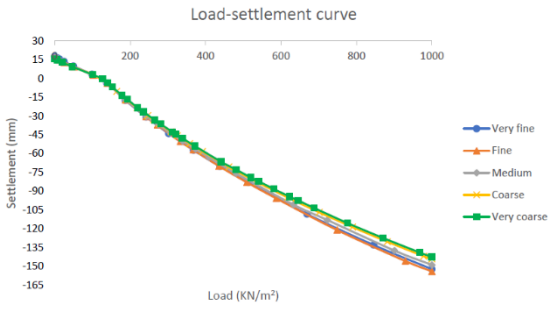


Figure 7: Load-settlement graph of the different mesh sizes

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 06, June-2020

ACKNOWLEDGMENT

The authors wish to thank the Department of Civil Engineering, Delhi Technological University. This work was supported in part by a grant from delhi Technological University, Department of Civil and Environmental Engineering.

REFERENCES

[1] Schanz T, Vermeer A, Bonnier P, (1999), "The hardening soil model: formulation and verification." Beyond 2000 Computional Geotechnics-10 years PLAXIS, Amsterdam, Netherlands, 281p.
[2] Brinkgreve RBJ, Al-Khoury R, (2020), "PLAXIS version 2020 reference manual".
[3] Braja, MD. (1999), "Principle of Foundation Engineering (4th edi.)". PWE Publishing Company, USA. pp243.
[4] Labuz, Joseph & Zang, Arno. (2015). "Mohr-coulomb failure criterion". The ISRM Suggested Methods for Rock Characterization, Testing and Monitoring: 2007-2014. 227-231.
[5] Hewitt, P. B. and Gue, S. S., (1994). "Piled Raft Foundation in a weathered sedimentary formation. Proc. Geotropica", Malaysia, 1-11
[6] Peck, Hanson and Thornburn (1974), "Modification of Teraghi and Peak Correlation by Peak, Hanson and Thornburn"
[7] Kulhawy, F.H. & Mayne, Paul, (1990). "Manual on Estimating Soil Properties for Foundation Design".
[8] Poulos, H.G., (2005). "Piled raft and compensated piled raft foundations for soft soil sites. Advances on designing and testing deep foundations". Geotech. Spec. Publ. (ASCE) 129: 214–235.
[9] Indian standards, IS:2131. (1981), "Method for standard penetration test for soils"
[10] Indian standards, IS:6403. (1981), "Code of Practice for Determination of Bearing Capacity of Shallow Foundations".
[11] Indian standards, IS:1904. (1986), "Code of practice for design and construction of foundations in soils : general requirements".
[12] Józsa, Vendel. (2011), "Effects of rarely analyzed soil parameters for FEM analysis of embedded retaining structures".
[13] Shield RT (1955)., "On Coulomb's law of failure in soils". J Mech Phys Sol 4:10–16
[14] Jerry A Yamamuro., Victor N,(2005), "Soil constitutive models, Evaluation, selection, and calibration". Geotechnical Special Publication No.128.
[15] Meyerhof, GG. (1956), "Penetration Test and Bearing Capacity of Cohesionless Soils". Journal of the Soil Mechanics and Foundation Division, ASCE. 82(1):1-19.
[16] Adachi, T., Oka, F., (1982). "Constitutive equation for normally consolidated clays based on elasto-viscoplasticity" Soils and foundations 22:57-70
[17] Broja, R.I., Kavaznjian, E, (1985). "A constitutive model for the σ-ε-t behaviour of wet clays". Geotechnique 35: 293-298.
[18] Russo, G., (1998). "Numerical Analysis of Piled Rafts. Int. J. Numer. Anal. Meth. Geomech", 22, 477-493.
[19] Prakoso, W. A. and Kulhawy, F. H., (2001). "Contribution To Piled Raft Foundation Design. Journal of Geotechnical and Geoenvironmental Engineering", ASCE, 127(1), 17-24.
[20] Reul, O. and Randolph, M. F., (2003). "Piled Rafts in Overconsolidated Clay: Comparison of in situ Measurements and Numerical Analysis". Géotechnique, 53(3), 301-315.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

Engineering Research Express

ACCEPTED MANUSCRIPT

Efficient strip to slot waveguide mode converter using fast quasi-adiabatic approach

To cite this article before publication: Nikhil Dhingra *et al* 2020 *Eng. Res. Express* in press https://doi.org/10.1088/2631-8695/aba819

Manuscript version: Accepted Manuscript

Accepted Manuscript is "the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an 'Accepted Manuscript' watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors"

This Accepted Manuscript is **© 2020 IOP Publishing Ltd**.

During the embargo period (the 12 month period from the publication of the Version of Record of this article), the Accepted Manuscript is fully protected by copyright and cannot be reused or reposted elsewhere.
As the Version of Record of this article is going to be / has been published on a subscription basis, this Accepted Manuscript is available for reuse under a CC BY-NC-ND 3.0 licence after the 12 month embargo period.

After the embargo period, everyone is permitted to use copy and redistribute this article for non-commercial purposes only, provided that they adhere to all the terms of the licence https://creativecommons.org/licences/by-nc-nd/3.0

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions will likely be required. All third party content is fully copyright protected, unless specifically stated otherwise in the figure caption in the Version of Record.

View the article online for updates and enhancements.

This content was downloaded from IP address 14.139.251.100 on 28/07/2020 at 07:00

IOP Publishing
Journal **XX** (XXXX) XXXXXX https://doi.org/XXXX/XXXX

Efficient strip to slot waveguide mode converter using fast quasi-adiabatic approach

Nikhil Dhingra[1], Geetika Jain Saxena[2], Kamal Kishor[3] and Enakshi Khular Sharma[1]

[1] Department of Electronic Science, University of Delhi South Campus, Benito Juarez Marg, South Moti Bagh, New Delhi - 110021, India
[2]Department of Electronics, Maharaja Agrasen College, University of Delhi, Vasundhara Enclave, New Delhi - 110096, India
[3]TIFAC-Centre of Relevance & Excellence in Fiber Optics & Optical Communications, Department of Applied Physics, Delhi Technological University, Bawana Road, Delhi – 110042, India

E-mail: ndhingra@south.du.ac.in

Received xxxxxx
Accepted for publication xxxxxx
Published xxxxxx

Abstract

We present an ultra-compact strip to slot waveguide mode converter to facilitate excellent fundamental mode coupling between the strip and slot waveguide over a broad wavelength range. The design utilizes tapered waveguide geometries for adiabatic mode conversion. The taper profiles are optimized using a fast quasi-adiabatic approach to achieve the adiabatic mode conversion within the smallest possible length. We numerically demonstrate a conversion efficiency of 99.4% at a wavelength 1.55 μm with a variation of <1.5% over a bandwidth of 100 nm and <5% over a bandwidth of 200 nm. In addition, the design is ultra-compact with a footprint of only 1.23×3.7 μm^2 and offers a high tolerance to the possible fabrication inaccuracies.

Keywords: adiabatic tapers, quasi-adiabatic dynamics, strip to slot converter, silicon photonics

1. Introduction

The slot waveguide can confine light within a nanoscale low index gap between two high index strips and hence facilitates strong light matter interaction. This property of slot waveguide has been utilized in a number of applications including optical modulation [1], all-optical signal processing [2,3], biosensing [4-6] and absorption-based sensing [7]. The use of polymer filled slot waveguide for modulation and all-optical signal processing allows efficient use of the non-linearity in polymers [1-3]. In biosensing applications, a strong interaction between the waveguide mode and the functionalized waveguide surface of the slot waveguide results in a better sensitivity with smaller device footprint [8,9]. However, scattering losses associated with slot waveguides are, in general, higher than that of the strip waveguides. Hence, the use of slot waveguides is better suited to the abovementioned applications while for most of the other photonic integrated circuits (PICs), the strip waveguide is a better choice due to its low loss and tight confinement. Also, the efficient designs of grating couplers and edge couplers, to allow coupling of light between the photonic chip and an optical fiber, are available for the strip waveguides [10]. To benefit from the advantages of both the waveguide geometries, there is a need for efficient coupling between the fundamental mode of the strip waveguide and the slot waveguide. Direct butt coupling results in a high loss due to a large mismatch between the fundamental mode of the strip and the slot waveguides and hence a mode converter is needed for efficient coupling of light between these waveguides. Few of the reported designs of the

 © xxxx IOP Publishing Ltd

Accepted Manuscript

strip to slot waveguide mode converters are solely based on adiabatic mode evolution [11-13]. For adiabatic mode evolution based V-shaped mode converter [11] and ultra-low loss converter [12], the coupling efficiency strongly depends upon the sharpness of the taper tip and hence poses challenges in terms of fabrication. The other design [13] based on adiabatic mode evolution is improved in terms of fabrication tolerance; however, the device footprint can be significantly improved by the appropriate choice of the taper profiles. The other category of strip to slot converter is based on split and combine technique. This technique requires the use of an addition splitter section prior to an adiabatic mode evolution based combiner section [14-17]. The technique utilizes either a y-splitter [14] or a MMI [15-17] to split the light into two strip waveguides followed by a tapered section for adiabatic evolution of the fundamental mode of these evanescently coupled strip waveguides into the fundamental mode of the slot waveguide. The performance of these designs is limited by the 3 dB response of the y-splitter or MMI. Also, the design with y-splitter has a very large device footprint and requires integrated heaters to compensate the phase error due to the differential length between the two arms [14].

We propose a compact and broadband strip to slot waveguide mode converter based on adiabatic mode evolution. In order to achieve adiabatic mode conversion within the smallest possible length, we optimize the taper profiles using a fast quasi-adiabatic (FAQUAD) [18,19] approach. Using the optimized tapers, a very high mode conversion efficiency is obtained over a broad wavelength range with a footprint of only 1.23×3.7 μm^2. The design also offers a very high tolerance to the possible fabrication inaccuracies.

2. Design criteria and optimization

The proposed strip to slot waveguide mode converter is designed for silicon-on-insulator platform with dimensions of the slot waveguide optimized specifically for biosensing applications [9]. Cytop is chosen as the cladding material as its refractive index ($n_c = 1.3335$ at $\lambda = 1.55$ μm [20]) matches well with the aqueous solutions used for biosensing and hence makes it an appropriate choice for cladding or microfluidic channel in integrated photonic biosensors [21-23]. Optimizing the mode converter with Cytop cladding also allows the freedom to use the mode converter within the sensing region with no significant change in the conversion efficiency. The cross-sectional view of the strip and slot waveguides, with corresponding dimensions and the field profile (E_x) of the fundamental quasi-TE mode, are shown in Fig. 1. The effective indices of the TE_0 mode of the strip waveguide and slot waveguide are obtained as 2.4277 and 1.6810, respectively.

It is clearly visible from the modal profile of the slot waveguide [Fig. 1], that the electric field is confined to the low index slot between the two silicon strips which is also evident from its low effective index. For direct butt coupling between strip waveguide and slot waveguide, with dimensions given in Fig. 1, a maximum coupling efficiency of only 70% is possible.

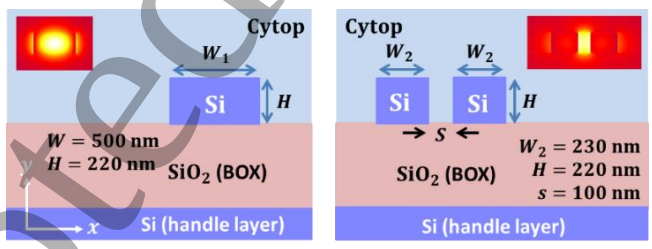


Fig. 1. Cross-sectional view of the cytop-clad silicon strip and slot waveguides with corresponding fundamental TE mode profile (E_x)

In waveguide based biosensing, the primary aim is to maximize the interaction between the waveguide mode and the functionalized waveguide surface. To find the optimum dimensions (W_2 and S) of the slot waveguide, we assume a biorecognition layer with refractive index 1.5 and thickness t directly above the waveguide surface (see inset of Fig. 2). As the sensitivity of a bio-sensor depends on the interaction between the propagating mode and the bio-recognition layer, the fill factor (FF), which represents the fraction of mode power confined in a given region, is calculated in the bio-recognition layer. For three values of S, equal to 80 nm, 100 nm, and 120 nm, the variation in the FF is shown as a function of W_2 in Fig. 2(a) and Fig. 2(b) for t = 5 nm and t = 10 nm, respectively. For each value of S, there exists an optimum value of W_2, for which the FF of the mode is maximum within the biorecognition layer. The FF increases with a decrease in S, however, the minimum value of S, which can be used is limited by the fabrication constraints. A spacing equal to or greater than 100 nm can be easily fabricated using

Accepted Manuscript

the 193 nm immersion photolithography and hence, we fix S = 100 nm. For S = 100 nm, the FF does not change much for values of W_2 between 210 nm and 230 nm. Hence, we choose W_2 = 230 nm for the ease of fabrication. From Fig. 2(a) and Fig. 2(b), it can be observed that the optimized dimensions of the slot waveguide do not change significantly with t.

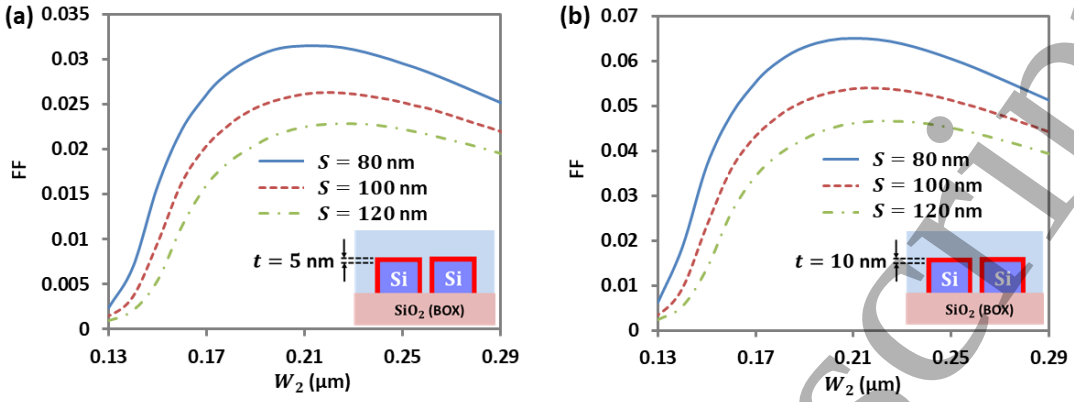


Fig. 2 Fill factor for the fundamental quasi-TE mode in the biorecognition layer of thickness (a) 5 nm and (b) 10 nm

The top view of the proposed strip to slot waveguide mode converter, consisting of two tapered sections (I and II), is shown in Fig. 3. In section I, a 230 nm strip is gradually coupled to the input strip waveguide by reducing the spacing from 500 nm to 100 nm. This ensures adiabatic coupling between the fundamental mode of the strip waveguide and the fundamental supermode of the evanescently coupled waveguides with a spacing of 100 nm at interface II. In section II, the width of the input strip waveguide is adiabatically reduced from 500 nm to 230 nm to form a slot waveguide.

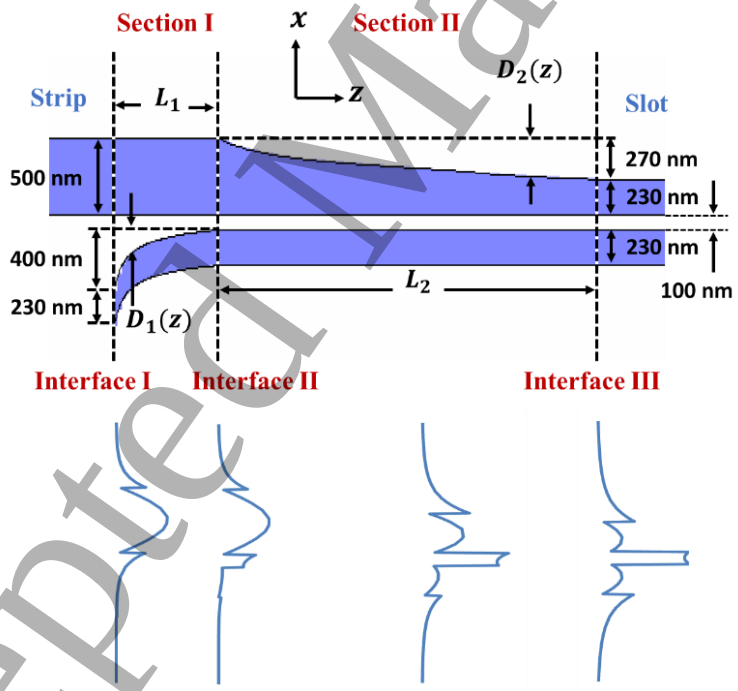


Fig. 3. Top view of the proposed design of mode converter with two tapered sections (I and II) along with the modal profiles of the fundamental quasi-TE mode at different positions along the length of the converter

The field profile (E_x), shown at different positions along the converter, shows the evolution of the fundamental mode as it propagates through the device. It is also possible to directly connect the 500 nm strip at interface II without the need for section I; however, this will result in the power loss at the interface II due to a small difference in the modal profiles of the fundamental mode of the strip waveguide and the fundamental supermode on the right hand side of interface II. Also, without the section I, the fabrication imperfections at the front edge of the 230 nm waveguide may affect the performance. Hence, we choose to design the mode converter with both tapered sections

Accepted Manuscript

I and II.

The tapering in both the sections should be slow enough to support adiabatic mode evolution, i.e., if the fundamental mode is excited at the input interface I, the power should remain in the fundamental mode as the geometry changes from a strip waveguide to a slot waveguide. A linear taper, in general, requires large length to ensure such an adiabaticity and hence results in a large device footprint. Different taper shapes [24,25] and techniques [26,27] to obtain taper profiles to support adiabatic mode evolution with shorter length of tapered sections have been suggested. However, none of these techniques provide a simple recipe to obtain the optimum taper profile for a specific application. In quantum mechanics, where the Hamiltonian changes with time, the adiabatic theorem defines an adiabaticity condition [28] in which a particle in the n^{th} eigenstate of the initial Hamiltonian remains in the n^{th} eigenstate of the changing Hamiltonian. An approach called fast quasi-adiabatic (FAQUAD) [29] can be used to optimize a single control parameter in the changing Hamiltonian of a quantum system to perform the process as fast as possible and still keeping it as adiabatic as possible. The optimization of the control parameter is done by maintaining the adiabaticity parameter to a small constant value over the entire process duration. The FAQUAD approach has been successfully translated to waveguide geometries, which vary along the propagation direction. It has been used in designing a variety of integrated photonic devices, based on z-varying optical waveguide structures, including a mode-sorting asymmetric Y-junction [30], a polarization splitter-rotator [31] and a 3-dB coupler [32]. For optical waveguides, the adiabaticity parameter (c_{nm}) for controlling the power transfer from m^{th} mode to n^{th} mode along a z-dependent waveguide geometry is given by [18,19]

$$c_{nm}(z) = \left| \frac{\iint \left[\frac{\mathrm{d}\boldsymbol{E_m}}{\mathrm{d}z} \times \boldsymbol{H_n} \right] \cdot \hat{z}\, \mathrm{d}x\mathrm{d}y}{(\beta_m - \beta_n)} \right| \tag{1}$$

where, $\mathrm{d}\boldsymbol{E_m}/\mathrm{d}z$ represents the derivative of the electric field of the m^{th} mode at a position z along the propagation length, $\boldsymbol{H_n}$ represents the magnetic field of the n^{th} mode with β_m and β_n the propagation constants of m^{th} mode and n^{th} mode, respectively. To ensure the adiabatic mode evolution, the adiabaticity parameter $c_{nm}(z)$ should be less than a desired small value ε throughout the length. For the proposed design, the FAQUAD protocol is used to distribute the adiabaticity uniformly over the entire taper length by controlling a single taper parameter. In our design, this taper parameter $D(z)$ is equals to $D_1(z)$ in section I and $D_2(z)$ in section II, as shown in Fig. 3. The Eq. (1) can also be represented in terms of $D(z)$ as

$$c_{nm}(z) = \left| \frac{\iint \left[\frac{\mathrm{d}\boldsymbol{E_m}}{\mathrm{d}D} \times \boldsymbol{H_n} \right] \cdot \frac{\mathrm{d}D}{\mathrm{d}z} \hat{z}\, \mathrm{d}x\mathrm{d}y}{(\beta_m - \beta_n)} \right| = F_{nm}(D) \frac{\mathrm{d}D}{\mathrm{d}z} \tag{2}$$

where, $F_{nm}(D)$ is given by

$$F_{nm}(z) = \left| \frac{\iint \left[\frac{\mathrm{d}\boldsymbol{E_m}}{\mathrm{d}D} \times \boldsymbol{H_n} \right] \cdot \hat{z}\, \mathrm{d}x\mathrm{d}y}{(\beta_m - \beta_n)} \right| \tag{3}$$

For ideal strip to slot mode conversion, the fundamental mode excited at interface I should propagate adiabatically throughout the length of the tapers to transform into the fundamental mode of the slot waveguide. In principle, a waveguide supports a few guided modes and infinite number of radiation modes. However, the radiation modes can be discretized to a finite number of modes using appropriate computational boundaries. The mode solvers in Photon Design [33] use reflecting boundaries to obtain the complete basis set in terms of finite number of guided and radiation modes. Hence, to ensure the adiabatic mode evolution of the fundamental mode, the sum of the adiabaticity parameters over all the N higher order modes, i.e., $c(z) = \sum_{n=2}^{N} c_{n1} = (dD/dz)F(D)$, where the parameter $F(D)$ is given by $F(D) = \sum_{n=2}^{N} F_{n1}(D)$, should be below a desired small value ε. The finite element method (FEM) and eigenmode expansion (EME) method [34], in Photon Design, are used for the calculation of $F(D)$. The variation in the parameter $F(D)$, represented by $F(D_1)$ for section I and $F(D_2)$ for sections

Accepted Manuscript

II, is shown as a function of D_1 and D_2 in Fig. 4(a) and Fig. 4(b), respectively. For a higher value of $F(D)$, a lower value of dD/dz is required to maintain constant value of $c(z)$ throughout the length of taper. Hence, the taper profile can be chosen in such a way that the slope of the taper is high where $F(D)$ is small and vice versa.

To find the optimum taper profiles $D(z)$, the adiabaticity parameter should be equal to a constant value ε throughout the length of the taper, i.e.,

$$\varepsilon = F(D)\frac{dD}{dz} \tag{4}$$

Integrating from 0 to z along the length of the taper results in

$$\varepsilon\int_0^z dz = \int_0^D F(D)\,dD \tag{5}$$

At $z = L$, where L is the length of the taper, $D = D_f$ and ε can be calculated as

$$\varepsilon = \frac{1}{L}\int_0^{D_f} F(D)\,dD \tag{6}$$

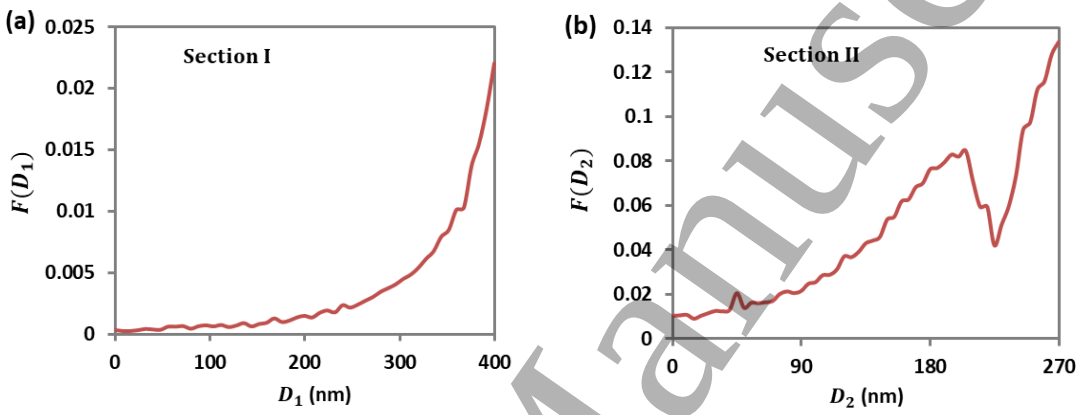


Fig. 4. (a) $F(D_1)$ as a function of D_1 [section I] and (b) $F(D_2)$ as a function of D_2 [section II]

Hence, by substituting the expression for ε from Eq. (6) into Eq. (5), the taper profile $z(D)$, which ensures a constant value of $c(z)$, can be obtained as

$$z(D) = L\frac{\int_0^D F(D)\,dD}{\int_0^{D_f} F(D)\,dD} \tag{7}$$

The optimized taper profiles for tapered sections I and II, obtained using Eq. (7), are shown in Fig. 5(a) and Fig. 5(b), respectively, along with the corresponding linear profiles for comparison. The parameter $F(D_1)$ and $F(D_2)$ and hence the taper profiles $D_1(z)$ and $D_2(z)$ are calculated at wavelength 1.55 μm. The use of these optimized taper profiles with constant $c(z)$ are expected to give rise to a significant reduction in the required length of the taper while maintain a low sensitivity towards wavelength and high tolerance towards the fabrication inaccuracies. The results shown in the section 4 validate these points.

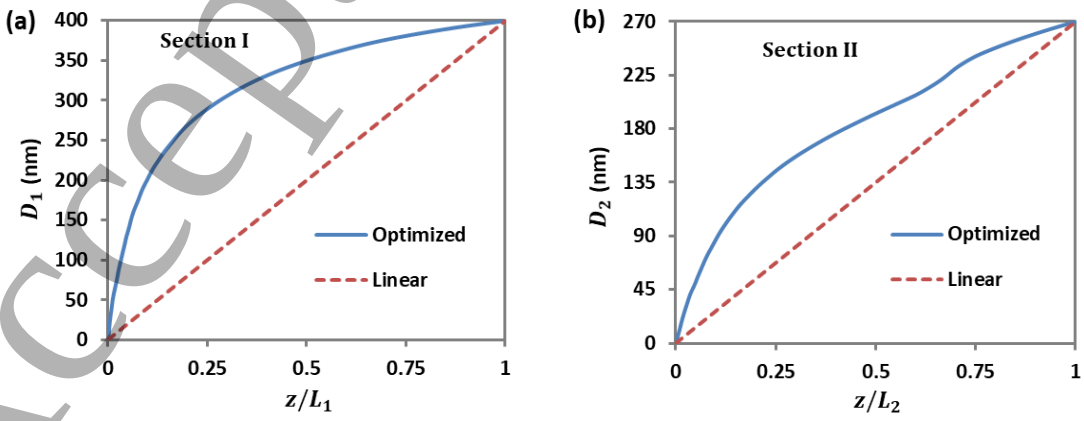


Fig. 5. Optimized taper profiles and corresponding linear profiles as a function of normalized position along the

Accepted Manuscript

length of the taper for (a) section I and (b) section II

3. Propagation through the optimized design

In order to estimate appropriate values of the length of the optimized tapers, we study the fractional power retained in the fundamental mode in propagating from interface I to interface II in section I and from interface II to interface III in section II as a function of lengths L_1 and L_2, respectively. For comparison, the corresponding study for the linear tapers is also carried out. The results shown in Fig. 6(a) and Fig. 6(b) for both linear and optimized tapers clearly show that optimized tapers perform better than the linear tapers at smaller lengths. At higher values of taper length, the conversion efficiency for both linear and optimized taper starts to converge to a maximum value. As the requirement is to obtain an efficient design with footprint as small as possible, the optimization of the taper profiles is important. For section I, although the first maximum in the Fig. 6(a) is observed at a taper length of 1.9 μm, we choose $L_1 = 0.79\,\mu\text{m}$, which provide sufficiently high value of fractional power (99.9%) in the fundamental mode at the output of section I. For section II, the length of the taper is chosen as $L_2 = 2.91\,\mu\text{m}$, corresponding to the first maximum (99.6%) in Fig. 6(b).

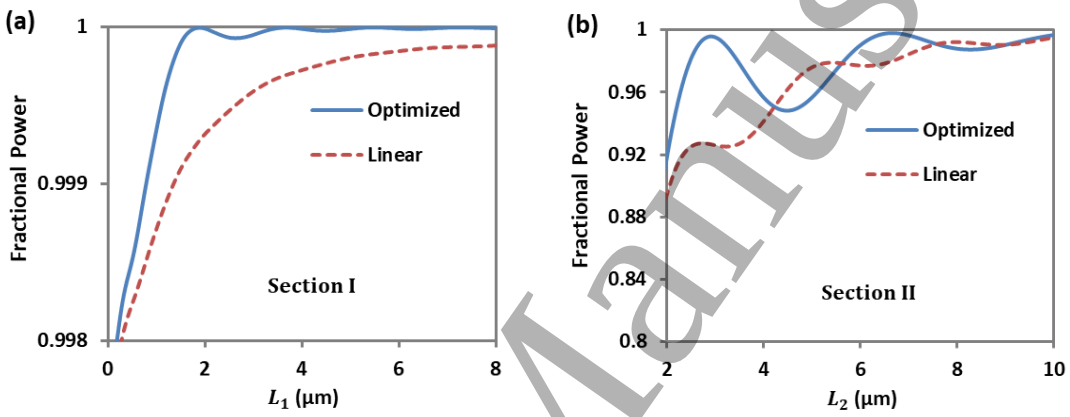


Fig. 6. Fractional output power for both optimized and linear tapers as a function of length of tapered section for individual (a) section I and (b) section II

Further, we look at the distribution of the adiabaticity parameter $c(z)$ along the length of the optimized tapers and compare it with the corresponding distribution for linear tapers of same length. The variation is shown in Fig. 7(a) and Fig. 7(b) with corresponding lengths $L_1 = 0.79\,\mu\text{m}$ and $L_2 = 2.91\,\mu\text{m}$. It can be clearly seen that the adiabaticity parameter $c(z)$ has a small value ε and is uniformly distributed over the entire length of the optimized tapers, as expected from the earlier discussion. A small noise in the $c(z)$ is due to the limited accuracy of the FEM solver. To maintain the value of $c(z)$ for the linear taper below the value obtained for the optimized taper, the slope (dD/dz) of the linear taper has to be reduced by an increase in the length of the linear taper by at least five times for section I and three times for section II.

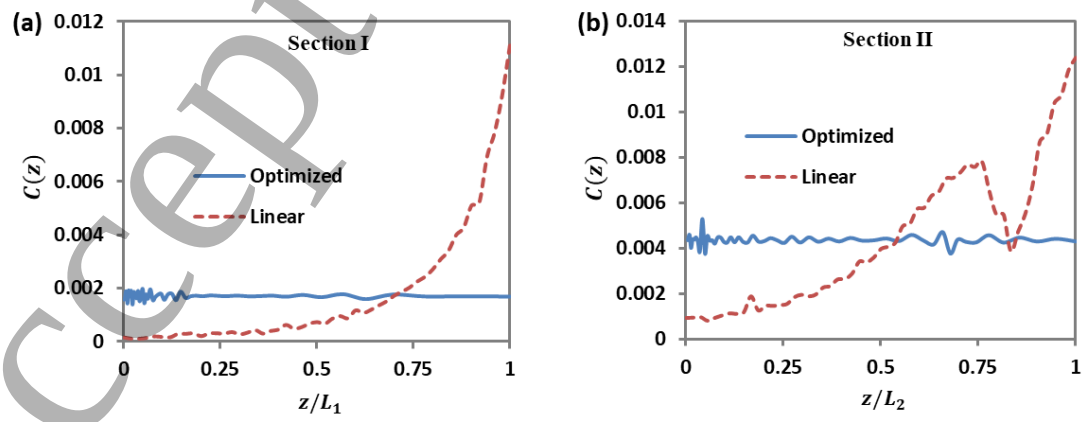


Fig. 7. (a) Distribution of parameter $c(z)$ throughout the length of taper for linear and optimized designs for (a) section I and (b) section II

Accepted Manuscript

The propagating field (E_x) through the proposed design, with $L_1 = 0.79\,\mu m$ and $L_2 = 2.91\,\mu m$, is shown in Fig. 8 for the input excitation by the fundamental strip mode at $\lambda = 1.55$ µm. It can be seen that the fundamental mode of the strip waveguide, excited at the input end, evolves adiabatically into the fundamental mode of the slot waveguide at the end of the converter. The conversion efficiency of the complete design is obtained as 99.4% at the design wavelength of 1.55 µm with a device footprint of only 1.23×3.7 µm^2.

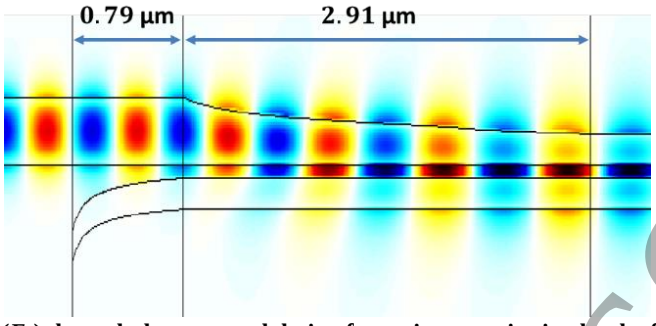


Fig. 8. Propagating field (E_x) through the proposed design for an input excitation by the fundamental mode of the strip waveguide

4. Wavelength response and fabrication tolerance

As the design utilizes the adiabatic mode evolution instead of conventional evanescent coupling, the design can be expected to be less sensitive to wavelength and fabrication variations. The mode conversion efficiency is plotted as a function of wavelength in Fig. 9. A variation of <1.5% in the conversion efficiency for a bandwidth of 100 nm and <5% for a bandwidth of 200 nm is observed which shows the broadband nature of the proposed design.

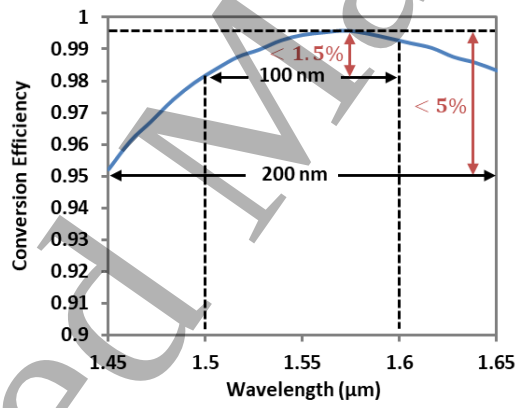


Fig. 9. Strip-to-slot waveguide fundamental mode conversion efficiency as a function of wavelength

The fabrication of the proposed design involves a single lithography and etching step which can be easily realized using the 193-nm immersion lithography followed by reactive ion etching [35]. The variation in the thickness of the device layer is entirely dependent on the SOI wafer. On the other hand, the width variation primarily depends on the lithography and etching process. Using 193-nm deep ultraviolet (DUV) lithography process, a linewidth accuracy of ±20 nm, and uniformity in the device layer thickness within ±10 nm can be easily achieved. The statistical results for a large-scale manufacturing shows the standard deviation in both the linewidth and thickness under 5 nm [36].

To study the fabrication tolerance of the proposed design, we consider a ±20 nm variation in the width and the thickness of the silicon device layer. These limits are more than sufficient to find the design tolerance of the proposed design to fabrication inaccuracies. First, we look at the conversion efficiency as a function of variation (ΔW) in width with $\Delta H = 0$ and then as a function of variation (ΔH) in height with $\Delta W = 0$ as shown in Fig. 10(a). Next, to calculate the worst-case conversion efficiency, we first vary the height and width by same magnitude simultaneously, as shown in Fig. 10(b), and then with different signs, as shown in Fig. 10(c). For the worst case, the conversion efficiency is found to be >96.5%. The results show the highly tolerant nature of the proposed strip

Accepted Manuscript

to slot waveguide mode converter to the possible fabrication inaccuracies.

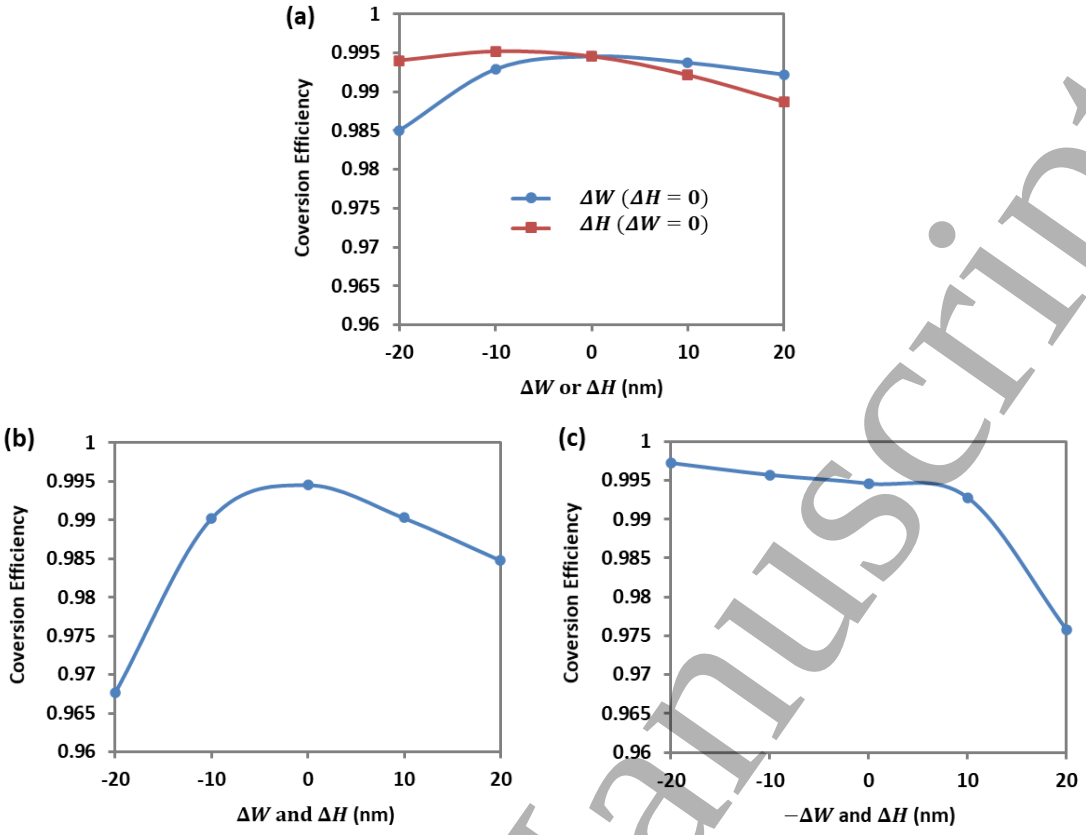


Fig. 10. Variation in the conversion efficiency as a function of deviation from the optimized parameters (a) deviation in width with optimized value of height and deviation in height with optimized value of width (b) deviation in width and height by same magnitude (c) deviation in width and height by same magnitude but with opposite signs

For the characterization of the proposed design, the grating couplers [37] need to be fabricated to allow in/out coupling from/to a single mode fibre (SMF). As the efficient designs of the grating couplers are not available for slot waveguide, an additional slot to strip converter needs to be fabricated as shown in Fig. 11(a) in order to allow coupling to the output SMF. As the loss associated with the proposed mode converter is very low, the inaccuracies in the measurement can arise due to a slight variation in the coupling efficiency of the grating couplers. Hence, to reduce the influence of such errors in the measurement, it is better to use a number of these sections in sequence [Fig. 11(b)] and to calculate the average response. To subtract the response of the grating couplers with respect to the fabricated design, a calibration structure consisting of two grating couplers connected through a strip waveguide is also need to be fabricated.

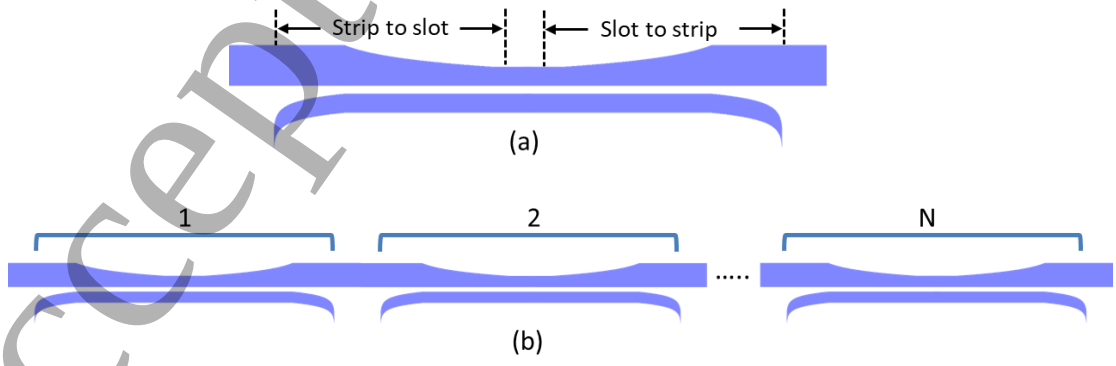


Fig. 11 (a) Mode converter pair consisting of a strip to slot converter and a slot to strip converter (b) Several cascaded pairs of mode converters

Accepted Manuscript

A tunable laser source, with operating wavelength range 1450 nm to 1650 nm, can be used as an input source. A polarization controller is required to control the polarization of the input light to allow efficient coupling from input SMF to the fundamental quasi-TE mode of the strip waveguide. To measure the wavelength response, the power can be coupled through the output grating coupler to an output SMF to observe the wavelength response using an optical spectrum analyzer.

The proposed design allows coupling between a standard strip waveguide and a symmetric slot waveguide, which confines the light within the narrow slot between the two silicon strips. However, there also exist other slot waveguide geometries including asymmetric slot waveguide [38], strip-loaded slot waveguide [1-3,39], plasmonic slot waveguide [40], and a slotted photonic crystal waveguide [41]. The mode converter for the asymmetric slot waveguide can be designed using two tapered sections by following a procedure similar to the one used for the symmetric slot waveguide. To facilitate the coupling between the strip waveguide and the other geometries of slot waveguides, the designs based on adiabatic mode evolution has been reported in the literature [39-41]. Each of these designs require certain number of tapers to facilitate the required mode coupling. The taper profiles in these designs can be optimized using the FAQUAD approach to allow efficient mode conversion with smaller device footprint.

5. Conclusion and discussion

Strip to slot waveguide mode converters are pivotal components in slot waveguide based integrated photonic devices [1-8,42]. The overall performance of these devices also depends on the performance metrics of the mode converters used to facilitate efficient fundamental mode coupling between the strip and slot waveguides. Hence, a design based on adiabatic mode evolution is proposed to obtain efficient fundamental mode coupling from a strip waveguide to a slot waveguide. A fast quasi-adiabatic approach is utilized to obtain optimized taper profiles which maintain a low constant value of adiabaticity parameter throughout the taper length. Using the optimized taper profiles, a very high (99.4%) conversion efficiency is obtained with small lengths of tapers, resulting in an ultra-compact footprint of only $1.23\times3.7\ \mu m^2$. In addition, the design is ultra-broadband with a variation of <1.5% in the conversion efficiency over a bandwidth of 100 nm and <5% over a bandwidth of 200 nm. The design also offers a very high tolerance to the fabrication inaccuracies.

Acknowledgement

This work was supported in part by the University Grant Commission, India, and in part by the University of Delhi South Campus, India.

References

1. Palmer, R., Koeber, S., Elder, D.L., Woessner, M., Heni, W., Korn, D., Lauermann, M., Bogaerts, W., Dalton, L., Freude, W. and Leuthold, J., 2014. High-speed, low drive-voltage silicon-organic hybrid modulator based on a binary-chromophore electro-optic material. *Journal of Lightwave Technology*, *32*(16), pp.2726-2734.
2. Koos, C., Jacome, L., Poulton, C., Leuthold, J. and Freude, W., 2007. Nonlinear silicon-on-insulator waveguides for all-optical signal processing. *Optics Express*, *15*(10), pp.5976-5990.
3. Koos, C., Vorreau, P., Vallaitis, T., Dumon, P., Bogaerts, W., Baets, R., Esembeson, B., Biaggio, I., Michinobu, T., Diederich, F. and Freude, W., 2009. All-optical high-speed signal processing with silicon–organic hybrid slot waveguides. *Nature photonics*, *3*(4), p.216.
4. Carlborg, C.F., Gylfason, K.B., Kaźmierczak, A., Dortu, F., Polo, M.B., Catala, A.M., Kresbach, G.M., Sohlström, H., Moh, T., Vivien, L. and Popplewell, J., 2010. A packaged optical slot-waveguide ring resonator sensor array for multiplex label-free assays in labs-on-chips. *Lab on a Chip*, *10*(3), pp.281-290.
5. Liu, Q., Tu, X., Kim, K.W., Kee, J.S., Shin, Y., Han, K., Yoon, Y.J., Lo, G.Q. and Park, M.K., 2013. Highly sensitive Mach–Zehnder interferometer biosensor based on silicon nitride slot waveguide. *Sensors and Actuators B: Chemical*, *188*, pp.681-688.
6. Luan, E., Shoman, H., Ratner, D.M., Cheung, K.C. and Chrostowski, L., 2018. Silicon photonic biosensors using label-free detection. *Sensors*, *18*(10), p.3519.
7. Butt, M.A., Khonina, S.N. and Kazanskiy, N.L., 2018. Silicon on silicon dioxide slot waveguide evanescent field gas absorption sensor. *Journal of Modern optics*, *65*(2), pp.174-178.
8. Dell'Olio, F. and Passaro, V.M., 2007. Optical sensing by optimized silicon slot waveguides. *Optics Express*, *15*(8), pp.4977-4993.
9. Milvich, J., Kohler, D., Freude, W. and Koos, C., 2018. Surface sensing with integrated optical waveguides: a design guideline. *Optics express*, *26*(16), pp.19885-19906.
10. Marchetti, R., Lacava, C., Carroll, L., Gradkowski, K. and Minzioni, P., 2019. Coupling strategies for silicon photonics integrated chips. *Photonics Research*, *7*(2), pp.201-239.
11. Wang, Z., Zhu, N., Tang, Y., Wosinski, L., Dai, D. and He, S., 2009. Ultracompact low-loss coupler between strip and slot waveguides. Optics letters, 34(10), pp.1498-1500.
12. Feng, N.N., Sun, R., Kimerling, L.C. and Michel, J., 2007. Lossless strip-to-slot waveguide transformer. *Optics letters*, *32*(10), pp.1250-1252.

Accepted Manuscript

13. Palmer, R., Alloatti, L., Korn, D., Heni, W., Schindler, P.C., Bolten, J., Karl, M., Waldow, M., Wahlbrink, T., Freude, W. and Koos, C., 2013. Low-loss silicon strip-to-slot mode converters. *IEEE Photonics Journal*, *5*(1), pp.2200409-2200409.
14. Mere, V., Kallega, R. and Selvaraja, S.K., 2018. Efficient and tunable strip-to-slot fundamental mode coupling. *Optics express*, *26*(1), pp.438-444.
15. Deng, Q., Liu, L., Li, X. and Zhou, Z., 2014. Strip-slot waveguide mode converter based on symmetric multimode interference. *Optics letters*, *39*(19), pp.5665-5668.
16. Xiao, J. and Xu, Y., 2016. Ultracompact and broadband silicon-based strip-to-slot mode converter. IEEE Photonics Technology Letters, 28(13), pp.1414-1417.
17. Deng, Q., Yan, Q., Liu, L., Li, X., Michel, J. and Zhou, Z., 2016. Robust polarization-insensitive strip-slot waveguide mode converter based on symmetric multimode interference. *Optics express*, *24*(7), pp.7347-7355.
18. Martínez-Garaot, S., Muga, J.G. and Tseng, S.Y., 2017. Shortcuts to adiabaticity in optical waveguides using fast quasiadiabatic dynamics. *Optics express*, *25*(1), pp.159-167.
19. Chung, H.C., Martínez-Garaot, S., Chen, X., Muga, J.G. and Tseng, S.Y., 2019. Shortcuts to adiabaticity in optical waveguides. *EPL (Europhysics Letters)*, *127*(3), p.34001.
20. Cytop technical brochure, AGC Inc., Tokyo, 2018, https://www.agcchem.com/documentation/technical-data/cytop/
21. Wong, W.R., Fan, H., Adikan, F.R.M. and Berini, P., 2018. Multichannel long-range surface plasmon waveguides for parallel biosensing. Journal of Lightwave Technology, 36(23), pp.5536-5546.
22. Leosson, K. and Agnarsson, B., 2012. Integrated biophotonics with CYTOP. *Micromachines*, *3*(1), pp.114-125.
23. Krupin, O., Wong, W.R., Béland, P., Adikan, F.R.M. and Berini, P., 2016. Long-range surface plasmon-polariton waveguide biosensors for disease detection. Journal of Lightwave Technology, 34(20), pp.4673-4681.
24. Ren, G., Chen, S., Cheng, Y. and Zhai, Y., 2011. Study on inverse taper based mode transformer for low loss coupling between silicon wire waveguide and lensed fiber. *Optics Communications*, *284*(19), pp.4782-4788.
25. Khoo, E.H., Liu, A.Q. and Wu, J.H., 2005. Nonuniform photonic crystal taper for high-efficiency mode coupling. Optics express, 13(20), pp.7748-7759.
26. Sun, X., Liu, H.C. and Yariv, A., 2009. Adiabaticity criterion and the shortest adiabatic mode transformer in a coupled-waveguide system. *Optics letters*, *34*(3), pp.280-282.
27. Horth, A., Cheben, P., Schmid, J.H., Kashyap, R. and Quitoriano, N.J., 2016. Ideal, constant-loss nanophotonic mode converter using a Lagrangian approach. *Optics express*, *24*(6), pp.6680-6688.
28. Schiff, L.I., 1998. Quantum Mechanics, 3rd ed. (McGraw- Hill).
29. Martínez-Garaot, S., Ruschhaupt, A., Gillet, J., Busch, T. and Muga, J.G., 2015. Fast quasiadiabatic dynamics. *Physical Review A*, *92*(4), p.043406.
30. Chung, H.C., Lee, K.S. and Tseng, S.Y., 2017. Short and broadband silicon asymmetric Y-junction two-mode (de) multiplexer using fast quasiadiabatic dynamics. *Optics express*, *25*(12), pp.13626-13634.
31. Chung, H.C. and Tseng, S.Y., 2018. Ultrashort and broadband silicon polarization splitter-rotator using fast quasiadiabatic dynamics. *Optics express*, *26*(8), pp.9655-9665.
32. Hung, Y.J., Li, Z.Y., Chung, H.C., Liang, F.C., Jung, M.Y., Yen, T.H. and Tseng, S.Y., 2019. Mode-evolution-based silicon-on-insulator 3 dB coupler using fast quasiadiabatic dynamics. *Optics letters*, *44*(4), pp.815-818.
33. FIMMWAVE/FIMMPROP, Photon Design Ltd, http://www.photond.com.
34. Gallagher, D.F. and Felici, T.P., 2003, June. Eigenmode expansion methods for simulation of optical propagation in photonics: pros and cons. In *Integrated optics: devices, materials, and technologies VII* (Vol. 4987, pp. 69-82). International Society for Optics and Photonics.
35. Selvaraja, S.K., Winroth, G., Locorotondo, S., Murdoch, G., Milenin, A., Delvaux, C., Ong, P., Pathak, S., Xie, W., Sterckx, G. and Lepage, G., 2014, April. 193 nm immersion lithography for high-performance silicon photonic circuits. In Optical Microlithography XXVII (Vol. 9052, p. 90520F). International Society for Optics and Photonics.
36. Lu, Z., Jhoja, J., Klein, J., Wang, X., Liu, A., Flueckiger, J., Pond, J. and Chrostowski, L., 2017. Performance prediction for silicon photonics integrated circuits with layout-dependent correlated manufacturing variability. *Optics express*, *25*(9), pp.9712-9733.
37. Marchetti, R., Lacava, C., Carroll, L., Gradkowski, K. and Minzioni, P., 2019. Coupling strategies for silicon photonics integrated chips. *Photonics Research*, 7(2), pp.201-239.
38. Spott, A., Baehr-Jones, T., Ding, R., Liu, Y., Bojko, R., O'Malley, T., Pomerene, A., Hill, C., Reinhardt, W. and Hochberg, M., 2011. Photolithographically fabricated low-loss asymmetric silicon slot waveguides. *Optics express*, 19(11), pp.10950-10958.
39. Liu, Y., Baehr-Jones, T., Li, J., Pomerene, A. and Hochberg, M., 2011. Efficient strip to strip-loaded slot mode converter in silicon-on-insulator. *IEEE Photonics Technology Letters*, 23(20), pp.1496-1498.
40. Chen, C.T., Xu, X., Hosseini, A., Pan, Z., Subbaraman, H., Zhang, X. and Chen, R.T., 2015. Design of highly efficient hybrid Si-Au taper for dielectric strip waveguide to plasmonic slot waveguide mode converter. *Journal of Lightwave Technology*, 33(2), pp.535-540.
41. Zhang, X., Subbaraman, H., Hosseini, A. and Chen, R.T., 2014. Highly efficient mode converter for coupling light into wide slot photonic crystal waveguide. *Optics express*, 22(17), pp.20678-20690.
42. Subbaraman, H., Xu, X., Hosseini, A., Zhang, X., Zhang, Y., Kwong, D. and Chen, R.T., 2015. Recent advances in silicon-based passive and active optical interconnects. *Optics express*, 23(3), pp.2487-2511.

Accepted Manuscript

Journal Pre-proof



Energy efficient transmission trends towards future green cognitive radio networks (5G): Progress, taxonomy and open challenges

Akanksha Srivastava, Mani Shekhar Gupta, Gurjit Kaur

PII: S1084-8045(20)30234-4

DOI: https://doi.org/10.1016/j.jnca.2020.102760

Reference: YJNCA 102760

To appear in: *Journal of Network and Computer Applications*

Received Date: 7 January 2020

Revised Date: 21 April 2020

Accepted Date: 23 June 2020

Please cite this article as: Srivastava, A., Gupta, M.S., Kaur, G., Energy efficient transmission trends towards future green cognitive radio networks (5G): Progress, taxonomy and open challenges, *Journal of Network and Computer Applications* (2020), doi: https://doi.org/10.1016/j.jnca.2020.102760.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier Ltd.

Energy Efficient Transmission Trends Towards Future Green Cognitive Radio Networks (5G): Progress, Taxonomy and Open Challenges

Akanksha Srivastava, Mani Shekhar Gupta and Gurjit Kaur

***Abstract*—With the growing cognizance about environmental concern and global warming-related to communication technologies, the researchers have been seeking some solutions to diminish the consumption of energy in the telecommunication industry. There is a remarkable advancement in mobile communication from simple voice-based devices to ubiquitous data-hungry smartphones. Nearly, 8 billion mobile subscribers are expected to be added by 2030 and also required an extra spectrum to serve. The existing static spectrum allocation-based technologies are not in a position to fulfill this extra spectrum requirement and handle this future traffic load. This volatile evolution of global traffic data urges research attention globally and can be handled by future cognitive radio networks. There is also a demand for providing fast speed and seamless services for which operators need to deploy more base stations continuously and increase the transmitting antenna power. As a result, a drastic hike in emissions of carbon dioxide into the environment, and exposure to harmful radiations in large amounts which is also authentically harmful to humans, animals, and birds. This increased energy consumption and scarcity of resources have made the energy as a geopolitical issue for 21 century. In this direction, this work contributes by introducing cognitive-based green communication technology to ensure the environmental and health concerns caused due to hike in the CO_2 level. In the second phase, an enlightening survey on futuristic approaches for making the future wireless networks green is covered with pros and cons. In the third phase, this paper provides a framework of research challenges with ongoing project activities on green communication which further requires attention from the research community.**

***Index Terms*—Green communication, Carbon dioxide emissions, Carrier aggregation, Massive MIMO, D2D communications, Green IoT, Cognitive radio network.**

I. Introduction

In the last few years, the mobile telecommunication industry has increased exponentially. The call for wireless applications is growing because of the astronomical boom of mobile subscribers, multimedia applications, and data applications. In this work, the word green cognitive radio (CR) communication is used to express an efficient method of the congregation of energy-efficient technologies at different levels to reduce the unfavorable effects of technology on the environment. Globally, by the year 2022, the internet user population will be increased from 45.3% to 59.7% [1]. In addition to this, Fig. 1 depicts the growth rate of different parameters from 2018 to 2025 [2]. The nuts and bolts of a telecommunication network are a mobile node, access network and a backbone core network. Base Stations (BSs) are the key component of the radio access network that connects mobile phone devices to a core network. So, with a growth rate in the number of mobile subscribers per day, the growth rate of BSs is also increasing. According to [3] it is expected that during period $2017-2021$, the BS market of Long Term Evaluation (LTE) system will increase by more than 17 percent compound annual growth rate (CAGR). Available literature work centering on improving both quality of service (QoS) and quality of experience (QoE), while ignoring the increased demand for energy for the mobile communication system. This increasing energy demand has motivated us to work on the subject of cognitive-based green communication with the objective of energy-efficient wireless communication. Based on our studies, we found two major goals of the green communications network. The first is to reduce the CO_2 emission from the system and second is to save energy by reducing transmitting power which will lead to reducing the operational expenditures (OPEX). The off-grid sites (to serve in remote areas) are the main cause of CO_2 emission. Mostly diesel-power generators are required for working on such sites. Presently, as the number of mobile subscribers is increasing, nearly 43% of the total CO_2 emission is contributed by the mobile communication system and this value will increase by 51% of the total in 2020 [4]. Due to the different segments of the mobile communication system, the CO_2 level is continuously increasing in the atmosphere. Fig. 2 represents the level of CO_2 emissions by a different segment of the communication system during 2010, 2015 and in 2020. The measure of data traffic related to a mobile node is expected to touch nearly 621.90 million Tbytes/year by 2025 while in the year 2012, it was 45 million Tbytes/year [5]. The electrical energy consumption by the ICT will be increased from 611 to $1,752$ Tera-watt hours by 2030 [6]. This rapid energy consumption rate by wireless communication networks will cause a serious problem in the future if no steps are taken to change the existing trend.

A. Motivation for Green Communication

The telecommunication industry expressions a serious energy consumption challenge. This industry is growing rapidly to serve the smartphone user anywhere, anytime and anyone. The simple reason is that every human is using dual a SIM card enabled phone. At the end of 2019, in the Asia Pacific there were 2.9 billion (nearly the world's half population) mobile subscribers [7]. India and China are the most populous

Journal Pre-proof

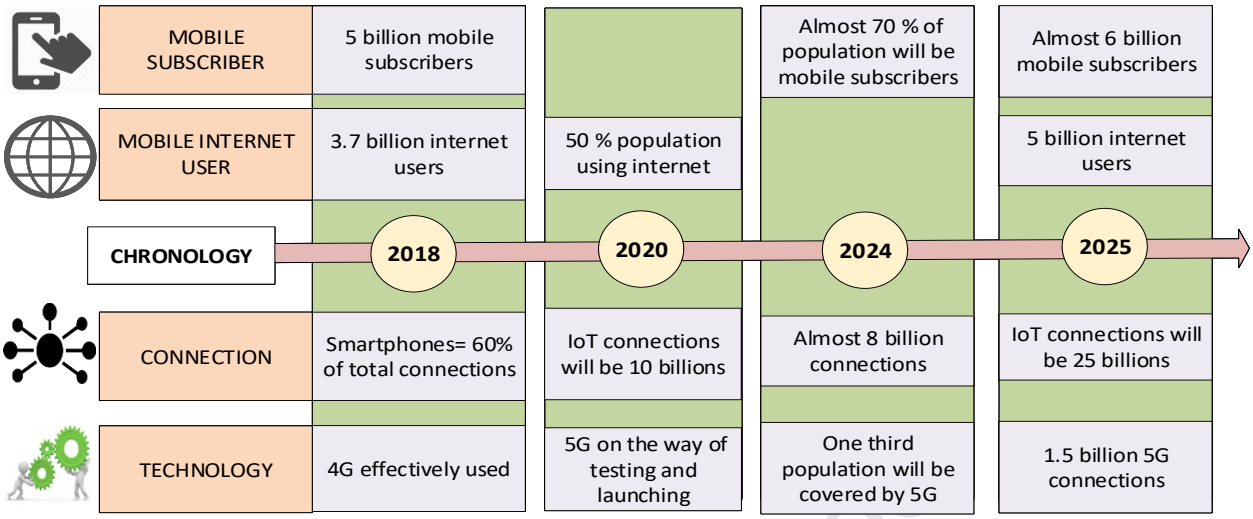


Fig. 1: Breakthrough in the field of telecommunication industry

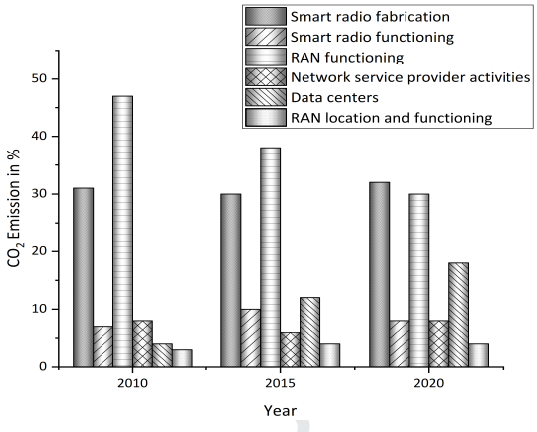


Fig. 2: Variation in carbon footprint produced by various parts of mobile communication system

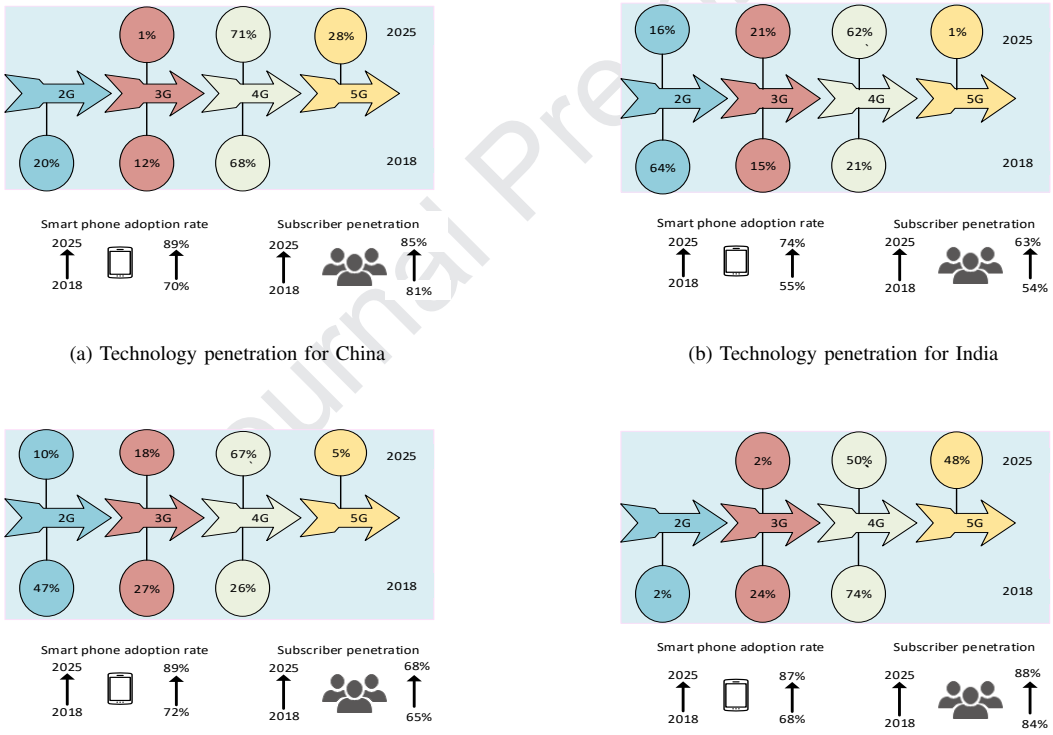
countries, which jointly covers nearly two-thirds of total mobile users. Nearly, 71.9 million mobile users are expected to be added in China by 2025, which is nearly 1.2 billion (85%) of the total population [8]. On the other hand, in India, it is expected to be nearly 207.9 million (22% of total Asia Pacific) by 2025. So, this makes a remarkable improvement in digital literacy and advanced mobile phones. In 2018, Indonesia was developed as a third-largest mobile phone market in the world (behind India and China) [9]. The smartphone adoption rate is to be expected very high (nearly 90% of total connections) for Indonesia by the end of 2025. Fig. 3 depicts the technology penetration of the four fastest-growing countries. Fig shows the percentage of the population which was using the different wireless generation (2G, 3G, 4G and 5G) in 2018 and how much this percentage value will be changed when we will reach 2025. In China and Japan, the 2G population will completely remove till 2025, on the other hand, some populations will present in India and Indonesia. In addition to this, Fig. 4 represents the addition of new mobile users globally till June 2019 in million [10] and Fig. 5 shows the mobile broadband adoption rate for different countries from 2015 to 2020.

The wireless communication networks have experienced a prodigious evolution from 1G to 5G networks. Nowadays 4G is effectively in use and 5G on the way of testing and launching. Recently, during the testing of the 5G network, 100s of birds dropped from the sky at Hague in the Netherlands [11]. The 5G antenna was situated at the Dutch railway station, to examine how large the 5G range was. At once, birds fell dead from the sky and the ducks that were swimming in nearby ponds seemed to respond very strangely. To escape from radiation, they were instantaneously inserting their heads inside water whereas some flew away. If they all had a healthy body, healthy blood, no effect of any virus and bacterial infection then the reasonable explanation behind this incident is using of microwaves in 5G having a serious impact on birds' hearts. It highly resonates with erratic pulsed microwaves (millions per second) which harm body organs [12]. Table II represents the upgradation of the various generation of wireless networks. This table reflects different characteristics, to discuss a tangible contrast among the generations of wireless networks. In 5G a drastic growth in the CO_2 emissions is presented.

To overcome this problem a pathway is designed in Fig. 6 which represents greenhouse gas (GHG) deductions across different sectors to reach 2030. Fig. shows that the amount of CO_2 was present almost 440 MMT in 2015 and the target is to reduce this value to 261 MMT till 2030. To curb such enormous emissions, various technologies have been discussed in this paper. This paper focuses on the mobile sector and provides comprehensive data on changing trends of information and communication technology (ICT) on the

TABLE I: Table of abbreviations

Abbreviation	Description	Abbreviation	Description
ASIC	Application-specific integrated circuit	AMPS	Advanced Mobile Phone Service
ATIS	Alliance for Telecommunications Industry Solutions	BDMA	Beam Division Multiple Access
CPLD	Complex Programmable Logic Device	CA	Carrier Aggregation
CAGR	Compound Annual Growth Rate	CRN	Cognitive Radio Network
D2D	Device-to-Device	EMR	Electromagnetic Radiation
ETSI	European Technical Standards Institute	FBMC	Filter Bank Multi Carrie
FPGA	Field Programmable Gate Array	GHG	Greenhouse Gas
IARC	International Agency for Research on Cancer	ITU	International Telecommunication Union
LTE-A	Long Term Evolution Advanced	MIMO	Multiple-Input Multiple-Output
OFDM	Orthogonal Frequency Division Multiplexing	PA	Power Amplifier
OPEX	Operational Expenditures	SAR	Specific Absorption Rate
SIM	Subscriber Identification Module	SM	Spatial Multiplexing
UDN	Ultra Dense Network	UMTS	Universal Mobile Telecommunications System
V2X	Vehicle to Everything	WHO	World Health Organization



(a) Technology penetration for China

(b) Technology penetration for India

(c) Technology penetration for Indonesia

(d) Technology penetration for Japan

Fig. 3: Technology penetration of four fastest growing countries in Asia Pacific from 2018 to 2025.

global platform. Emission of CO_2 is required to be declined.

B. Key contribution of survey

This work coveres, an up-to-date survey and technical aspects of advances which is providing future research directions to the foremost study on energy-efficient green communication in cognitive radio networks (CRNs). The key contributions of this work can be outlined as follows:

- A complete survey on green wireless communication networks have been covered in a single paper, with consideration of energy-efficient transmission.
- The futuristic technologies like cognitive radio, carrier aggregation, Terahertz communication, Internet of Things (IoT), massive MIMO (multiple-input multiple-output) and mm wavelength are briefly reviewed to prepare for advancing recent research contributions. These future

TABLE II: Upgradation and contrast among various generations of wireless networks in green communication aspect

Technology	1G	2G	3G	4G	5G
Frequency band	800 Mhz	850/900/1800/1900 Mhz	1900 /2100 Mhz	1.8/2.6 Ghz	30-300 Ghz
Power density ($Watt/meter^2$)	4	4.5-9	4.5-10	10	10
Power level of MS	Low	GSM 850/900 :35 dBm GSM 1800/1900 :22-37 dBm	20-32 dBm	23 dBm	High
Power level of BS	Low	Macro 46 dBm, Micro 13-30 dBm	32-38 dBm	44-49 dBm	High
CO_2 emission	23 Mt CO_2	30 Mt CO_2	85 Mt CO_2	171 Mt CO_2	236 Mt CO_2
Carbon footprint per mobile	10kg	15kg	20kg	24kg	30kg
Power optimization methods	Modulation techniques such as SSB-FM used to save power.	Moved towards digital communication, Sleep mode at BS and small powered radio signals used.	Sleep mode at BS and relay technique is used.	In OFDMA and SC-FDMA, PAPR minimization used small cells concept used.	D2D, UDN, C-RAN, IoT used.
Date rate	2.4 Kbps	100 Kbps	2.4-30 Mbps	100-200 Mbps-1.5-3 Gbps	10-50 Gbps
Access technology	AMPS	GSM-GPRS	WCDMA, UMTS, CDMA 2000, HSPDA	LTE-A, WiMAX	BDMA, FBMC
Spectral efficiency (bps/Hz)	0.46	1.4	2.7	4.27	10.6
MIMO/Massive MIMO	SISO	SISO	SIMO, MISO	MIMO	Massive MIMO
WLAN Technology	802.11	802.11b	802.11ag, 802.11n	802.11ac, 802.11an	802.11ad
BS density per square kilometer	Very low	Low	4-5 BSs	8-10 BSs	40-50 BSs
OPEX	-	-	Low	High	High
SAR value	-	-	High	Higher	Expected to reduce
Used antenna	Low frequency antenna	High frequency antenna	High gain and clip antenna	Slot and Patch antenna	Phased array antennas

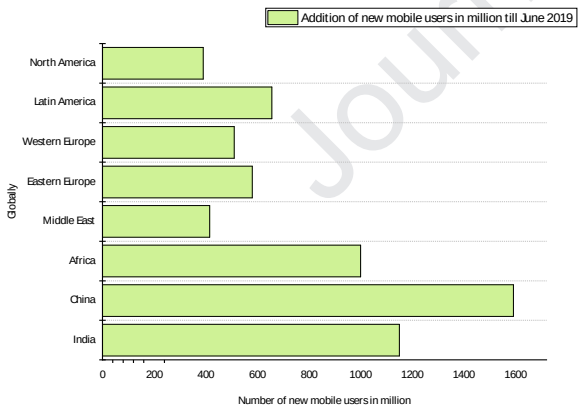


Fig. 4: Addition of new mobile users globally in million till June 2019

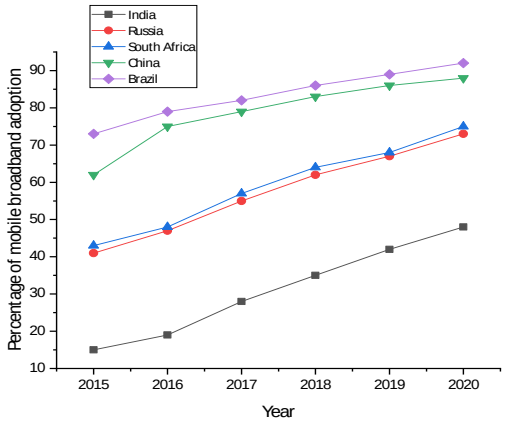


Fig. 5: Mobile broadband acquisition rate

technologies for 5G are proficient for improving the energy efficiency of the next generation wireless networks with high QoS and low GHG emissions.

- Starting with the impact of the increasing number of mobile users globally, their demands for different services and their adverse effects on the environment are discussed in detail.
- Several energy efficiency metrics used for the calculation of different parameters at different levels of communication system have been illustrated.
- Various future research directions and challenges in the pathway of green network are also covered to handle the implementation challenges.

C. Paper Organization

The progression of the work is maintained as in Fig. 7. In Section II, the recent facts and figures are briefly discussed

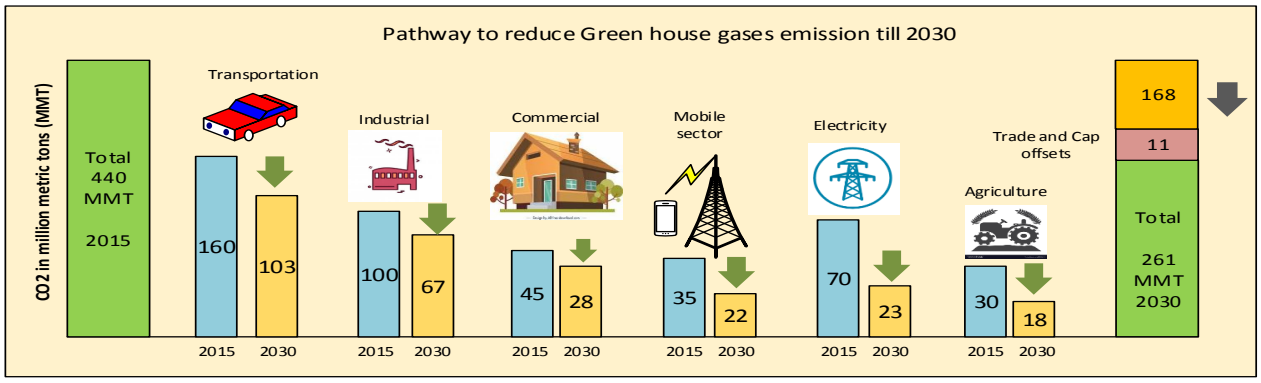


Fig. 6: Aim to reduce GHG across various sectors to reach 2030

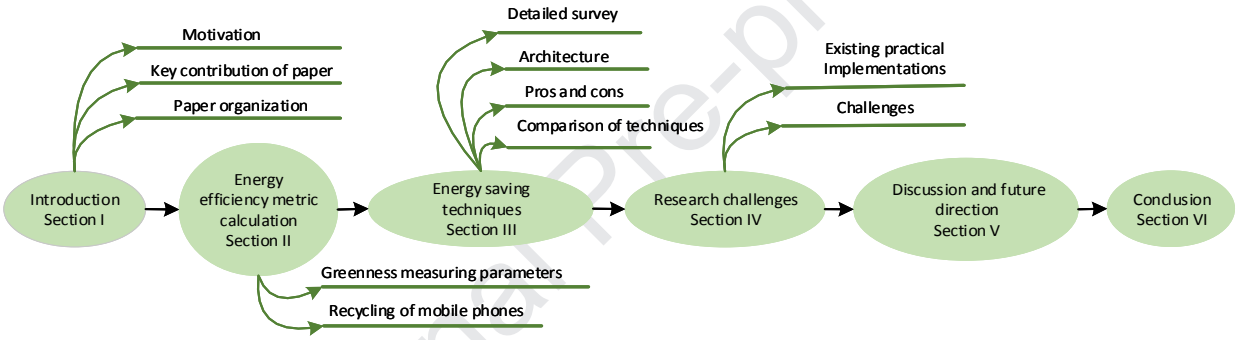


Fig. 7: Paper organization

with energy efficiency metrics for greenness calculation. The main purpose of this section is to familiarize with the vital facts and the expressive concern of an energy aware green CRN research which is the subject of attention. In Section III, various energy aware techniques have been discussed with their concept, detailed survey, merits and demerits and also discussed how they are capable of supporting green networking for next generation (5G) networks. A comparison is performed to highlight the role of each techniques to make communication energy-efficient. Various research challenges associated with green communication have been discussed in Section IV. Future research directions have been explored in Section V. Finally, conclusion is presented in Section VI.

II. Energy Efficiency Metrics for Greenness Calculation

A. Energy Efficiency Measurement (EEM)

As we know, the cellular system is the largest component of ICT sector. So, EEM statistics have encouraged both academia and industry researchers to evolve methods to cut down the energy expenditure of cellular system and make the wireless network system greener. The primary goals of energy-aware green CRNs are:

- Improving EE.
- Improvement in the tradeoff between energy-saving and seamless connectivity.
- Reducing unnecessary GHG emission.
- Proper utilization of resources.

Fig. 8 represents that in a wireless communication network BS consumes the maximum amount of energy as compared to the other component of the network [13]. The number of BSs are continuously increasing to provide better QoE expected by mobile users. Thus, energy consumption will also increase proportionally. Nowadays this is a very crucial subject for research to find out some effective methods to minimize energy consumption by fixed and dynamic components of BS while maintaining the QoS. The internal components like digital signal processors, power amplifier (PA), feeder cables and cooling system of BS are considered under the fixed components while network planning, deployment and management based on proper load balancing are under the dynamic components. For the conversion of a base station into a green base station, several schemes have been suggested in

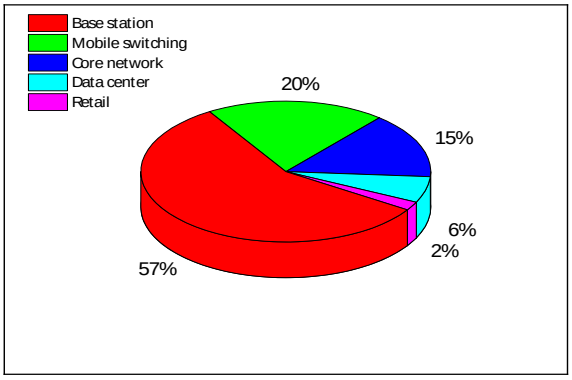


Fig. 8: Energy consumption w.r.t. different components of communication system

[14]. A game theory-based scheme considers heterogeneous networks for the reduction of power consumption at the base station level is discussed in [15]. A novel network architecture having energy-efficient techniques based on the physical layer and cross-layer of cellular networks is represented in [16]. Different research activities and projects based on energy-efficient communication has been considered in [17].

The various components in wireless networks consume different amounts of energy. So, for calculating and comparing the amount of energy which is consumed by the various components and a communication system as a whole, EE metrics are used. Researchers have been proposed different levels (facility, component, and system) for measuring EE metrics [18]. The high-level systems such as data centers are considered under the facility level [19]. Presently, with the upsurge in research activities on green energy efficient wireless networks, some standard organizations like European Technical Standards Institute (ETSI) and Alliance for Telecommunications Industry Solutions (ATIS) are making efforts to describe diverse EE metrics for wireless networks. The EE metrics consist of three major metrics as follows:

1) Energy Efficiency: EE is a very essential metric to measure in green wireless CRN [20]. This can be obtained by dividing network output energy to network input energy consumption [21] and expressed as

$$\eta = \frac{\text{Network output energy}}{\text{Network input energy consumption}} \tag{1}$$

here η represent EE. The best EE can be achieved by minimizing total power consumption and maximizing the data rate.

2) Secrecy Energy Efficiency: This is another important metric consider under the system level. The high value of this metric provides better security in the CRN [22] [23]. This is given by

$$S_{EE} = \frac{Secrecy\ rate}{Power\ consumption\ at\ \text{network}} \tag{2}$$

here secrecy rate (SR) is defined as the difference of data transmission rate of the transmitter-receiver link to the transmitter-eavesdropper link and S_{EE} represents secrecy energy efficiency.

3) Area Energy Efficiency: The A_{EE} metric is considering for a dense deployment [15]. This can be obtained by dividing overall energy efficiency to the macro cell area and given by

$$A_{EE} = \frac{Energy\ efficiency}{Total\ area\ of\ macrocell} \tag{3}$$

here A_{EE} is area energy efficiency.

The readers can discover different taxonomy of green communication metrics which are summarized in Table III.

B. Recycling of Mobile Phones

Usage of recyclable stuff for manufacturing of next generation devices can help to make the environment eco-friendly. For example cell phones are manufactured by some scanty natural resources likes plastic, copper, etc and also contain some non-biodegradable materials. This can surge the GHG effect if it did not correctly dispense when the cell phones are in a dead state. In Australia, almost 24 million cell phones are existing in cupboards and drawers which are not currently used [35]. Presently, the whole world is fighting against vast energy consumption and GHG effect. So, the trends for recycling cell phones are ever-increasing because 90% material of cell phones is recyclable. Fig. 9 represents the cell phone recycling process with a percentage of CO_2 emitted at each stage.

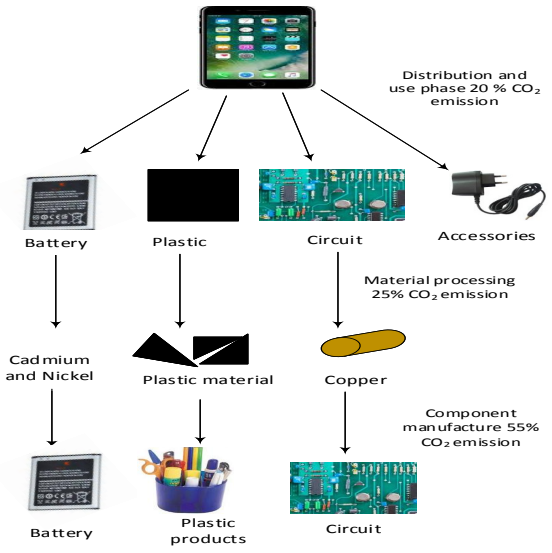


Fig. 9: Cell phone recycling process with CO_2 emission

III. Energy Efficient Techniques

In this section, several traditional and recent advanced energy efficient communication approaches for future CRNs (5G) are considered with their detailed overview in an informational manner. Various energy-efficient communication approaches are highlighted in a classification tree as shown in Fig. 10. These different approaches are introduced by researches to handle various inspiring topics allied with green CRNs.

TABLE III: Summary for several EE metrics evaluation in green communication networks

Level	Metric	Remark	Unit	Reference
Facility level	Power usage efficiency	It gives the ratio of total power consumption at facility level to the total power consumption at equipment level.	PUE ≥1	[24], [25]
Facility level	Data center efficiency	Reciprocal of PUE	DCE in percentage	[24]
Element level	Telecommunications energy efficiency ratio (TEER)	Ratio of maximum throughput to total power consumption.	$Gbps/Watt$	[26]
Element level	Telecommunication equipment energy efficiency rating (TEEER)	Calculated by summing the amount of energy consumed by the equipment for various traffic conditions.	$-\log(\frac{G_{bps}}{Watt})$	[27]
Element level	Energy consumption rating (ECR)	Ratio of energy consumed to the adequate system capacity.	$Watt/Gbps$	[18]
Element level	ECR-weighted (ECRW)	Evaluated in the similar manner as ECR but energy consumption is in ECRW is calculated as $0.25E_i$+$0.4E_h$+$0.35E_f$ here E_i, E_h and E_f represents Energy consumption in idle mode, half load and full load.	$Watt/Gbps$	[18]
Access point level	Energy consumption rate	Energy used to transmit a piece of information.	J/bit	[28] [15]
Access point level	Spectral efficiency	For a communication system, the transmitted information rate to a given frequency	$b/s/Hz$	[29] [30]
Access point level	Power efficiency	Ratio of output power to input power	-	[18]
Access point level	Energy consumption index	Ratio of (BS/AP) input power to the key performance indicator. ECI evaluate the power utilization efficiency of a BS. ECI inversely proportional to EE.	-	[31] [32]
System level	Power per bit per area	The average power used in terms of the average transmission rate and the coverage area.	$W/bps/m^2$	[33] [32]
System level	Energy per bit per unit area	The energy consumed for transferring the bits for a particular coverage area.	$J/bit/m^2$	[25]
System level	The power consumed per unit area	Ratio of power consumed to unit area	W/km^2	[34]
System level	Performance Indicator in urban areas (PI urban)	Ratio of total number of mobile users to the total power consumed by the base station/ access point.	$Users/Watt$	[34] [18]
System level	Performance Indicator in rural areas (PI rural)	Ratio of coverage area to the total power consumption by the base station/ access point.	$km^2/Watt$	[34]
System level	Energy reduction gain	Defined as the derivative of energy consumption gain metric.	ERG in percentage	[26]

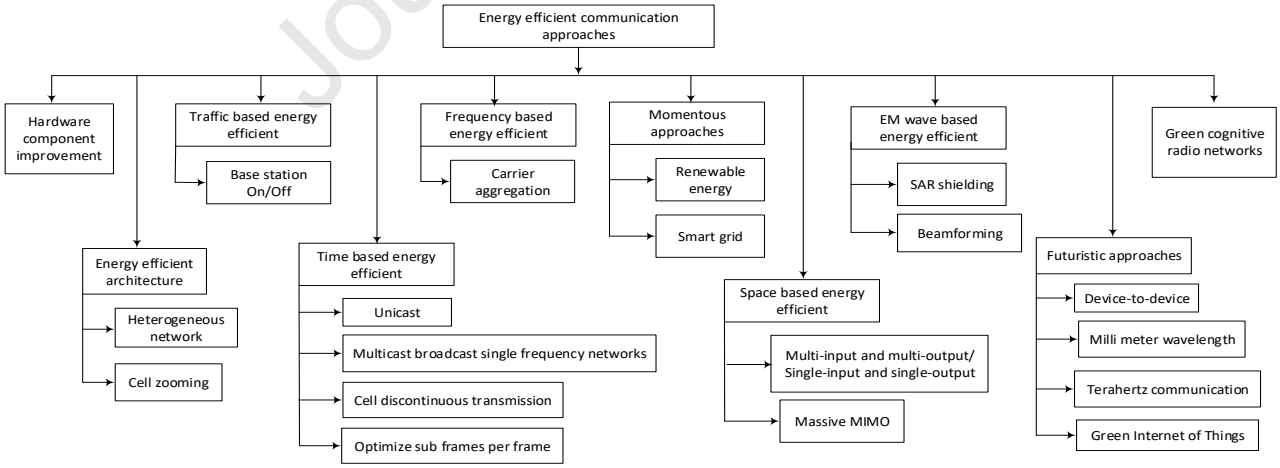


Fig. 10: A classification tree of different energy efficient communication approaches

A. *Hardware Component Improvement*

In the telecommunication system, the PA consumes maximum amount of energy (nearly 60%) in the BS. Therefore, for OPEX cost reduction a highly proficient PA is necessary for green CRNs. In [41], various techniques for improving EE related to PAs and cables used in the feeder and cooling components are discussed. An operating point of the most efficient PA is adjacent to the saturation region. In [42], the author has

TABLE IV: Approaches to enhance the EE of power amplifiers

Approaches	Description	Energy saved	References
Doherty designs	Unique design method which is able to offer significant enhancements in EE.	30 to 35%	[18], [36]
Switched mode power amplifier (SMPA)	SMPA works on the principle that the active device works in saturation, therefore it works as a switch, so either current or voltage waveforms are alternatively reduced to decrease overlapping. This leads to minimizing power dissipation. If a transistor is considered as an ideal switch, 100% of efficiency can be attained by appropriate designing of the network.	80-90%	[19]
Enhancement in digital pre-distorted Doherty architecture and gallium nitride (GaN) amplifiers	GaN is a special material that is characterized by large breakdown voltage, high current density, and high frequency. Excellent power dissipation and thermal conductivity are used in the construction of power amplifier transistors, which enhance the efficiency of the system.	50%	[37]
Inverse Class F	Considered for WiMAX network having carrier frequency approximately 2.45 GHz.	74%	[38]
Envelope tracking designs	To track the input signal magnitude envelop this technique uses amplifier's drain bias.	Up to 60%	[37]
Class-AB with digital pre-distortion	Class AB with digital pre-distortion amplifiers, display gain roll-off at small input power and at saturation this property helps reduce energy.	50%	[39]
Light Radio Cubes	Alcatel-Lucent has been designed as an innovative BSs in 2012 which reduces a large amount of energy consumption as compared to the existing radio access network devices.	50%	[40]
Class J Amplifier	Due to the presence of nonlinear capacitance, the Class-J amplifier represents high efficiency. This nonlinear capacitance generates a harmonic voltage which reduces the phase difference between current components and fundamental voltage.	70 to 90%	[19]

discussed that PA operates in the linear region due to non-constant envelope signals of orthogonal frequency-division multiplexing (OFDM) modulation and nonlinear effects ranged from 6 to 12 dB below saturation region. For designing of PA, the following factors are having major importance:

- Wider bandwidth for communication.
- Follow linearity property, to provide higher-order modulation technique.
- Improved BS energy efficiency to achieve OPEX.
- Higher average output power levels.

The higher value of DC power supply is required for the operation of PAs. During a day, the traffic load is not constant like in-office hours traffic load is high and in the night it is low. So, a large amount of power is wasted at night because the working of PAs does not depend on traffic load. To overcome this challenge two techniques can be used. In first, the PAs can be deactivated in some time slots with no data transmission. In second, an operating point of PA can adjusted to reduce devouring of power and also optimize the power efficiency for different traffic load scenarios (low, medium and high). The major advantage of this technique is that it saves the maximum amount of energy but high replacement cost is its limitation. Table IV summarizes a survey of various approaches that have been used to improve the hardware component of the communication system and also represents how much amount of energy can be saved by these approaches.

Instead of several endeavors, an improvement in hardware approach for saving energy at BS is not in a position to achieve the desired result. Ineffective utilization of resources also results in a large amount of energy wasted. So, for designing energy-efficient green CRNs, exploitation of both network-level and equipment level approaches are required. In the succeeding subsection of the paper different approaches are discussed to improve EE.

B. Energy-Efficient Architecture

Energy-efficient architecture considers an efficient approach for both energy efficiency improvement and energy saving. This approach is based on the cooperation among BSs to offer the best QoS to mobile subscribers and save energy at low traffic. In this subsection, two techniques such as heterogeneous networks (HetNets) and cell zooming are discussed.

1) Heterogeneous Networks (HetNets): HetNets represents a layered architecture that has different networks for providing the service to the same mobile subscriber. Considering the small size cell (like femto, pico, and microcells) to reduce consumption of energy and enhance the throughput of the network. These cells differ in data rate, size, capacity, output power, coverage and also emit a different amount of CO_2 as indicated in Fig. 11 (the size of the cloud). As the diagram represents that macrocells are having the largest coverage area and least CO_2 emission while femtocell is having less area but emits the maximum amount CO_2. The principle of the HetNet approach is the deployment of macrocells for providing complete network coverage, while small cells will come in an operational state when the traffic demand increases in that area. In [48], [49] author considers both scenarios heterogeneous and homogeneous (microcells and macrocells) networks having a different number of microsites, based on data traffic situations for calculating EE. The results indicated that better EE can achieve in HetNets deployment. Fig. 12 represent architecture in which different networks are deployed to form HetNets. The deployment of more microsites in network increases the throughput of the network. In fig., various links like (macro-macro, macro-micro, macro-pico, macro-femto etc.) are present.

Table V summarizes the various schemes and algorithms used in the available literature to improve EE in HetNets. The table also showing the description and percentage of

TABLE V: Algorithms/Schemes for improving energy efficiency using Hetnet

References	Objective	Description	Energy saved	Scheme/Algorithm
[43]	Multi-layer nature of HetNets is analyzed which shows energy consumption depends on the traffic requirement and BS hardware components.	Consider the classical 0-1 Knapsack issue. The scheme is analyzed to reduce consumption of energy by adapting the BS on/off situation.	35%	On/off scheme
[44]	Micro BSs have been Deployed in the area where macro BSs were situated and calculated saved power at the various radius of micro BSs.	Analyze the placement of various BSs based on a new power-saving metric.	Maximum energy is saved in Greedy BSs deployment algorithm as compared to others	Region-based, Grid-based and Greedy BSs deployment algorithm
[45]	Deployment of the heterogeneous network which is having different size cells (micro, pico, and femto) for achieving green efficient system.	An energy-saving method is proposed that permits femtocell BS to completely switch off its communication and processing when not involved in an active call.	37.5%	Voice traffic model
[46]	To calculate the EE gains	Implementation of femtocells within a microcell, while maintaining QoS. The results show that the usage of femto cells minimize the power consumption of the integrated network.	70%	Fast fading and path loss model
[37]	To analyze the impact of energy efficiency.	Deployment of different picocell in a macrocell.	60%	-
[47]	Investigate the area power consumption in terms of network performance metric	Discuss the effects of many deployment policies on the power consumption of mobile system.	55%	Power consumption models

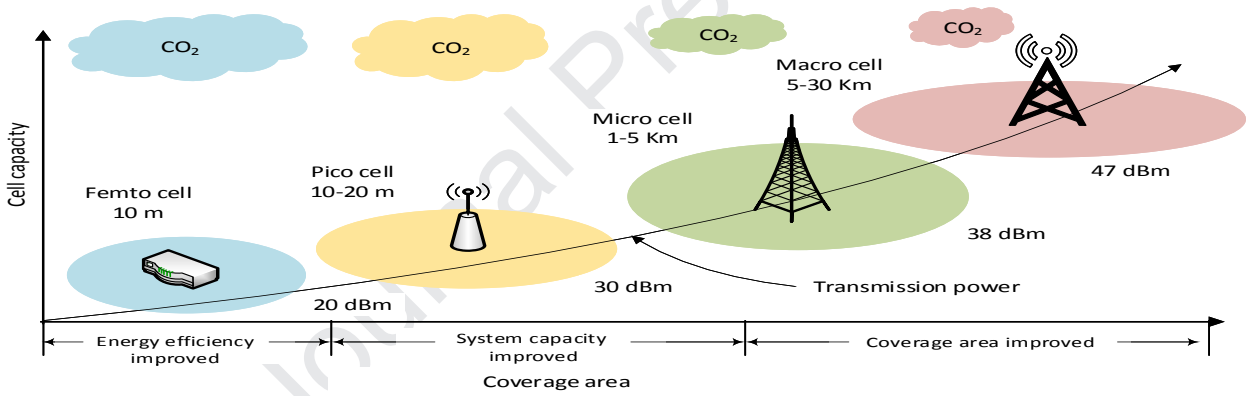


Fig. 11: Power consumption by different cell size for HetNet

energy saved by these schemes. A combination of different networks and smaller size cells provides an improvement in EE, coverage, and capacity because of the short transmitter-receiver distances. There are also some challenges in this technique, including the organization of the HetNets interfaces and the dead zone problem.

2) Cell Zooming: Cell zooming technique not only saves energy but also improve network performance. In this technique, the size of the cell varies with variation in traffic load, state of channel and requirement of mobile subscribers. In the condition of high traffic load, the congested cell will be in the zoom-in mode and low traffic cells will be in zoom out mode to provide the service to the mobile subscribers of the congested cell. In cell zooming algorithm the time is represented by $'T'$ (called cell zooming period). There are three stages in each period such as coordination, transition and serving phase as shown in Fig. 13. A cell zooming server (also known as the brain of the system) will sense the traffic situation and then decides the occurrence of cell zooming operation. A specific control message is released after the sensing process. After coordination with neighbor cells, the required cell will zoom in/zoom out by cell zooming server under the serving phase. The working of the cell zooming approach is represented in Fig. 14 having a seven cells cluster cellular network. Six nearby cells are covering one central cell as shown in Fig. 14a. The position of BSs is at the center of each cell. In a cell, mobile nodes are located arbitrarily. The central cell will perform a zoom-in operation when the cell is congested by the movement of mobile nodes into this cell. Thus, the size of the cell is minimized as shown in Fig. 14b. On the other hand, The central cell will perform the zoom-out operation, in the case when the movement of the mobile node will out from the central cell. As from Fig. 14c. congestion can be avoided by the zoom-in operation of neighboring cells.

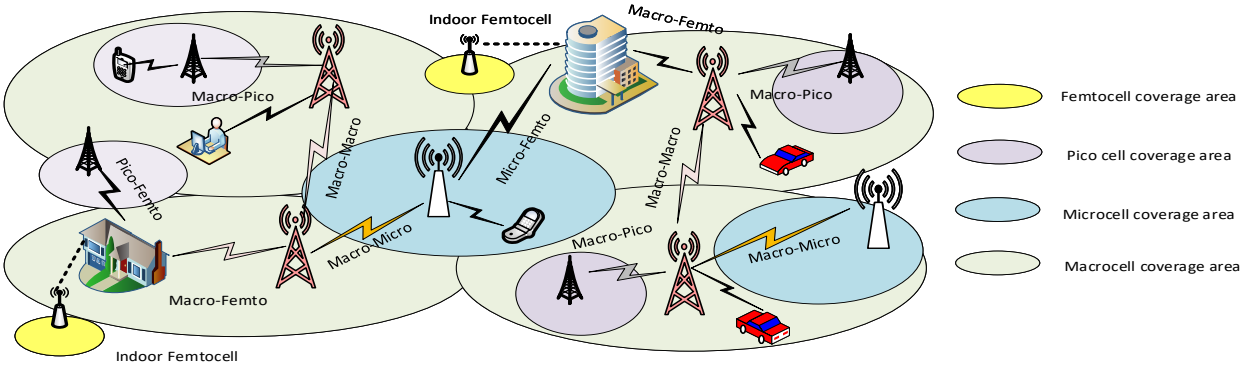


Fig. 12: Architecture for energy efficient HetNets

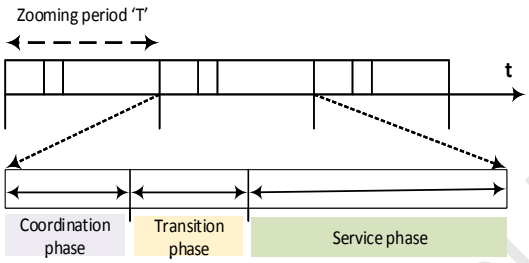


Fig. 13: The framework of cell zooming process

The mobile users will get service by nearby cells when the central cell is less congested and goes in sleep mode as shown in Fig. 14d. Table VI represents the literature based on many algorithms which are used for cell zooming technique. Almost 50% of energy can be saved by cell zooming technique.

There are various benefits of the cell zooming technique. It provides proper load balancing and reduces the blocking probability. This technique also enhances the battery life of the handset and throughput but intercell interference is the main disadvantage of this technique when some active nearby cells perform zooms out operation simultaneously. In cell zooming operation, coverage holes (no network coverage at some places) may occur. The newly arrived mobile user can be served by performing zoom in operation by nearby cells to cover these unreachable areas.

C. Traffic based Energy Efficient Approaches

The traffic based energy-efficient approache is discussed in the following subsections.

1) Base Station ON/OFF Technique: In wireless networks, energy can be saved while maintaining QoS by turning off idle wireless resources and BSs. The traffic load condition varies with space and time in the coverage region. So, underutilized BSs can be dynamically turned off. The user can be served by remaining BSs which are in an active mode of operation [54].

To ensure global coverage, the telecom operators are expanding to almost remote areas. Fig. 15a shows the data of off-grid and poor-grid BS from 2014 to 2020. Due to the absence of electrical grids, the off-grid BSs are expensive. In 2009, a telecom company named China Mobile has minimized the expenditure of energy (nearly 36 million kWh/year) by applying BS sleep control. In [55], an optimization approach related to reducing the consumption of energy is suggested with QoS constraints. The model is a combination of signal propagation models and a set of actual sized universal mobile telecommunications system (UMTS). The outcome consists of very useful constraints like guaranteed service rates, full coverage, minimum delay and accomplishes a significant reduction in network energy almost 35 and 57%. The amount of consumed energy depends on the BS type and application scenario. From Fig. 15b, 15c, and 15d, it is cleared that how much energy percentage is consumed by the separate parts of different BSs.

In a few cases, the power utilized by the BS is considered as zero in sleep mode but in actual it consumes very less power as compared to the active state of BS. However, the BS never goes in completely off state, it may still utilize some power like detection power [56]. This type of power consumption can not be ignored, especially in BSs that do not have on-site cooling systems and installed in a small location. Sleep mode is further divided into many sleep levels. In [57], an opportunistic sleep concept is discussed with an idea of BS sleep only for a certain period to improve the system capacity. The range of energy saving depends on various traffic load conditions and lies from 13% to 41%. In [58], an author describes four status of the BS and their respective power consumption ratio as represented in Table VII.

The BS with sleep/wake-up strategy is discussed in [59]. According to the traffic density, BS switches among active, sleep and off mode. Presently, a new module named as sleep/wake-up is augmented in each BS. Nearly 72.9% of energy can be saved by using this module. In [60], the distance between BS and the mobile node are considered as criteria

Journal Pre-proof

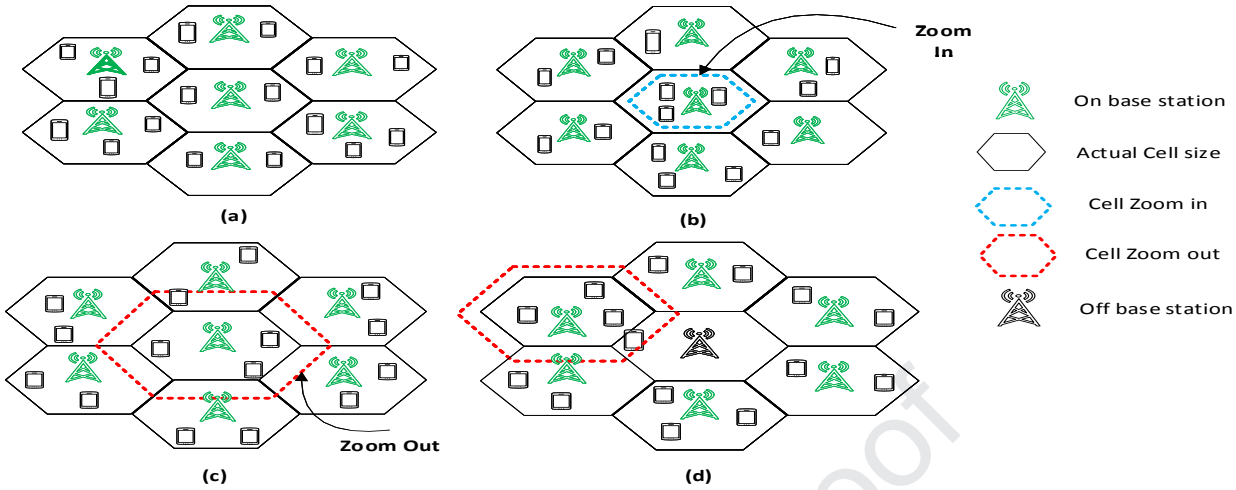


Fig. 14: Concept of cell zooming (a) Typical layout of cell (b) Cell zoom in because of high congestion in network traffic (c) Cell zoom out to serve mobile users (d) Data transfer in sleep mode via another nearby cell.

TABLE VI: A literature survey on Cell Zooming Technique

Objective	Description	Energy saved	Scheme/Algorithm	References
To achieve user traffic request expands network EE and optimize user service by adjusting the coverage of cell	For cell coverage author describes Cell Zooming Factor (CZF), which gives the relationship between small cell and large cell densities. To optimize the CZF and minimize area power consumption game theory is applied.	41%	Adaptive cell zooming scheme, cell sleeping scheme	[50]
To reduce the consumption of energy at BS and maintain proper load balancing of traffic	Analyze the adaptive network traffic load balancing technique to overcome the energy consumption issue.	55.75%	Adaptive traffic load balancing	[51]
Divide the area into more small cells to minimize energy consumption	The author expressed that if the number of cells increases then more cells can be switched off, therefore blocking probability will reduce and EE can improve.	40%	cell switch off scheme	[52]
To reduce the extreme traffic of cellular network	The author considers Relaying, Physical Adjustment, BS Cooperation and BS Sleeping for implementing cell zooming concept.	55%	Adaptive cell zooming scheme	[53]

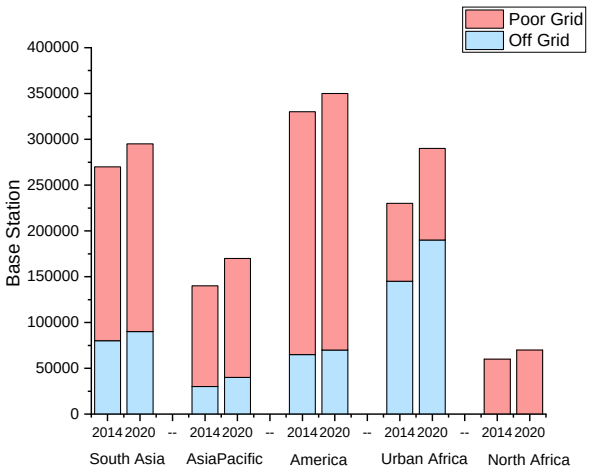
TABLE VII: Power consumption ratios status in several mode

S.no.	Base station status	Power consumption ratios
1	On	100
2	Standby	50
3	Sleep	15
4	Off	0

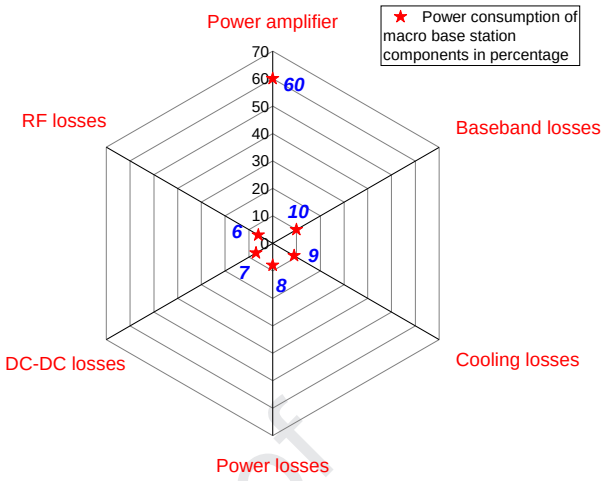
for decision making about making BS in on/off state. This technique reduces the energy consumption of nearly 29%. The switching period of the BS (switching from on to off and off to on state) is associated with QoS and plays an important role [61]. In [62] the author investigates that the processing time, required for handoff is ranged from 50 ms and 100 ms. The cooperation among BSs is necessary to serve the mobile users belonging from switched-off cites. A flowchart is proposed, related to BS on/off operation to attain an energy-efficient solution as represented in Fig. 16. The BS on/off technique achieves a remarkable trade-off between QoS and energy saving. To make the energ-efficient system, the uncovered region related to inactive BSs is covered by increase the transmission power of active BSs. Since the BS power is limited so an issue related to network coverage occurs. The battery life of the mobile users is also affected by this approach because of the requirement of high power at the receiver end to maintain connectivity with nearby BS.

D. Time based Energy Efficient Approach

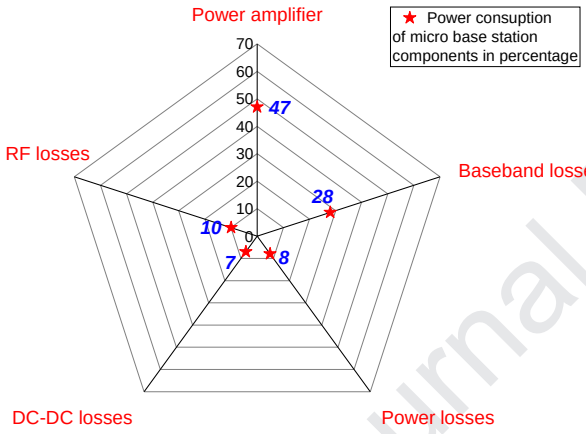
In the time domain approach, the PA of LTE base station is temporally turned off during low network traffic in the downlink and idle channel state. This power-saving solution is tightly associated with the LTE system frame structure. The concept of the frame structure, reference signals (RS) and control signals have investigated in [63]. LTE downlink uses orthogonal frequency division multiple access (OFDMA)



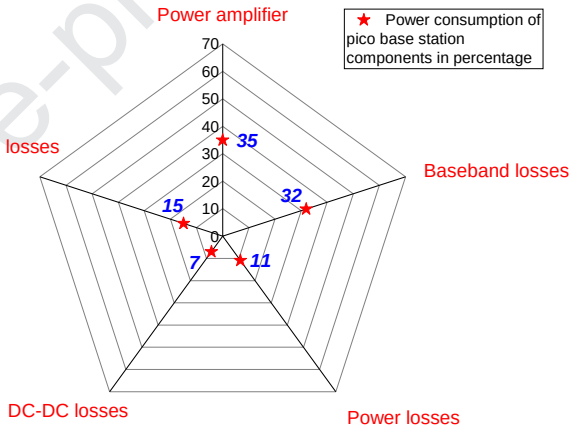
(a) Increment of base station from year 2014-2020



(b) Energy consumed by the different part of Macro base station



(c) Energy absorbed by various parts of Micro base station



(d) Energy absorbed by various part of Pico base station

Fig. 15: Energy consumed by different parts of different BS

approach. In this, an RS signal performs two main tasks in the network such as channel estimation for the mobile node and search neighbor cell for handover [64]. There are mainly three techniques to minimize the number of RSs.

1- Normal unicast:- The LTE frame consists of 7 symbols per slot. The RSs, applies in 0, 4, 7 and $11th$ OFDM symbols at every subframe. This pattern of RSs keeps on slot $1-4$ and $6-9$ of subframe and does not apply on slot 0 and 5 of the subframe. Slots 0 and 5 are used to transmit control signals [65].

2- Multicast Broadcast Single-frequency Networks (MBSFN):- The services like mobile TV and DTH use the MBSFN technique in the LTE system. In this, the pattern of RSs keeps only in slot 1. So, the number of RSs is reduced.

3- Cell Discontinuous Transmission (DTX)/Microsleep:- In DTX approach, the RSs signals are not transmitted in slot $1-4$ and $6-9$ of subframe if no mobile node is found in the active state [63].

Fig. 17 represents that each frame of the LTE system includes 10 subframes (from 0 to 9) and there are 7 OFDM symbols in one slot. In a single slot, four RSs are used for communication purposes. Some amount of energy is required for the transmission of RSs. In cell discontinuous transmission technique no RS is sent for idle and low traffic load condition so the maximum amount of energy can be saved in this technique as compared to the normal unicast and MBSFN technique in which RSs are sent at low traffic conditions.

In [66], a dynamic subframe allocation technique is introduced for multimedia broadcast multicast services with MBSFN. In [67], an upcoming network traffic load is predicted by MBSFN. The BS turning off periods are calculated with the help of MBSFN based energy-saving technique [68]. In this technique, the needful resources are pre-calculated based on traffic load. Thus, more energy is saved by keeping off underutilized resources. This technique also has features

Journal Pre-proof

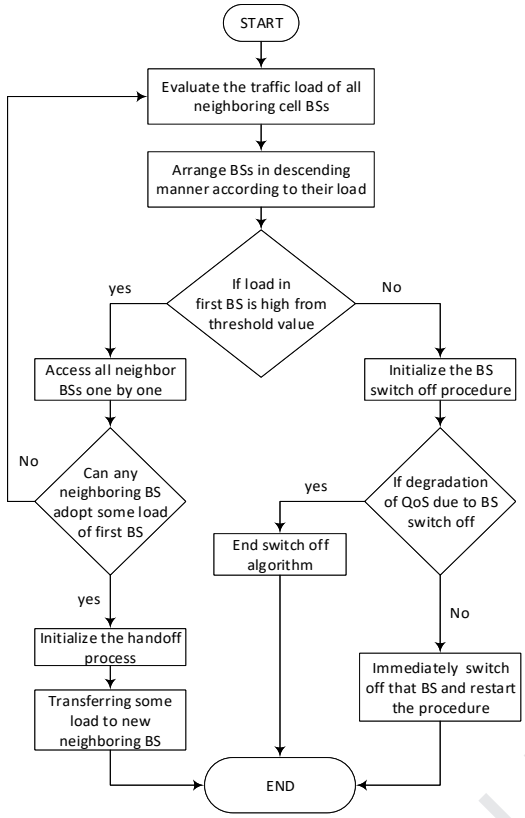


Fig. 16: Flow diagram of proposed BS sleeping operation

like backward compatibility, delay and process time [69]. In addition to this, Fig. 18 shows the trade-off between on period of PA in a frame to the energy-savings for various time-domain approach. The maximum amount of energy can be saved by the extended cell DTX technique. This technique gives the best results for rural areas where traffic is low and idle case scenario. This technique also saves energy at components levels like cooling systems, DSP units and ADC/DAC devices in turn-off time duration. Some mobile node faces troubles due to the large number of RSs during synchronization operation with BS. Thus, the mobile node may fail to admit in DTX mode. This reduces the battery life of the mobile node and causes a negative effect. There is also difficulty arises during control signal decoding.

E. Frequency based Energy Efficient Approaches

Energy-efficient approache based on frequency is discussed in the following subsection.

1) Carrier Aggregation: The LTE-A (Long Term Evolution Advanced) system uses the bandwidth of up to 100MHz in different frequency bands [70]. Such frequency bands are 450–470 MHz, 698 – 960 MHz, 1710 – 2025 MHz, 2110 – 2200 MHz, 2300 – 2400 MHz, 2500 – 2690 MHz and 3400 – 3600 MHz define by the International Telecommunication Union (ITU) [71]. The carrier aggregation (CA) technique used by the LTE-A, exploits and supports various frequency bands. The concept of the CA technique is to amalgamate two or more component carriers (carrier frequency) into one frequency to amplify the channel capacity and data rate. The frequencies belonging to the same or different frequency bands can be combined in this technique [72]. The concept is represented in Fig. 19. In [73] author considers multi-cell and single cell scenario to analyze the impact of CA on overall power, send by eNodeB. A relation among assigned bandwidth, cell radius, and the total send power by the eNodeB is expressed as:

$$eNB_{TP} = \frac{2R^{\alpha}(2^{\frac{D\times\sigma}{BW}-1})\mu \times BW \times N_0}{(\alpha+2)kd_0(-\ln[1-P_{r(out)}]^{\frac{1}{M}})} \quad (4)$$

Where eNB_{TP} represents the total send power of eNodeB and R is the cell radius. α gives the value of channel path loss exponent, D is the minimum data rate for better QoS. The total number of active users is represented by σ. The term BW, μ, T and K represent transmission bandwidth, modulation coefficient, noise temperature and Boltzman constant respectively. Here $K = 1.38X10^{-23}m^2kgs^{-2}K^{-1}$. The transmission power in both scenarios decreases by nearly 65% by applying the CA technique and EE of the network improves. According to [74], the carriers are combined into groups in CA technique and separate PAs serve every group individually. The limitation of this approach is that the BS with total carriers must have individual PA connected to every group of carriers.

CA supports aggregation of licensed carrier and unlicensed frequency spectrum. Hence unlicensed bands are also utilized. Scaling of the carrier also facilitates the expanded coverage. This technique also helps in enhancing network efficiency and optimizes user experience. Multiplexers are required at transmitter end to combine the carrier components and RF filters are required at the receiver end to isolate carrier components. PAs and switches are specially designed to fulfill the need of CA-based RF system. These additional components raise the cost of the apparatus and device manufacturers and also increases system complexity.

F. Momentous Approach

1) Renewable Energy Resources: Usage of renewable energy resources (RER) in the mobile communication system for making the system greener is a very remarkable approach. With growing wireless communication, infrastructure needs a high amount of electricity. These RER is used for powering cellular system and having many advantages. The major advantage and popular justification behind using RER are to make the environment eco friendly and reduce the emission of GHG. Currently, India has more than 320000 mobile phone towers, which consume almost 3 billion liters of diesel per year. This switching from diesel to solar and other alternative energy sources will result in deduction of nearly 6 million tons of CO_2 emissions and save nearly 1.5 billion in operating expenses for wireless communication tower companies. Switch to RER can produce millions of carbon credits that could balance the OPEX on telecommunication towers. Areas like islands and deserts where the installation of a smart power grid is difficult, the RER can be considered as a problem

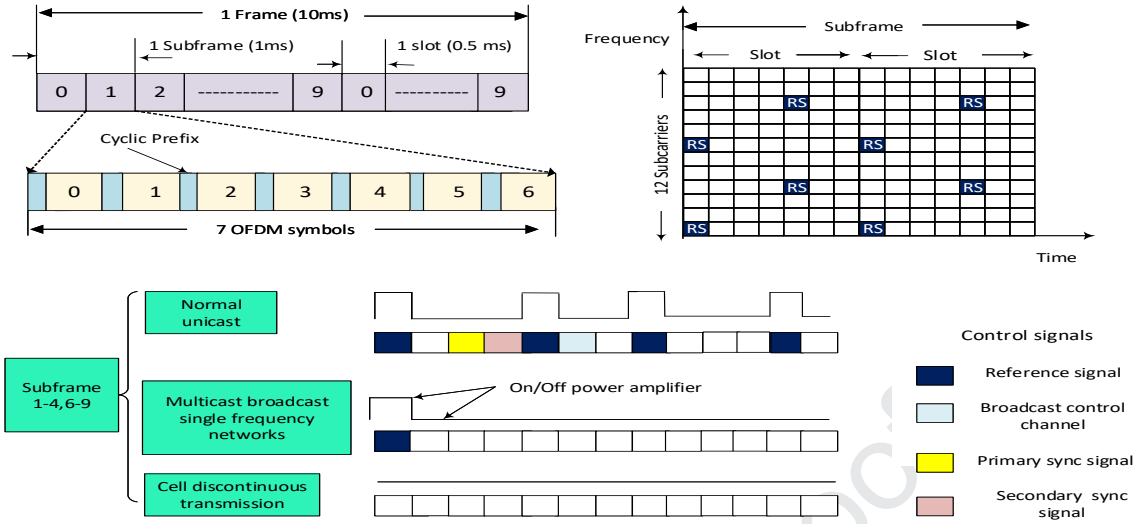


Fig. 17: Reduction of RSs in LTE frame for idle and low traffic load condition

TABLE VIII: The value of power consumption for various cellular BSs

Base Station Type	NTRX (number of transceiver chains)	PA (W) power consumed at power amplifier system	BB (W) power consumed at baseband system	Cooling (%)	Mains Power(%)	DC-DC (%)	RF (W)	Total per TRx (W)	Total Power for NTRX (W)
Macro	6	128.2	29.6	10	9.0	7.6	12.8	225.0	1350
Micro	2	6.2	2.6	0.0	9.0	39.0	54	72.3	745.8
Pico	2	0.13	4.0	0.0	11.0	4.4	6.8	7.2	14.8
Femto	2	1.1	2.5	0.0	11.0	9.0	0.6	5.3	10.5

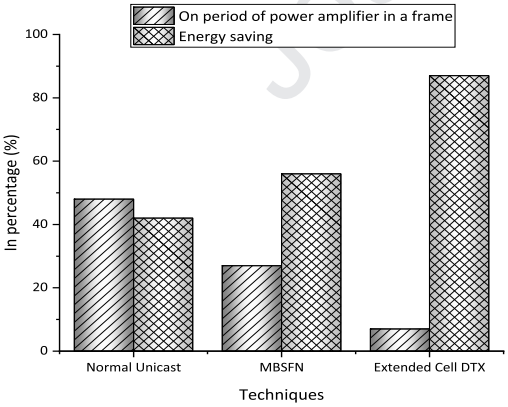


Fig. 18: Tradeoff between active time of power amplifier and energy saving

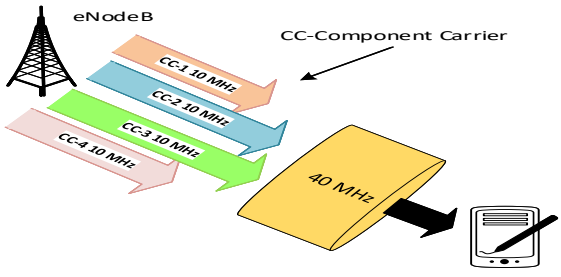


Fig. 19: Concept of carrier aggregation

solving solution [75]. Until now, in such types of areas, mobile operators depend on diesel generators for working of BS which is one of the most pollutant origins of energy and also very expensive because of its price and transportation [76]. The several RER likes sunlight, wind, biomass energy are not only sustainable but also clean in nature and can be used in future networks. The GSM Association has been initiated a program named as "Green Power for Mobile" to support industries related to telecommunications and also to establish several RER for proper operation of 640000 new and prevailing off-grid BSs [77]. With this initiative, nearly 0.347% of the global diesel consumption saved.

In [78] the author describes how RER can reduce the practical realization and working costs of upcoming future

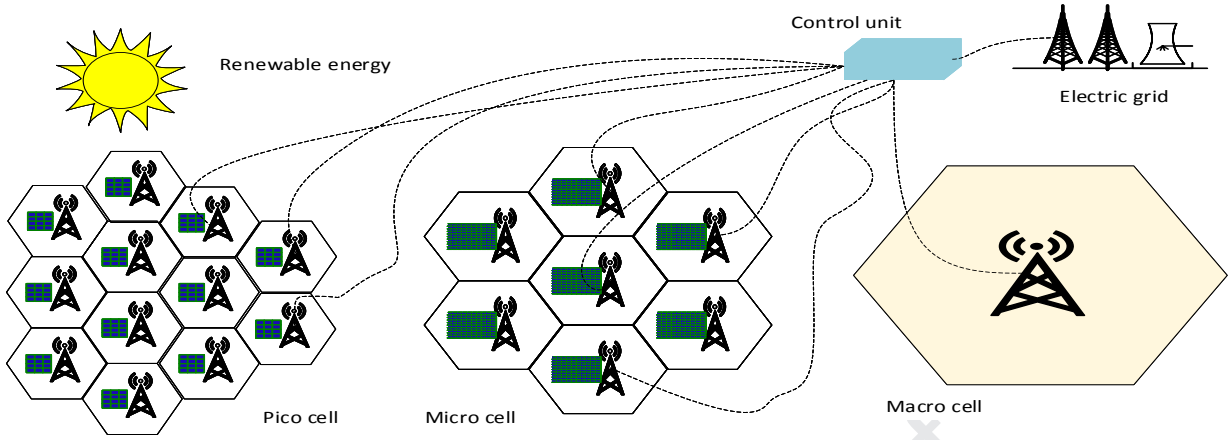


Fig. 20: Addition of renewable energy resource with cellular system

wireless communication systems. Sunlight is an excellent RER used for electricity generation. Another RER used for powering the BSs are wind power, fuel cell, hydro-power and hybrid (solar and wind) system, etc. Fig. 20 depicts the utilization of RER with the wireless communication system. Here solar panels are connected with BSs of the macrocell and microcell to convert solar energy into electrical energy [76]. The major components responsible for power consumption in BS are PA, baseband unit, RF module, cooling system, DC power supplies and losses incurred in feeder [79]. The power consumption in BS can be determined as

$$P_{consumed\ at\ BS} = N_{TRX}\frac{P_{BB}+P_{PA}+P_{RF}}{(1-\xi_{DC})(1-\xi_{MS})(1-\xi_{cool})} \quad (5)$$

here N_{TRX} is the number of transceiver chains. The P_{BB}, P_{PA} and P_{RF} denote power consumed by baseband system, PA and RF system, respectively. The ξ_{cool}, ξ_{MS} and ξ_{DC} represent losses associated with the cooling system, main power supply and DC power supply, respectively. The P_{PA} can be given by the ratio of antenna feeder power output to the addition of PA efficiency (η_{PA}) and feeder loss (ξ_{feeder}) and expressed as

$$P_{PA} = \frac{P_{output}}{\eta_{PA}(1-\xi_{feeder})} \quad (6)$$

Table VIII represents the consumption of power (as a percentage of the overall consumption of power) associated with various parts of BS. The consumption of BS power is divided into two types named as static and dynamic. In first, the power consumption remains constant. In the second, power consumption depends on the number of transceivers deployed and traffic-load. In BS the PA consumes maximum power about 50 to 60% of the total power, and linearly depends on traffic-load. This can be expressed as

$$P_{Consumed} = N_{TRX}(P_{static} + \tau_p P_{Dynamic}) \quad (7)$$

where $0 < P_{Dynamic} \leq P_{\max}$

Here τ_p is the slope of traffic load dependent power consumption. The term $P_{Dynamic}$ and P_{static} represents static and dynamic power respectively.

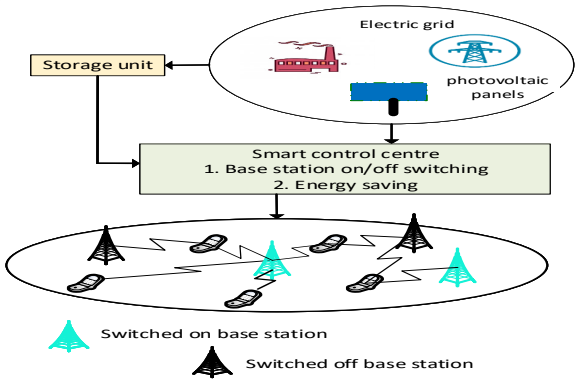


Fig. 21: Future wireless network powered by smart grid

2) Smart Grid: This is an integrated communication system which has various major subsystems such as wired/wireless communication channels, the smart control center (also work as a transmitter/receiver) and software-based management systems. Fig. 21 represents the architecture of future green wireless communication networks concurrently powered with the help of a smart grid and locally deployed RER. The key component of the architecture is the smart control center which sets up the connection between operator and smart grid to optimize wireless network operations like resources management, taking decisions about turning on/off the BSs periodically and storage of auto-generated energy amount etc. For a power supply, the BSs cluster is linked with the electric grid and also connected to the locally generated RE, which generates energy by the array of photovoltaic panels. Due to the irregular production of the RE, the system is connected

TABLE IX: Comparison among various techniques in terms of energy saving

	SISO	SIMO	MISO	MIMO
Merits	Simple circuitry	Known as receiver diversity	Transmitter diversity	Improvements channel throughput
	At short distance SISO system is more energy efficient than others	Under low traffic SIMO system is more energy efficient than others	Less battery consumption	Improve EE for long distance communication
	No additional processing required	Overcome the effects of ionospheric fading and interference	Multiple antennas are combined to decrease errors and increase data speed	Reduce co-channel interference
	Diversity not required	Easy to implement	Reduce the problems of intermittent reception	Improve channel capacity and channel robustness
	$C_{SISO} = B\log_2(1+\frac{S}{N})$	$C_{SIMO} = A_r B\log_2(1+\frac{S}{N})$	$C_{MISO} = A_t B\log_2(1+\frac{S}{N})$	$C_{MIMO} = A_t A_r B\log_2(1+\frac{S}{N})$
	here C, B and S/N are respectively capacity, bandwidth and signal to noise ratio	here C, B, S/N and A_r are respectively capacity, bandwidth, signal to noise ratio and no of receiving antenna	here C, B, S/N and A_t are respectively capacity, bandwidth, signal to noise ratio and no of transmitting antenna	here C, B, S/N, A_t and A_r are capacity, bandwidth, signal to noise ratio, no. of transmitting and receiving antennas respectively
Demerits	Limited performance	Additional processing required at receiver	Additional processing required at transmitter	Requires high processing
	Channel bandwidth is limited	Problem of battery drain	Channel capacity is not Improve	Complex circuitry
	Low throughput	High cost than SISO	High cost than SISO	High cost

with a storage unit, which stores the extra amount of produced RE that can use when RE is not generated, so that the uneven production of RE can be compensated. There is an additional connection between the smart grid and the storage unit shown in the figure. Hence, this storage unit may be used to harvest energy produced from the grid. The arrows represent the possible direction of energy flow.

There are several benefits of using a smart grid in next-generation wireless networks. For mobile operators, the smart grid can be work as a tool for the optimization of energy acquirement, bound their CO_2 emission and reduce their energy bills. The main functions performed by the smart grid are i) To integrate control and information technologies for increasing the generation. ii) Efficient allocation and utilization of energy. iii) To serve consumer's mobile with safe and secure functionality of the smart grid. iv) To design demand-based management that provides controlling choices and secure information to consumer's mobile. In [80], the cost of power consumption with maintaining consumers is minimized by the integration of the smart grid and RER. In [81] a novel combined architecture of RER and electrical grid for wireless networks is designed to utilize the advantages of network operators and smart grid.

In [82], a real-time energy management system based on the smart grid is suggested with the beamforming technique to limit the cost of energy consumption by using locally generated RER. In cooperation with a smart grid, the network state depends on several factors like mobile subscriber density, day of the week (weekend, holiday, working), day time (noon, office, busy hour, midnight) and user's behavior profile, etc.

Smart grid often encourages cost-effective energy saving and high penetration of local renewable energy generation to support future green communication network and also, minimizes the overall emission of total harmful CO_2 with improved QoS. The smart grid system is complex and has number of problems that required to be handled in order to set up a consistent and efficient communication system.

G. Space based Energy Efficient Approach

The various space based energy efficient approaches are discussed as:

1) Multiple-input and Multiple-output (MIMO): This is the most promising approach considers for designing of future networks. Without increasing the transmission bandwidth and power, this technique has overcome the problem of fading and improves the data rate, capacity, and throughput as compared to other antenna systems.The fading effect at the receiver end can be minimized using diversity coding concepts in the MIMO system. Available literature indicates that EE of MIMO system is highest than the other antenna transmission like SISO (Single-input/Single-output), MISO (Multiple-input/Single-output) and SIMO (Single-input/Multiple-output) due to the existence of a larger number of antennas. Table IX shows a comparison among these techniques in energy-efficient communication perceptive.

In [83], three important parameters such as capacity, path loss and delay are considered to make energy-efficient green BS. A trade-off between diversity gain (transmitting similar signals through multiple paths) and multiplexing gain (sending independent signals) is discussed in [84]. In [85] author stated that energy can save at BS by employing steerable directional antennas work on frequency 28 GHz and 38 GHz. In [86], a resource allocation scheme is introduced to decrease the amount of data transmission energy. According to [87], [88]

energy saving MIMO technology depends on traffic demand in that area, so by proper traffic load adaptation, the performance of the MIMO system can improve. In [89], the number of reference signals in MIMO system is decreased in frequency and time domain to make transmission energy-efficient. In [90], a switch-off RF and PA units based novel approach is introduced to reduce energy consumption. It means, at low traffic system, switching from M-antennas to 1 antenna reduces the consumption of energy by the factor 1 by M.

2) Massive MIMO (M-MIMO): M-MIMO is one of the advanced technology of MIMO having several antennas at BS. This technique requires shorter wavelengths (higher frequencies) because the system needs to physically pack more antennas into a small area than the other mobile network [91]. This approach is not only energy efficient but also spectrum efficient. The main advantage is that a BS can serve multiple subscribers simultaneously within the same spectrum. This results in better multiplexing and array gain at the same time. Now the question is for green massive MIMO system, how many antennas are required? In [92] author suggested that in the M-MIMO system, a BS can have X antennas and subscribers have Y single-antenna. A subscriber can scale down uplink power to X and $\sqrt{X}$. In [93], a power consumption model is investigated for obtaining an energy-efficient optimal solution by installing $150 - 250$ antennas to serve a large number of mobile subscribers. Fig. 22 depicts the concept of M-MIMO where an array of antennas is situated in BS and serves multi users simultaneously. Table X summarized a literature survey on M-MIMO techniques and table XI provides a comparison between MIMO and massive MIMO considering energy saving prospect.

H. Electromagnetic Radiation based Energy Efficient Approach

Now a day's mobile communication networks have become an important part of human life. Enormous efforts are dedicated to developing mobile communication services without thinking about the adverse effects of its on human. It is always a topic of discussions among people about the installation of BSs near the residential areas. The Electromagnetic Radiation (EMR) from these BSs causes long-term effects on animals, birds, and humans. The literature stated that BS's EMR causes low risk if antennas are not directly in contact. But the long-term effects of this radiation are still concerning. Paradoxically, BS is considered as the main source of EMR exposure, even though there is a greater health risk from the emissions by mobile phones because those antennas are directly in contact with the human body usually head and ear when it operates. This EMR can interpose the working of our body organs particularly that of our brain and also the reason for some serious health problems such as depression, insomnia and a shortened attention span [102]. A report of the World Health Organization (WHO) shows the classification of RF emissions as a carcinogen of the group 2B category, which is a category shared by lead and pesticides. This study suggests that excess usage of cellphone leads to the development of malignant tumors, especially of the brain. Some research also indicates that exposure to non-ionizing radiation can lead to skin problems, fatigue, headaches, allergies, Parkinson's disease, Alzheimer's disease and more [103]. To overcome this problem, Specific Absorption Rate (SAR) are explored to minimize EMR. Thus, these techniques provide the direction for the designing of next generation low-EM exposure to wireless networks.

EMR Metrics– For finding the direction of EMR of any antenna, two factors named as directivity of an antenna and transmit power are used. The antenna which radiates uniformly in all direction is known as an omnidirectional antenna on the other hand directional antennas radiates in a particular direction. The high value of antenna gain causes stronger the signal in a particular direction. According to the electric and magnetic field theory, depending upon the distance from the transmitting antenna two regions exist such as Far-field and Near-field regions. The two main metrics are used in this field for calculating the EMR, first, is SAR considered for near-field and the second is power density for the far-field. Fig. 23 shows EMR fields. Characterizing the EMR fields in the reactive area is a little bit complicated. The human body easily absorbs the energy transmitted within the reactive near-field region. In the reactive near-field region, the field strength changes inversely with the cube of separation from transmitting antenna i.e., $1/(distance^3)$.

SAR– SAR is evaluated when the body is influenced by EMR. Since the phone is generally near to the body, so EMR from communicating node is typically in the near-field. The value of EMR exposure in the near-field region is calculated for a given period t -

EMR exposure $(J/kg) = SAR.t$

$$SAR = \frac{\delta}{\delta t}\left(\frac{\delta E_A}{\delta M}\right) = \frac{\delta}{\delta x}\left(\frac{\delta E_A}{\Psi \delta V}\right) \tag{8}$$

Where E_A gives the value of energy absorbed by the body, V, M and Ψ represents volume, mass and density of the body, receptively [104]. Different body parts have different SAR value because of different absorption rate as represented in Table XII. *EM Radiation Reduction Techniques–*The target of the 5th generation network is to increase the spectral efficiency, throughput and data rate as well as it also focuses on criteria like improving latency, energy efficiency and reducing EMR exposure from the wireless networks. In this paper some approaches are discussed that are using to decrease the transmit power and EMR exposure levels and SAR of wireless networks because these parameters are directly related to each other. These EMR reduction techniques are:

1) SAR Shielding: Environment and human health [111] have been unpleasantly affected by the increased EMR in the surrounding. The absorption of this EMR by the human body is defined as SAR. For human health higher SAR is very injurious. Different manufacturers have a different value of SAR and this value is governed by several environmental bodies around the world. SAR shielding is used to overcome this problem. In this technique, a metamaterial/ferrite is sandwiched between the human head and mobile, to minimize SAR [112]. The conductivity of ferrite materials is very small therefore there is a small current-induced [113]. The

Journal Pre-proof

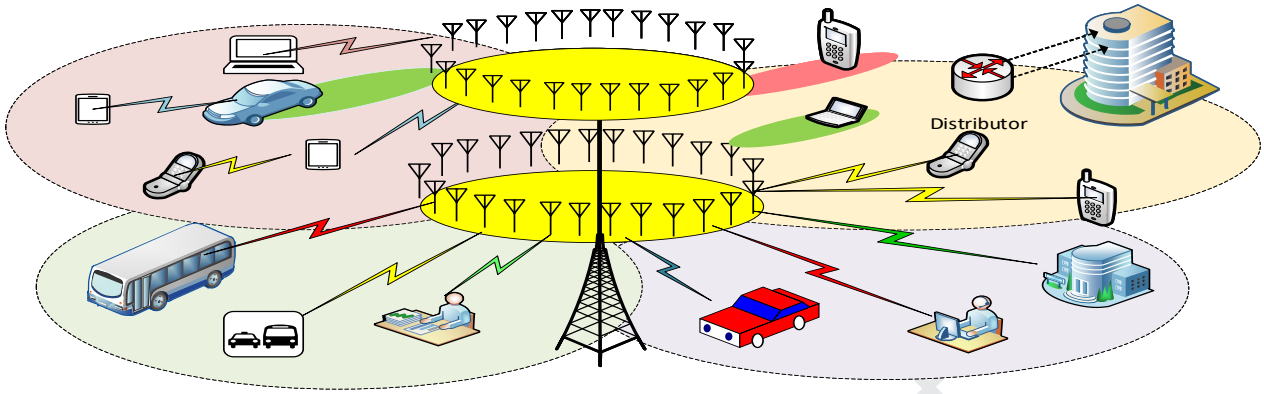


Fig. 22: Massive MIMO antenna array layout

TABLE X: A summary on Massive MIMO techniques

Proposal	Methodology	Scheme/Algorithm	Route	Remark	Reference
To design cell-free MIMO downlink, which provides the service of various single antenna users in the same frequency-time resource.	Work on spectral efficiency with the consideration of the consequence of channel approximation errors, power control, non-orthogonality of pilot channels and backhaul power consumption.	Conjugate beamforming scheme with consideration of local channel state information.	Downlink channel	Design optimal power control algorithm which increase EE as well as access point power constraint and user spectral efficiency.	[94]
Energy utilization in massive MIMO system is optimized.	For a massive MIMO network consider single-cell and multi-cell scenarios.	Bisection searching algorithm	Uplink channel	Optimizing pilot sequences length and find optimal parameters for energy efficient massive MIMO system.	[95]
Work on joint power allocation in source, channels and relay to maximize the EE.	Massive MIMO relaying technique has been used to enhance wireless PHY security, even with no instantaneous legitimate channel state information.	Energy-efficient power allocation scheme	Downlink channel	Energy-efficient joint source and relay power allocation schemes maximize the EE.	[96]
This paper presents the performance of MIMO cellular systems with pilot contamination.	Consider uniformly distributed user location scheme, find out the area EE in a cell.	Power consumption model	Uplink channel	Investigate closed-form estimation of the area spectral efficiency for multi cell multiuser massive MIMO systems.	[97]
Investigate the zero-forcing beamforming, maximum ratio transmission and scaling laws of EE.	Circuit power, transmit power, pilot contamination, and channel estimation errors are considered for improving the EE.	Realistic channel models' scheme	Downlink channel	-	[98]
Design cell free architecture with the consideration of user centric approach to improve coverage.	An interaction between alternating and sequential optimization is considered.	Hybrid analog/ digital beamformers	Uplink/ Downlink channel	To maximize global EE downlink power control algorithms are used.	[99]
Multi-user, multi-cell, state is considered in which every user work to improve their EE.	Integration of antenna arrays are considered	Distributed power control algorithm	Uplink/ Downlink channel	Evolutionary game theory (EGT) is analyzed to maximize the fairness between users in the network.	[100]
Enhancement in system EE, spectral efficiency and reduce energy consumption for massive MIMO system.	Use full duplex method to improve spectral efficiency and EE.	Self-backhauling and energy harvesting jointly used	Uplink/ Downlink channel.	System EE improved by reducing multi-tier interference	[101]

magnetic field region of an incident EM wave is revoked by this metamaterial/ferrite. This metamaterial has negative permittivity. When metamaterial/ferrite is layered between the human head and mobile model, the value of SAR reduces to 47.68% over 1 gram of tissue on the other hand 27% and 52% over 1 gram of tissue at frequency 900 and 1800 MHz.

TABLE XI: Comparison between MIMO and Massive MIMO

Parameters	MIMO	Massive MIMO
Energy Efficiency	Improved EE	100 times improved than MIMO because of the focus of radiated energy on UE.
Capacity	Improved Capacity	Five times improved than MIMO and coverage of lower frequency bands.
System Hardware	Less complex using up to 10 antennas	More complex using upto 100 antennas
Channel modeling	Easy channel modeling	Complex channel modeling
Battery lifetime	Reduces battery lifetime of MIMO based system because of faster processing of complex and computationally intensive signal processing algorithms.	More reduce battery lifetime as compared to conventional MIMO.
Battery spectrum efficiency	Improved SE	Improved SE because of large multiplexing gain as well as antenna array gain
MIMO based system	WiMAX (IEEE 802.16), WLAN (IEEE 802.11), 4G systems, LTE	Advanced WLAN (802.11ac, 802.11ad), LTE Advanced, 5G system.
Operation Mode	Operate in frequency division duplex (FDD) and time division duplex (TDD) both.	Massive MIMO was originally invented for the TDD operation, but can potentially be applied also in FDD operation.
Throughput	Having high downlink and uplink throughput because of multiple antennas and spatial multiplexing technique.	Large improvement in throughput

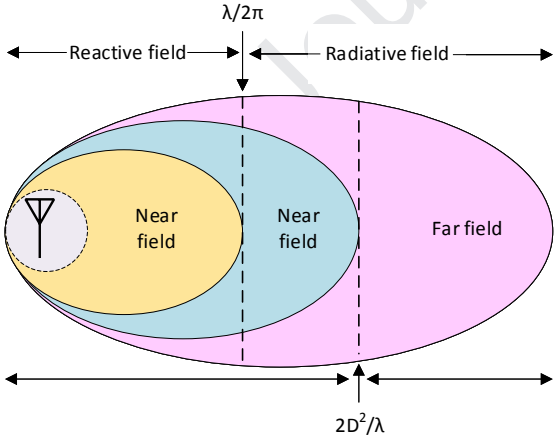


Fig. 23: Effect of EM radiation fields

Recently, the German Federal Office has published a list of top 16 smartphones emitting the most radiation with a SAR. The Xiaomi Mi A1 releases the most radiation with a SAR of 1.74 watts per kilogram while the ZTE Axon Elite and Samsung Galaxy Note 8 have emerged as the lowest radiation-emitting

TABLE XII: Different SAR calculation for different body parts

Metric	Description	Calculative Formula
Whole Body Averaged SAR (WBA-SAR)	Defined as the total power take in by the body to the mass of the body.	$WBA_{SAR} = \frac{1}{M_B}\int_{R_B} SAR(\delta m)$ Where M_B, R_B shows the mass and region of the body
Specific Organ Averaged SAR (SOA-SAR)	This depicts the absorption of EM fields by the tissue or specific organ of the body.	$OSA_{SAR} = \frac{1}{Mo}\int_{Organ} SAR(\delta m)$ Here $Mo, Organ$ shows the organ mass and organ under consideration.
Spatial Peak Averaged SAR (SPA-SAR)	Discussed as maximum local SAR averaged over organ specific mass.	-

phones with a SAR of 0.17 watts per kilogram.

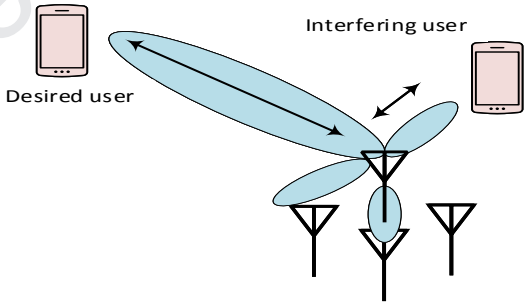


Fig. 24: Principle of Beamforming.

2) Beamforming: A direct approach to reduce power consumption from the mobile communication system is beamforming. An array of antennas is used in this technique to focus EMR towards every desired mobile phone user direction. In the beamforming approach, each transmitting signal of the mobile subscriber is multiplied by complex weights. These weights are adjustable and have a signal's phase and magnitude in a specific direction. By using the DSP algorithms, the exact location of the user can be tracked in the beamforming technique. So, the signal strength in a specific direction maximizes while interference, EMR exposure and transmit power reduces [114]. Fig. 24 shows the principle of beamforming. In figure, the beam formation is in the direction of the desired user. Table XIII represents a detailed survey of the beamforming techniques in terms of energy saving.

I. Futuristic Approach

In this subsection, various important futuristic approaches like Device-to-device (D2D) communications, Mm wave and Green IoT is discussed in an informational fashion.

TABLE XIII: Survey of Beamforming technique

Proposal	Remark	Reference
Authors investigate the effect of beamforming for SAR improvement.	Reducing the absorption of EM radiation by placing nulls in the head direction. Results indicate that a SAR reduction can be achieved at least 10 dB in this method.	[105]
To overcome the problem of EM radiation exposure radiated from mobile equipment, a less complex, SAR-aware optimum beamformer is investigated.	The SAR-aware beamformer for multi-antenna terminals is used to improve transmission power constraints and SNR. Results indicate that the SAR-aware beamformer performs better than the traditional method and provide SNR gain up to 5 dB.	[106]
Propose hybrid beamformers to fulfill the objective of reducing power while maintaining users' QoS requirements.	The author presents a phase array switching framework to comprehend the eigenmode selection beamforming. Results show that the investigated eigenmode selection method can attain a significant amount of power savings.	[107]
An energy-efficient multi-cell multi-cluster coordinated joint beamforming scheme is investigated.	The author works on antenna specific higher power constraints to overcome the problem of non-linear distortion in power amplifiers. It represents the relation between beamformers and binary antenna selection variables and numerical results illustrate the gains achieved in the form of EE.	[108]
In this paper user scheduling problem and joint multicast beamforming was discussed and another goal is to minimize the total transmitting power of multiple channels.	An algorithm based on Convex relaxation is proposed. Extensive mathematical researches illustrate the benefits of the suggested co-design model over random scheduling and fixed scheduling in terms of power ingesting.	[109]
At the transmitter end, a digital-beamforming method depends on DSP was investigate	The solution presented high throughput and improved data rate for a cost-effective 5G wireless access network which operates at frequency 28 GHz and also represents an improvement in EE.	[110]

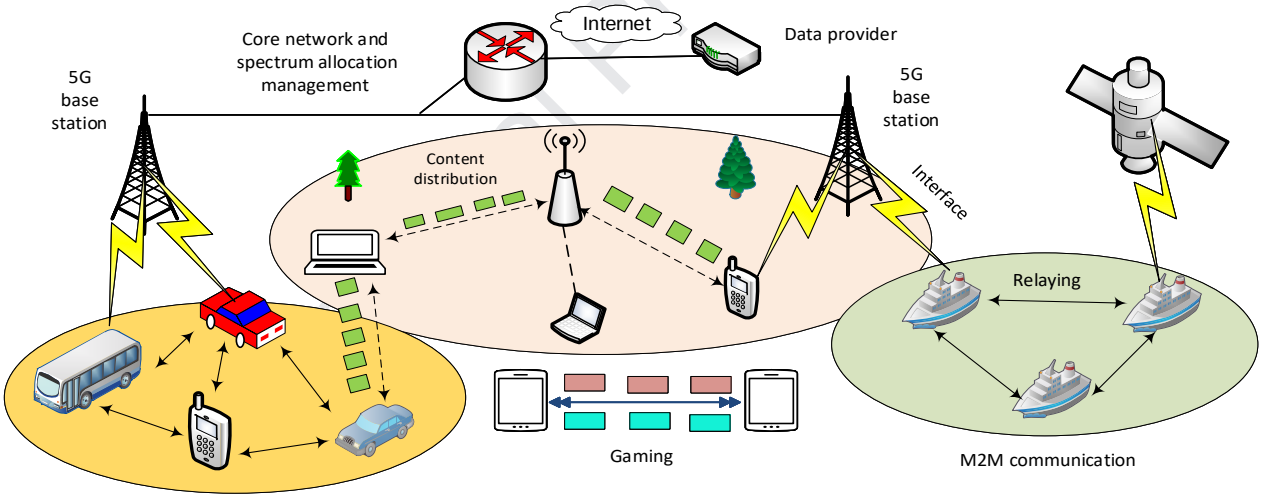


Fig. 25: Energy efficient D2D communications model

1) D2D Communications: D2D transmission is one of the effective technical approaches to reduce the consumption of power and improve the data transmission rate. In this technique two physically separated nearby located nodes can directly communicate with each other with low transmit power [115]. Thus, a remarkable amount of energy can be saved at BSs. In the case of wireless network unavailability and network congestion, D2D communications offer alternative communication. The D2D approach is recognized as a public safety network for future wireless communication by FCC (Federal Communications Commission) because of low cost and high data rates.

An energy-efficient high-level D2D communications model is represented in Fig. 25. Here different users are linked with BS via a cellular interface. These users are also communicating directly via D2D links between nodes. Several Radio Access Technologies (RATs) including 3GPP evolution like Bluetooth and Wi-Fi standards are used to establish direct links between communicating nodes such as mobile phones and laptops. In Vehicular to Infrastructure (V2X) communication, a dedicated 802.11p standard operates in the frequency band of 5.9 GHz is used by cars in an intelligent transport system (ITS).

TABLE XIV: Comparison of existing surveys of D2D in terms of energy efficiency

Proposal	Methodology	Technology	Route	Evaluation	Achieved Performance	Reference
Enhance power efficiency	Exhaustive search	LTE-A	Downlink	Simulation on system level	20% energy can be saved	[116]
Improve power efficiency	Heuristic algorithm	LTE	Uplink	Simulation on system level	30% power efficiency improved	[117]
Increase power efficiency	Linear programming	OFDMA	Uplink/Downlink	Simulation on system level	40% to 80% energy saved	[118]
Enhance EE	Linear programming	OFDMA, LTE-A	Uplink/Downlink	Simulation on system level	50% energy saved	[119]
Improve EE	-	LTE	Uplink	Simulation on numerical based	Improved EE from 0.9 bps/Hz/mW to 22 bps/Hz/mW	[120]
EE management	Inducement mechanism based on open market and sealed market model for D2D (auction algorithm)	OFDM	Downlink	Framework for cooperative and D2D communications	37% energy saved	[121]
Analysis of EE	-	LTE	Uplink	-	Energy saved bps/Joule	[122]
Analysis of EE	Distributed resource allocation algorithm	LTE	Uplink	Tradeoff between EE and SE	Energy saved in bits/Hz/J	[123]

Depends on the spectrum usage, this approach is divided into two categories named inband D2D and outband D2D. In first, the identical licensed band is exploited by both D2D and cellular networks. The major drawback of inband D2D communications is interference prompt by D2D to the cellular link. The unlicensed spectrum is utilized by D2D links in outband D2D communications. The reason behind the usage of the unlicensed spectrum is to limit the interference between cellular links and D2D network. But, more interface requires to connect different wireless technologies like ZigBee, Bluetooth and, Wi-Fi [124].

In [125], joint optimization of adaptive power allocation and uplink subcarrier assignment-based approach is investigated for cellular and D2D network to achieve improved EE. Various power controlling algorithms used for D2D communications have been explained in [126] [17]. Power and resource block allocation using game theory has been discussed in [127]. The author represents a cross-layer optimization scheme in conjunction with nash equilibrium under the accurate availability of channel state information (CSI) [128]. In [129], the author suggests a joint scheme of power control and resource allocation. The adoption of a two-layer method is involved, to achieve an optimal solution for energy efficient architecture. Table XIV provides a complete survey of D2D in terms of energy efficiency.

2) Mm Wave: For next-generation networks, Millimeter waves are the most promising approach. For delivering fast multimedia services, high-quality audio-video, and real-time services, a large amount of bandwidth is required. Due to a large number of consumers, the traffic congestion problem in the network will be arisen by 2030. To solve this problem the fifth generation network uses millimeter wavelength. The signals are operating between 30 and 300 GHz and being shifted to a higher spectrum. On the other hand, this technology has a large penetration loss, high signal attenuation and extremely susceptible to blockage effects which lead to the limited coverage area. Spatial multiplexing technique is used

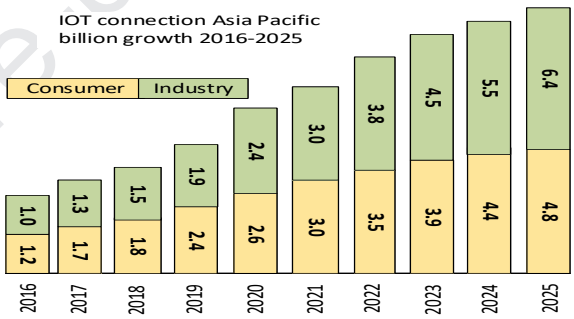


Fig. 26: Growth rate of IoT in Asia Pacific

for transmission and reception because the wavelength is very small. To obtain the high capacity, fast data rate and reduction in energy consumption, an array with directional transmission between MS and BS is used in this technique [130] . Table XV provides a complete survey of mm wave technique in terms of energy efficiency.

3) Terahertz communication band: In the upcoming decade, it is expected that the data rates will be increased up to 12 Gbps or more. This requirement of a high data rate cannot be fulfilled by a licensed band spectrum [137]. In addition to this mmWave technology alone cannot realize the need for a high data rate since there is only 9 GHz bandwidth for data communication. Free Space Optical (FSO) communication systems may be one of the solutions with minimum operating power for this demand but it requires direct line of sight (LOS) communication between the transmitter and receiver, which makes it impractical for mobile communication. So the most promising solution to achieve the demand of high data rate is Terahertz (THz) communication band which is least considered frequency region, where the data can be transmitted over non-line-of-sight (NLOS) links. Terahertz Band com-

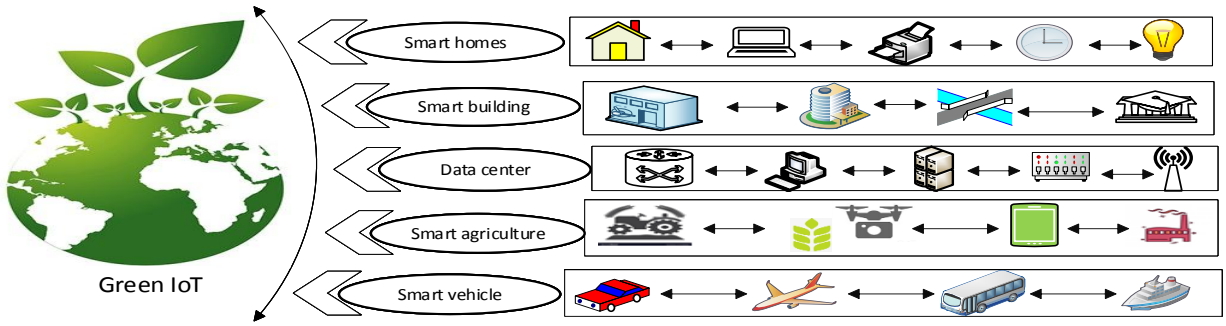


Fig. 27: Green IoT Architecture

TABLE XV: Comparison of existing surveys of Mm wave in terms of energy efficiency

Proposal	Key Points	Consideration	Model	Reference
Deployment of optimal BS is investigated	To reduce interference of two BSs, precoding between multi-BS and single BS coordination technique and different antenna angles are considered.	Attenuation coefficients and packet loss ratio	Path loss model	[83]
Non-line of sight and Line of sight link model investigated.	Green metrics, power efficiency, network latency,and area energy efficiency are calculated.	73 GHz frequency band has better SE and 28 GHz frequency band has better EE.	LOS and NLOS model	[131]
Downlink power control algorithms are investigated to enhance the EE	Path loss is calculated for different scenario.	licensed and unlicensed both applications	Large-scale, small scale path loss models and LOS probability model	[132]
Adaptive power control method is investigated	Interference between cellular and D2D user is calculated.	Radio resource allocation and management in wireless networks.	Power allocation mechanism	[133]
Implementation of full duplex relaying	Circuit power consumption and relay interference elimination.	Energy efficiency oriented solutions.	Joint beamforming	[134]
Beam-alignment protocol based energy-efficient architecture deployment.	For higher range, fractional search decreases power expenditure by 2dB with respect to bisection.	optimality in fractional search method	Markov decision model	[135]
Dynamic BS sleep control and RF chain activation are investigated.	Considering centralized and distributed both schemes for designing traffic aware dynamic scheduling architecture.	property of hybrid precoding	Greedy algorithm	[136]

munication will improve the spectrum scarcity and capacity of wireless communication networks. In the upcoming future THz bands are predicted for the development of various applications in both nanoscale communication and classical networking field [138]. The frequency region above and below the THz band is the infrared and microwaves. The THz frequency band (between 0.1 THz and 10 THz) will provide ultra-high-speed data transmission within the areas up to 10 m (PAN). The necessity to deal with this rapid progression of wireless services and related applications has required a large research activity that explores the optimal options for designing of user-oriented context-aware next generation (5G) wireless communication network. The 5G network is not only expected to improve the performance of information and communication devices but also capable to solve many pervasive issues of this field. The several applications related to 5G green communications in THz band are represented in Fig. 28.

4) Green IoT: The goal of the Internet of Things (IoT) is to link millions of users with each other using sensors like RF identification, and is the most evolving technology for making the foundation of the green CRNs. Large energy is required for exchanging information among millions of users. Fig. 26 represents the growth rate of IoT in the Asia Pacific in terms of consumers as well as industry. The goal of Green IoT is to enhance an EE of IoT devices and making the environment healthy. In [139] author projects that the usage of RFID will grow exponentially in the upcoming years. So, if this forecast is even nearby the reality, then a large amount of energy consumption will arise due to the more battery power will be demanded by all active RFIDs. So, to solve this problem, this work focuses on making the IoT technology green by applying numerous approaches. The IoT framework management involves identification, sensing,

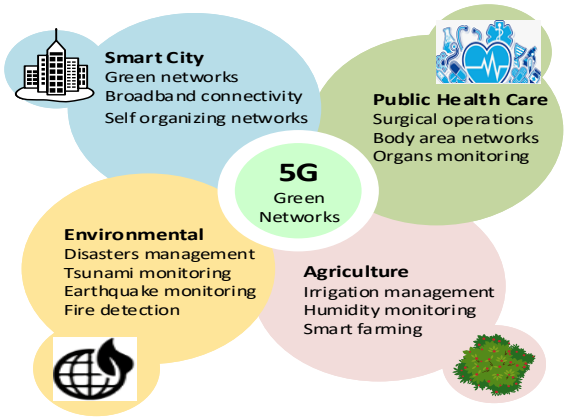


Fig. 28: Future applications related to 5G green networks

analyzing, communication, computation facilities and semantics. Identification is the initial process which guarantees that needed information sends to the accurate destination. Sensing works on collecting information and data, then sends to data centers. Considering different parameters and conditions, this collected data is then scrutinized for different services and operation purposes. The different parameters sensed by sensors are real time traffic, humidity, pressure, rate of pollution, temperature, etc. Combining different networks and objects for performing any task will comes under communication. Technology likes Bluetooth, Wi-Fi, etc are used to complete communication. The microprocessors, microcontrollers, embedded systems, FPGA, CPLD, ASIC, and various software applications perform the computation. Green IoT concept visualizes in Fig. 27. The figure shows that for the design of a smart city every element of a smart city like smart homes, smart agriculture, smart building, etc are connected by the IoT. People will use the internet for different purposes like traffic control, traffic monitoring, medical purpose, etc [140].

The main motive of green IoT is to make the environment green from its designing level to its implementation level. There are two ways to achieve this objective. The first is that the IoT will aid in eradicating the greenhouse effect produced by present applications and second is to optimize itself to reduce the greenhouse effect. Different scientific approaches for designing of green IoT are discussed in [141]. A framework is proposed for constructing the green IoT in [142]. Moreover, green cloud computing [143], green sensor networks [144], green RFIDs and green data-centers plays an important role in designing green IoT.

The IoT is developing technology for future green networks that have capability to change the way we observe IT industry. The green IoT approach is going to have an enormous influence on how we handle with convinced glitches in our everyday life and it is surely going to make our survives easier and better. There are five principles of achieving green IoT. Considering a small network size will save a large amount of energy and improved EE. Select only desired information by proper sensing and eliminate extra data also helpful for energy saving. The integration of different networks and implementation with proper policy making will lead towards green IoT and reduce carbon footprints. There should be a proper tradeoff between parameter is also very necessary. Fig. 29 describe the classification of green IoT techniques and Table XVI represents a detailed comparative survey of these techniques.

J. Green Cognitive Radio Networks Approach

According to the report of CISCO on data uses, in 2017−22, the telecommunication traffic originated by mobile subscribers is predictable to have nearly tenfold [155]. For new mobile subscribers and upcoming data-hungry devices, the availability of unused frequency bands is very low. Because, a report published by the FCC says that, nearly complete accessible bandwidth for wireless network communications has been almost allocated [156], [157]. Also, a study utters that nearly 75% of the assigned bandwidth is not properly used due to the static spectrum allocation scheme [158]. This generates an abundant gap between the existing requirement of additional bandwidth and presents an allocation scheme. The CRNs are the most promising, emerged and key enabling technology of future networks. This technology provides the freedom to utilize underutilized frequency bands by allowing Secondary Users (SUs) to access licensed frequency bands which are not used by Primary Users (PUs) opportunistically [159]. When the primary user comes back to the channel, the SU has to switch to another unoccupied channel belonging from the different network for consistent communication. The performance of SU has affected by the appearance of high prioritized PUs. Fig. 30 represents an idea behind green cognitive radio networks (GCRNs). The Fig.30a shows existing static spectrum allocation policy which fails in heavy traffic load situation and frequency bands are highly utilized. In Fig. 30a, the number of mobile subscribers are more in-network 'A' and cause heavy traffic load, whereas the other networks B, C, D and E have less traffic load with underutilized frequency bands. Fig. 30b, represents the difference between CR and non-CR technology. Fig. 30c represents an optimal utilization of available frequency bands (called white space) by using dynamic spectrum allocation policy. The mobile node can switch from one network to another network after making the decision using the cognitive property. A dynamic spectrum allocation policy requires key functions like dynamic spectrum environment, spectrum sensing, sharing and management respectively, in GCRN as shown in Fig. 30d.

It is very challenging to design a green CRN because it involves the optimization of dynamic spectrum allocation as well as optimal exploitation of green energy. The various functionalities are performed in the energy-efficient CR.

- Spectrum Sensing and Analysis- For designing the energy-efficient spectrum sensing approach the main task is to analyze the performance indicators to find out the presence of the primary user. Various types of spectrum sensing with energy efficiency related issues are represented in Fig. 31.

Journal Pre-proof

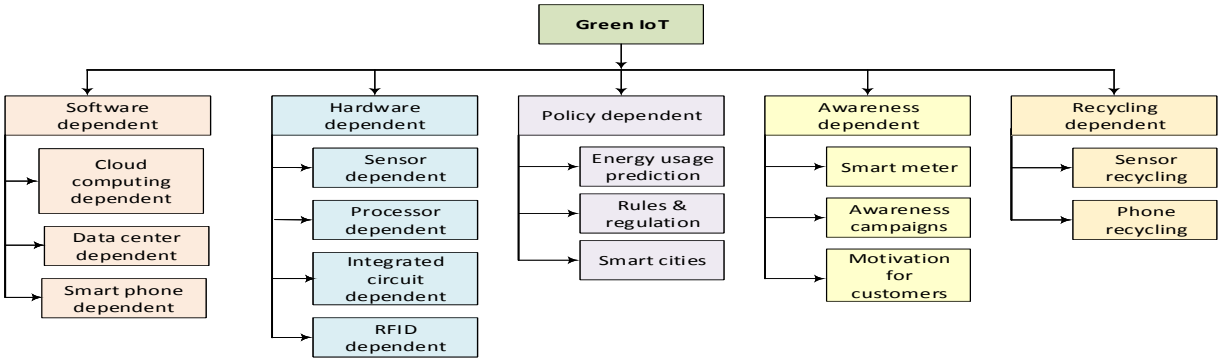


Fig. 29: Classification of Green IoT techniques

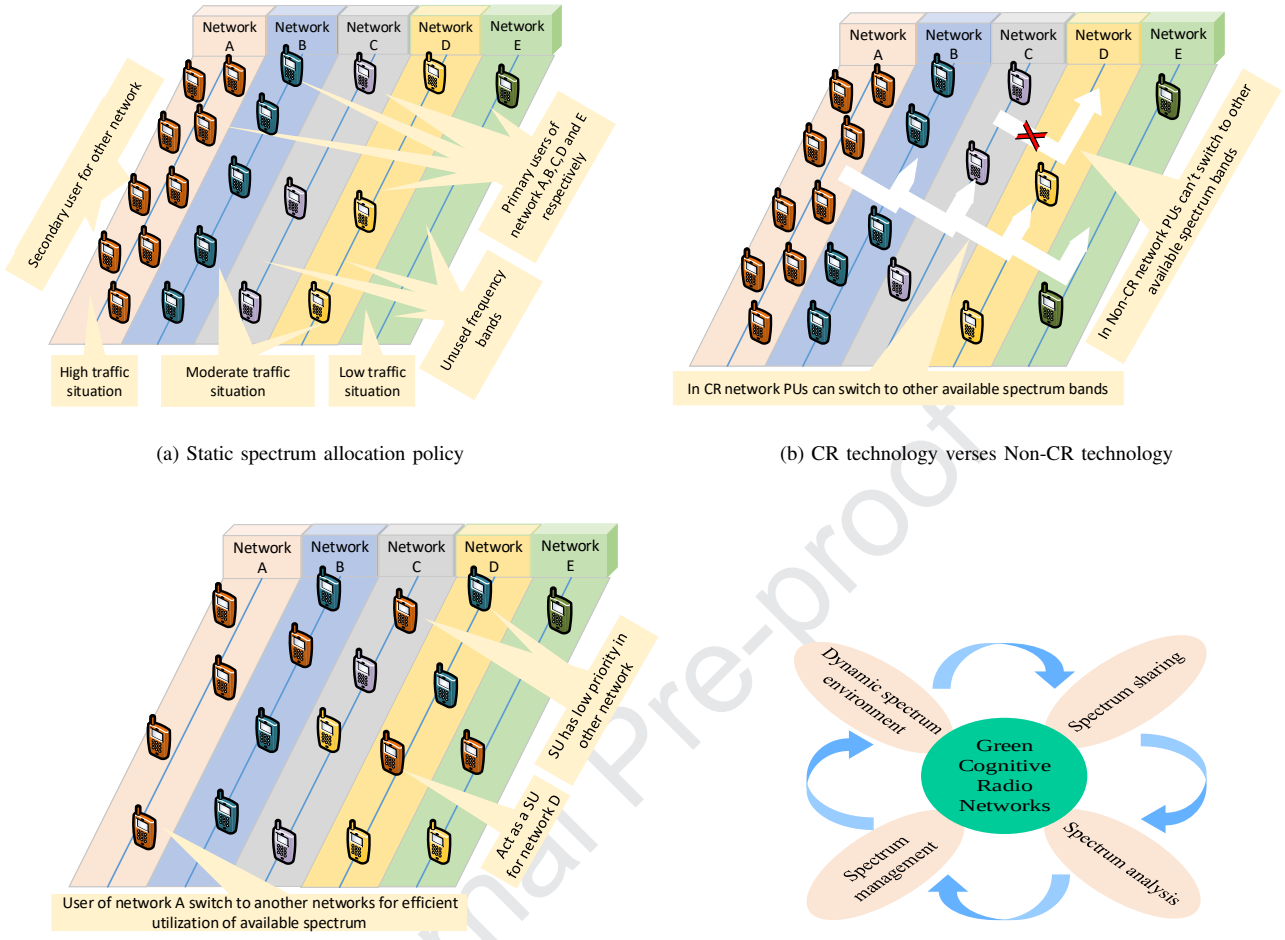
TABLE XVI: Comparative survey of green IoT techniques

Approach	Consider Technology	Energy saving process	Tradeoff	Real time implementation	Metric	Reference
Hardware dependent	Relay	Load balancing among different nodes	Communication delay	Possible	End to end delay	[145]
Software dependent	Sensors	Selective sensing	Privacy, energy overheads	Highly possible	Energy efficiency	[146]
Policy dependent	Information management	Information collected from various section of building to develop energy efficient guidelines.	Data security	Highly possible	Data efficiency	[147]
Software dependent	Cloud computing	To decrease data traffic dynamic packets are downloaded from access points.	QoS may be compromised	Possible	Packet transmission rate	[148]
Software dependent	Smart phones	Calculation of energy consumed by smart phone applications	Security and data privacy	Possible	Energy efficiency	[149]
Software dependent	Data centre	Workload sharing among physically separated data centers	Highly complex	Not possible	Data center efficiency	[150]
Hardware dependent	Integrated circuits	Inbuilt sensors on-chip are used to decrease network traffic load	High cost	Possible	Radio efficiency	[151]
Awareness dependent	Smart metering	Track various types of devices and evaluate the energy consumption rate.	Security and data privacy	Possible	Power efficiency, transmitted power	[152]
Policy dependent	Smart buildings	Guidelines and approaches to reduce energy consumption.	User dependency	Not possible	Energy per bit per area	[153]
Hardware dependent	RFID	Passive sensors are used for evaluation	Communication delays	Possible	Spectral efficiency	[154]

- Green Relaying in Cognitive Radio- In CRNs, authentic end-to-end data transmission among secondary users need a large amount of power, which leads to harmful interference to primary users. Relay technology is capable to boost the overall performance by saving energy. Because of the shorter transmission distance, the path loss is very low in this technology.
- Cooperative Cognitive Radio Networks- Cooperative communication is considered as the most promising solution to tackle the issue of multipath fading. This provides energy-efficient communication and also reduces the detection time due to the usage of spatial diversity. Cooperative communication protocols are working in two modes named as Decode-and-Forward and Amplify-and-Forward. The cooperative communications in CRNs, can be classified into two classes such as cooperation among secondary users and cooperation between the primary user and secondary users respectively.

Now the question is how do CRNs reduce power consumption and provide green communication? So, the answer is for wireless communication system bandwidth efficiency is always a pivotal concern and there exists a rich state of the art on this topic, focusing on improving bandwidth efficiency. The solution lies under Shannon's capacity formula which describes the trade-off between power and bandwidth. The capacity various logarithmically with power, but linearly with bandwidth. It reflects that to decrease power, larger bandwidth will be required or in other words, handle the spectrum dynamically and optimally. This is the principle of CR. Table XVII having a detailed survey of cognitive Radio technique in terms of energy saving.

A complete comprehensive survey of all techniques dis-



(a) Static spectrum allocation policy

(b) CR technology verses Non-CR technology

(c) Dynamic spectrum allocation policy

(d) Key function for Green CR networks

Fig. 30: Concept of Cognitive Radio

cussed in this paper is given in Table XVIII and continue in Table XIX.

IV. Challenges and Existing Practical Implementations of Green Communication

The framework to enhance the EE of telecommunication the system has a holistic view which considers possibly all facts and figures of an ecosystem for green communication. The several challenges related to green communication are listed as:-

- For estimating any energy efficient solution, the required energy for the manufacturing of a communication system (called embodied energy) has a very significant role.
- The surge of EE at a single stage of the communication system might be the reason for the decline at another stage of the system. For instance, the usage of a cognitive radio system may decrease the transmission energy by using proper bandwidth but drain additional power in the process of sensing the spectrum hole of the second network.
- There should be a respectable trade-off between EE and QoS. Like the aforementioned usage of BS sleep modes technique, more energy can be saved during off mode of some cells. But, it is compulsory to stipulate the cell waking time in advance as well otherwise, the QoS may degrade.
- There should be also a proper trade-off between flexibility and EE which depicts in Fig. 32.

The energy-efficient solution depends on certain guidelines, concepts, parameters, energy consumption models and considerations. The energy consumption of the telecommunication system is load-dependent and implementation-dependent. These are the challenges for modeling the energy- efficient system. The EE metric for a green telecommunication system is generally expressed in the form of performance per unit of energy. The performance usually measured in terms of efficiency and throughput for the energy-efficient communication system. Other key challenges for designing the green energy-efficient system are cross-layer adaptation, system reconfiguration, load balancing, and multi-domain scheduling. It is necessary to gauge EE from a new perspective for

TABLE XVII: Comparative survey of CR technique

Objective	Description	Energy saved	Scheme/Algorithm	Reference
Cultivating the EE of cellular networks	The author investigates the cooperation between BSs and television stations in sending UE relying on an opportunistic usage of the TV spectrum. The unoccupied TV spectral frequency band is utilized in an opportunistic manner.	41%	Explores the joint employment of cooperation techniques and cognition	[160]
For CRNs examine the EE with the relation of area spectral efficiency	A constant base station power consumption model is considered for both perfect and imperfect detection in the case of cognitive as well as primary users.	35%	Stochastic geometry approach	[161]
The author has described the method to attained maximum EE in a spectrum sharing atmosphere.	With consideration of EE, the trade-off between throughput and transmit power in a centralized CRNs is explored.	28%	Spectrum-sharing model	[162]
The effect of interference power, interference gain on energy consumption and optimal sensing time of the channel is described.	For multiband CRNs power allocation and optimal sensing time algorithm are proposed.	30%	Multiband sensing-based scheme	[163]
Research work is dedicated to energy-efficient optimum power allocation tactics for green CR in fading channels.	For both delay-insensitive and delay sensitive CR, EE improvement schemes are investigated with the consideration that the channels are suffered from fading and primary user coexists with a secondary users.	25%	Convex optimization methods and Dinkelbach's method	[164]

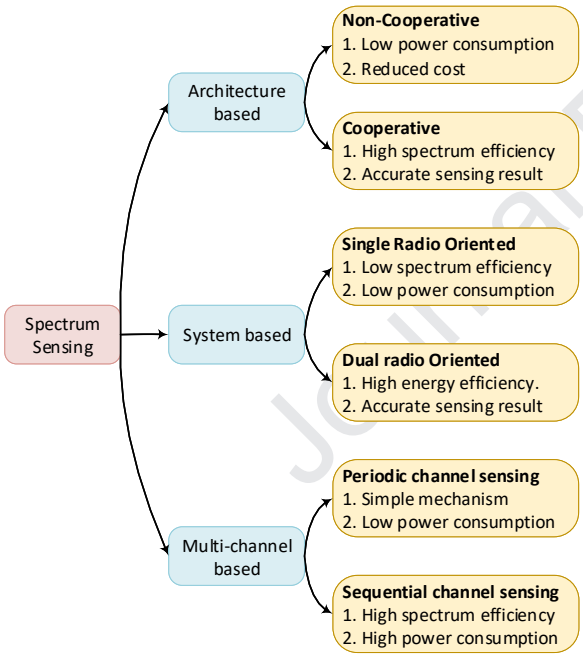


Fig. 31: Various spectrum sensing techniques

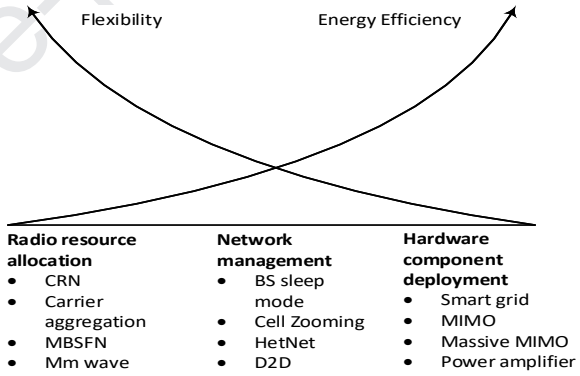


Fig. 32: Tradeoff between flexibility and EE

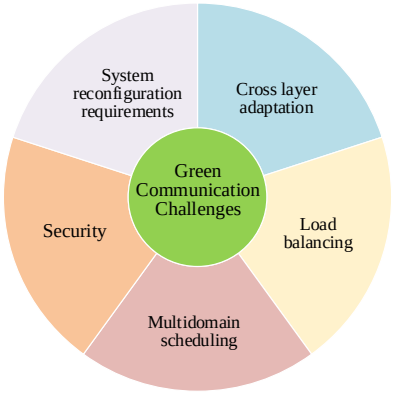


Fig. 33: Green communication related challenges

designing and developing novel energy-efficient architecture. The pathway represented in Fig. 6 to implement in the real world for making communication green is also having several challenges that need to be paid consideration too in this research area.

Several challenges, related to green communication for next generation networks are represented in Fig. 33.

There are some international projects with the aim of exploring possible ways to reduce CO_2 emission, transmitted power of telecommunication system, health effects, and EMR exposure are discussed in Table XX

Journal Pre-proof

TABLE XVIII: Assessment of the Energy-Efficient strategies

Approaches for Green communication	Techniques	Metrics for Green performance evaluation	Saved Energy/Reduced power	Parameter Consideration	Applied Domain	Merits	Demerits
Hardware Component Improvement	Doherty architecture and GaN based amplifier	EE, PA Efficiency	Up to 51%	Peak-to-average power ratio	At infrastructure level	Energy can be saved in a large amount	Less flexible
	Switch mode power amplifier	Energy Consumption Rate	70%	Linearity	At infrastructure level	A large amount of energy can be saved	For hardware replacement large cost required, some upper limits for hardware implementation
Energy Efficient Architecture	Hetnet	CapEx and OpEx		Interference, handoff	Always macro cells remain in active mode, while small cells will come in active mode when traffic increases.		Required high radio interference management
	1. Microcell-Macrocell	CapEx and OpEx	45%	Interference	At infrastructure level	Low implementation cost	Handoff management
	2. Picocell-Macrocell	CapEx and OpEx	62%	Handoff	At infrastructure level	Prolonged handset lifetime	Handoff management
	3. Femtocell-Macrocell	CapEx and OpEx	80%	Interference	At infrastructure level	Higer SNIR	Interference management
	4. Femtocells (urban area)-Marocell (Suburban area)-Portable femtocell (rural area)	CapEx and OpEx	Reduced by 83% to 87%	Handoff	At infrastructure level	Improvement in EE and low transmission power	Handoff management
	5. Femtocell (Highly populated urban area)-Picocell (Less populated urban area)-Microcell (suburban areas)-portable femtocells (rural areas)	CapEx and OpEx	79-80.2%	Interference	At infrastructure level	Low path loss	Interference management
	6. Macrocells (Fully coverage area)-femtocells, picocells, and microcells (border regions)	CapEx and OpEx	10%	Handoff	At infrastructure level	Improve coverage	Handoff management
	Cell Zooming	CapEx and OpEx		Problem of dead zone, coverage holes	BS coverage area continually changes with variation on the real time traffic		Maintain tradeoff between energy saving and performance
	1. Zoom out Active cell in Commercial area	CapEx and OpEx	30%		At infrastructure level	Inexpensive implementation and testing	Handoff management
	2. Zoom out Active cell in Residential area	CapEx and OpEx	38-40%		At infrastructure level	Improve throughput	Interference management
	3. Zoom out Active cell in uniformly distributed area	CapEx and OpEx	25-40%		At infrastructure level	Proper traffic load balancing	Handoff management
	4. Zoom out Active cell in Office area	CapEx and OpEx	45%		At infrastructure level	Improvement in EE	Interference management
Traffic based energy efficient	Base station ON/OFF	Sleep mode energy saving, Energy consumption index, performance indicator, Energy consumption gain		Battery life of UE, coverage	Mobile users transfer connection from off BS to on BS	Energy savings depend on the time period over which each of the BSs is switched off	Maintain trade-off between energy saving and QoS, less accurate in modelling
	1.Distance based evaluation		29%				
	2.Real time Traffic based evaluation.		30-35%				
Time based energy efficient	1. Unicast	Electricity cost, state of battery	38.5-45%	Proper synchronization	LTE BS	High energy reliability, deactivate unused subframe	Degrade battery life of UE

Journal Pre-proof

TABLE XIX: Assessment of the Energy-Efficient strategies

	2.MBSFN		56%				
	3.DTX		80%				
	4.Optimize subframes in each frame		90%				
Frequency based energy efficient	Carrie Aggregation	Latency, gain and axial ratio	Power reduces up to 65%	Compatibility, Complexity	4G LTE	Power reduce by double factor band	Because of early hand-off, radio link failure increases, Difficult for two networks to work together
Momentous Approach	Renewable energy	Cost efficiency	10%, energy savings depend on BSs installation and location, speed of wind	BSs location, other natural energy resources	At infrastructure level	Less expensive, long-term solution for off grid	High replacement cost, for on grid BS limited gain
	Smart Grid	Cost of electricity, on-grid energy consumption	Up to 30%	Resource management, QoS	At infrastructure level	Energy savings through reducing consumption, Improved environment	Required proper infrastructure
Space based energy efficient	MIMO	Spectral efficiency, Energy efficiency	55%	Effect on the UE	At infrastructure level	Improved capacity and spectral efficiency	High circuit power, reduces battery lifetime
	Massive MIMO	Transmission Power, Energy consumption	Depends on number of switched off antenna	Spatial multiplexing, space time block coding	At infrastructure level	Large throughput	Complex channel modeling, more loop interference and training symbols
EM Radiation based energy efficient	SAR Shielding	Power density, Whole body averages SAR	A large amount of energy can be saved	Signals coming from the sensor array	At UE level	Save body organs and tissues from harmful waves	Enhancement depends on a special material and design
	Beamforming	SAR exposure	A large amount of energy saved	Signals coming from the sensor array	At antenna level	Improve system coverage and capacity, improved SINR in the desired direction	High hardware complexity. Advanced processing DSP chip is required. High Battery drain due to large power requirement
Green Cognitive Radio Network		Throughput, network energy cost	Upto 50% by proper spectrum management	Spectrum hole, unutilized bandwidth	Performance depends on real time load distribution information of whole network	User oriented, low cost, various application	Performance depends on spectrum unavailability, Delay is involved due to sensing operation
Futuristic Approach	D2D method	Average sum rate, BS power consumption	15-40% Depends on the number of intermediate devices	Switch off the devices for a deterministic traffic	Device on off depends upon the distance between the transmitter and receiver	Low transmitter and circuit power	Packet encoding and decoding is required, Interference management, D2D mobility management, Synchronization among devices
	Mm wave	Data rate, outage probability, coverage probability	Up to 20%	Privacy, security	At antenna level	Cost-effective	High circuit power and signaling overhead required atmospheric loss, pathloss and shadowing
	IoT	Energy efficiency, end to end delay, weighted throughput	Up to 30%	Compatibility, complexity	At infrastructure level	Reduce human effort, improve security	Complex network, device recognition and difficult to locating in IoT

V. Discussion and Forward Way Research Direction

Various energy-efficient approaches examined in previous sections are valuable to design energy-efficient green communication system and network. Integrating all these approaches into a combined approach can lead the revolt in the field of green communication. The next-generation networks are anticipated to realize the necessities of the mobile subscribers as well as deliver a solution for energy-efficient green communication.

Green communication networks have been broadly subjugated in bestowing a plentiful range of real-time multimedia applications. The Fig.34 demonstrates the extensive summary of upcoming future networking technologies likes Software Defined Networks, Supermassive MIMO, Terahertz Communication, Ultra-Dense Networks, IoT Communication, Quantum Communication, Beamforming and NOMA. However, the future research direction and issues related to green communication networks are also summarized in Fig.34. On the basis of studied literature the future research directions can be summarized as:

- In supermassive MIMO, usage of multiple antennas requires very high power consumption and a great amount of overhead. This problem can be overcome using the spatial diversity concept.
- The EE of terahertz communication in multiple cell scenario requires to be explored to eradicate interference-effect.
- The consumption of high power in communication system circuits is a critical issue in the usage of renewable energy for designing a green communication system.
- The cost factor involved in the practical realization of IoT technology is another significant research subject for the future.
- The power control strategy in hardware and efficient algorithms in the software of the IoT system can be used to achieve optimum SNR and reduce interference.

To realize the green communication network environment, an intelligent advanced energy-efficient technique requires to be developed. Different energy-efficient techniques have been immensely deliberated in this survey. However, there are still new research areas for more effort to be achieved in this field. This part of the section delivers future research directions that entail the consideration of the research community. These new directions are discussed as:

1) Cooperative Heterogeneous Network (HetNet) Architecture: The upcoming generation of networks is likely to provision incorporation of various networks having diverse services and protocols. So, vertical handover, management of interference, enhancement in network capacity with the attention of energy efficient communication system is an important research topic. The tradeoff between spectral and energy efficiency and a tradeoff between network coverage and QoS in HetNet network also requires more investigation.

2) Green Networks: Energy-efficient approaches are the main prerequisite for the designing of GCRN. Thus, a new research area to minimize the EMR effect by the green networks is desired to be investigated. Two key research issues associated with GCRN are (1) decrease in consumption of power and energy overhead instigated by communication among mobile nodes and (2) optimization of dynamic spectrum allocation techniques to improve EE without distressing QoE of mobile subscribers.

3) Experimental Testbeds for Green Networks: For analyzing an impact of suggested energy-efficient approach for green networks, it is vital to test these approaches on realistic testbeds. In this way, the hardware realization issues can be identified and would be able to check the viability of these proposed approaches. The absence of mature green networks modules in prevailing simulators offer new research direction to enterprise an experimental testbed to provision green networks.

4) Security: Auspicious of protected communication in green networks is a significant design objective for consistent advance next-generation wireless networks. It is a challenge to design an energy efficient approach for green communication to handle the EMR and GHG effect jointly with the utilization of RER for saving energy. In the available literature, energy-efficient approaches are focusing on the security of either network or mobile nodes. Secured green communication for both networks and mobile nodes, along with energy-efficient encrypted technique is required in future wireless communication systems.

5) Green networks Model for LTE/Wi-Fi coexistence: LTE network supports the modern wireless networks like carrier aggregation, massive MIMO, HetNet, D2D network with multiple cells of small size. The adaptive energy-efficient approach designed for green networks would be a promising technology in such advanced wireless networks. For example, HetNet comprises multiple cells of different sizes (termed as femto, pico, and microcells). So, the management of interference causes more complexity in networks. The coexistence of two different networks with combining their capabilities and energy- saving proposals would help in the formation of green networks, which is still an energetic research direction and will receive further attention.

Fig. 35 represents the advanced technologies i.e. next step towards futuristic wireless networking. These technologies are blockchain, machine learning, artificial intelligence, big data, advanced beamforming.

VI. Conclusion

This paper scrutinized the trends of the telecommunication system in the last decade which visualize a move towards following energy- efficient green communication for the purpose of designing next-generation networks. For each energy efficient technique, an outline of the existing state of the art research, merits, demerits, open issues, challenges, and possible forward way research direction has been investigated.

Lessons Learned by the survey

- *Lesson 1*: Select an appropriate EE metric for performance measurement at facility, component, access point and network level. Green mobile communication is an important chunk of corporate social obligation which

Journal Pre-proof

TABLE XX: Ongoing projects

International projects/ Consortium on green communication	Web Reference	Approach	Participated country/ infrastructure vendors	Research outcome
FANTASTIC	http://fantastic5g.eu/	Develop a framework based on AI components	European countries	Present a multi-service air interface for enhancing reliability and energy efficiency
GREENNETS	http://www.greennets.eu/	For designing energy efficient architecture advanced tools and software platforms introduce in the market	Germany, Poland, Lithuania	Decrease CO_2 footprint, power consumption and exposure to electromagnetic smog.
TRIBE	http://tribe-h2020.eu	To improve social behavior of users towards EE	European Commission	Reduce energy consumption and emissions between 15% and 30%
CogNet	https://5g-ppp.eu/cognet/	Applying machine learning method for network management in 5G	European Commission and European ICT industry	Design an intelligent system of visions and action, supporting green energy-efficient communication
EUFORIE (European Futures for Energy Efficiency)	http://euforie-h2020.eu	Works on multi-scale analysis for enhancing EE	Finland	Achieve secure, clean, green and energy efficient environment
5 Green	https://wireless.kth.se/5green/	Physical layer signaling, traffic adaptive solutions and higher layer protocols used for minimizing energy consumption	Finland, Italy	Achieve eco-friendly energy efficient 5G networks
GreenTouch	http://www.greentouch.org	Design energy efficient architecture and work on network specifications	-	Target to achieve 98% energy reduction till 2020
5GPPP	https://5g-ppp.eu/	Design energy efficient architecture with converging capabilities for both fixed and mobile accesses	Europe	Saving up to 90% of energy per service provided
LEXNET (Low EMF Exposure Future Networks)	http://www.lexnet.fr/	Focus on cross layer techniques to reduce the EM radiation exposure	European countries	Target to achieve reduced EM radiation exposure up to 50% without compromising QoE and QoS
HORIZON 2020	https://ec.europa.eu/programmes /horizon2020/en/what-horizon-2020	Apply security mechanisms on physical layer	-	Focus on privacy and security issues of 5G architecture
mm MAGIC	https://5g-mmmagic.eu/project/	Consider 6-100 GHz range, novel frame structure, cooperative and adaptive beamforming for designing and addressing the challenges of mobile radio access technology	Key European operators (Telefonica, Orange) and Nokia Samsung, Lucent Ericsson, Alcatel-Huawei, Intel)	Cost and energy-efficient solutions for 5G millimeter wave deployments
The International EMF project	https://5g-mmmagic.eu/https://www.who.int/peh-emf/project/en/	Study the static and time varying magnetic and electric fields between 0-300 GHz frequency range and adverse effects of these on health.	WHO member states	To minimize health hazards because of EM radiation exposure
Green Wireless Communication	https://sri-uq.kaust.edu.sa/Pages /greenwireless.aspx	Progress and performance evaluation of novel wireless channel estimation methods, transceiver design optimization for BS sleeping strategy.	Kingdom of Saudi Arabia	To minimize energy consumption of NGNs.
Wireless@MIT Center	http://wireless.csail.mit .edu/Energy	Work on low-power handsets, energy-improving sensors and wireless medical devices.	MIT Computer Science and Artificial Intelligence Laboratory, Cambridge	Analyze software and hardware components for energy-efficient mobile systems.

endeavors to minimize greenhouse gases and carbon footprint in green ICT services for customers. The government should propose protocols, rules, and regulations to certify the service provider as a green service provider. So, mobile operators have to combat global warming by decreasing their CO_2 emissions.

- *Lesson 2*: The joint application of the BS sleep method with hybrid beamforming provides better cost reduction and flexibility. The increased demand for mobile data and the internet needs a paradigm shift towards broader bandwidth accessibility in GCRN and mm-Wave frequency band. Massive MIMO contains a larger number of antennas which minimize transmitted power. The main attention lies on the utilization of relay for the future green networks as well as small size cells and their role in the formation of energy-efficient communication systems.

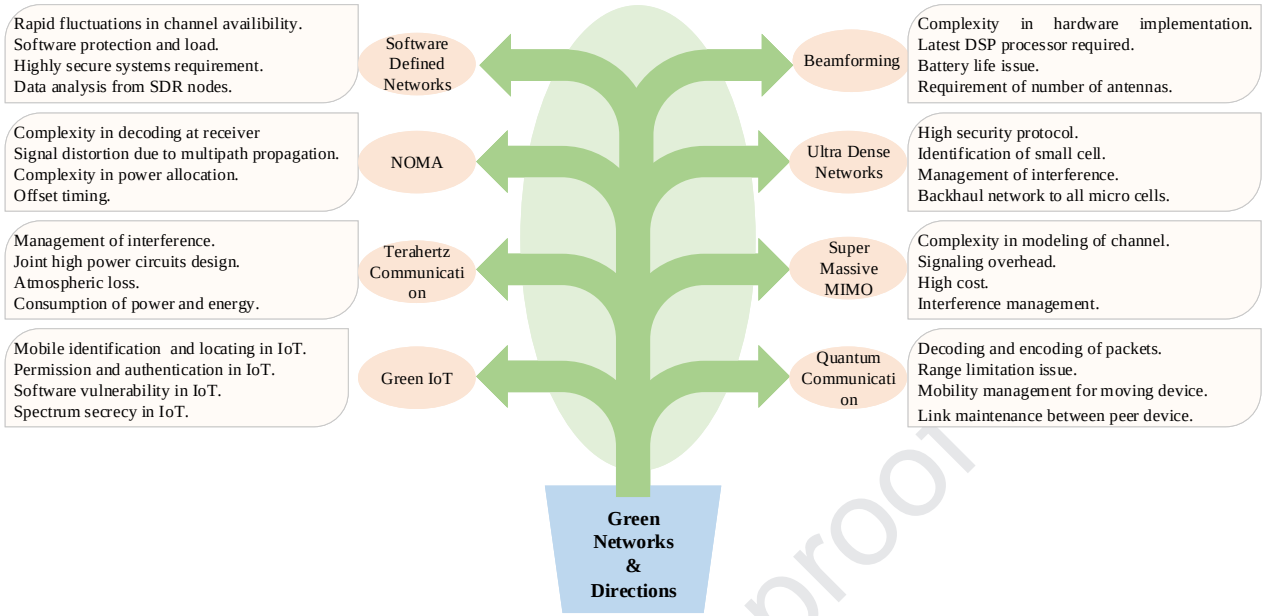


Fig. 34: Green communication future networks and related research directions

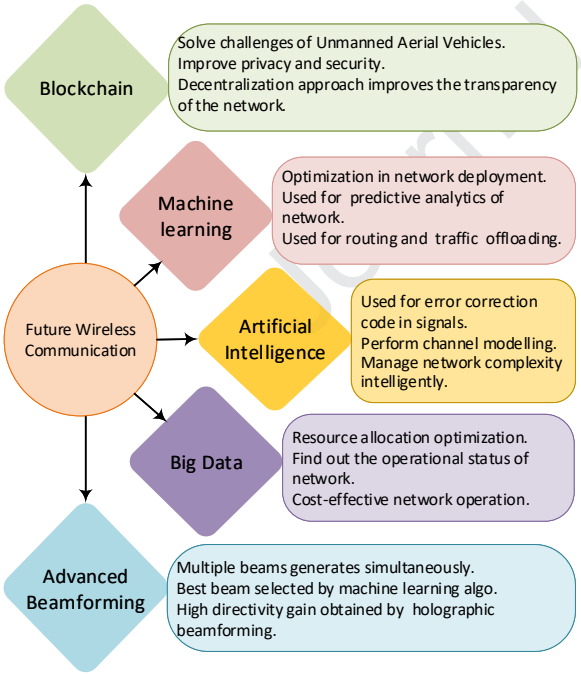


Fig. 35: Advance communication technologies i.e. Next step towards future wireless networks

The green cognitive radio networks have a great potential to handle inevitable demands of extra spectrum in next generation wireless networks. But have some limitation in applying methods in existing communication systems. The most prominent limitations of the current methods are such as practical hardware realization, interference and handoff management, proper traffic load balancing among different candidate networks, non-zero latency in real-time services and applications, appropriate densification limit of a network (how much density of the network is adequate?), lack of self-optimizing and self-configuring network devices, complex circuitry, requirements of system reconfiguration, cost factor and privacy and security management.

- *Lesson 3*: ICT attempts to achieve the target to reduce 13% of the global GHG emissions by 2035 with the help of– i) Development of miscellaneous IoT applications to support efficient energy management ii) Advancement of multimedia applications and mobile broadband and iii) Designing and development of system software by introducing functionality of the network, to analyze the behavior of the network's traffic and improving spectral efficiency gains. To pursue energy-efficient green communication, the impact of the mobile life cycle on the carbon footprint is discussed.

Acknowledgments – The authors would like to thank the Women Scientists Scheme-A under Department of Science and Technology Government of India for its financial support of this work under File No: SR/WOS-A/ET-154/2017. The authors also would like to thank the anonymous reviewers for their feedback.

REFERENCES

[1] (01-11-2018) Internet user population will be increased. [Online]. Available: https://www.cisco.com

[2] (01-02-2019) Growth rate of different technologies. [Online]. Available: https://www.gsma.com/r/mobileeconomy/
[3] (11-11-2018) Growth rate of base station. [Online]. Available: https://www.telecomlead.com
[4] (01-06-2008) gas emmission. [Online]. Available: https://www.theclimategroup.org
[5] I. Humar, X. Ge, L. Xiang, M. Jo, M. Chen, and J. Zhang, "Rethinking energy efficiency models of cellular networks with embodied energy," *IEEE network*, vol. 25, no. 2, pp. 40–49, 2011.
[6] S. Lambert, W. Van Heddeghem, W. Vereecken, B. Lannoo, D. Colle, and M. Pickavet, "Worldwide electricity consumption of communication networks," *Optics express*, vol. 20, no. 26, pp. B513–B524, 2012.
[7] (01-02-2019) Population of mobile subscriber. [Online]. Available: https://www.gsma.com/r/mobileeconomy/asiapacific/
[8] (01-02-2019) Population of mobile subscriber in china. [Online]. Available: https://www.gsma.com/r/mobileeconomy/china/
[9] (01-02-2019) Population of mobile subscriber global. [Online]. Available: https://www.gsma.com
[10] (01-06-2019) addition of mobile users globally. [Online]. Available: https://www.ericsson.com
[11] (11-04-2019) 100s of birds drop from the sky at the hague in netherlands. [Online]. Available: http://www.radiationdangers.com
[12] (03-11-2018) Negetive impact on body organs. [Online]. Available: https://gizadeathstar.com
[13] S. S. Sandhu, A. Rawal, P. Kaur, and N. Gupta, "Major components associated with green networking in information communication technology systems," in *2012 International Conference on Computing, Communication and Applications*. IEEE, 2012, pp. 1–6.
[14] Y. Kremenetskaya, S. Markov, and S. Morozova, "Application of hybrid millimetre wave technology for green wireless communications," in *2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT)*. IEEE, 2018, pp. 649–652.
[15] R. Mahapatra, Y. Nijsure, G. Kaddoum, N. U. Hassan, and C. Yuen, "Energy efficiency tradeoff mechanism towards wireless green communication: A survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 686–705, 2016.
[16] D. Feng, C. Jiang, G. Lim, L. J. Cimini, G. Feng, and G. Y. Li, "A survey of energy-efficient wireless communications," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 167–178, 2013.
[17] P. Gandotra and R. K. Jha, "A survey on green communication and security challenges in 5g wireless communication networks," *Journal of Network and Computer Applications*, vol. 96, pp. 39–61, 2017.
[18] T. Chen, H. Kim, and Y. Yang, "Energy efficiency metrics for green wireless communications," in *2010 International Conference on Wireless Communications & Signal Processing (WCSP)*. IEEE, 2010, pp. 1–6.
[19] L. Suarez, L. Nuaymi, and J.-M. Bonnin, "An overview and classification of research approaches in green wireless networks," *Eurasip journal on wireless communications and networking*, vol. 2012, no. 1, p. 142, 2012.
[20] T. A. Khan, A. Yazdan, and R. W. Heath, "Optimization of power transfer efficiency and energy efficiency for wireless-powered systems with massive mimo," *IEEE Transactions on Wireless Communications*, vol. 17, no. 11, pp. 7159–7172, 2018.
[21] M. R. Mili, L. Musavian, K. A. Hamdi, and F. Marvasti, "How to increase energy efficiency in cognitive radio networks," *IEEE Transactions on Communications*, vol. 64, no. 5, pp. 1829–1843, 2016.
[22] Y. Jiang, Y. Zou, J. Ouyang, and J. Zhu, "Secrecy energy efficiency optimization for artificial noise aided physical-layer security in ofdm-based cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 12, pp. 11 858–11 872, 2018.
[23] J. Ouyang, M. Lin, Y. Zou, W.-P. Zhu, and D. Massicotte, "Secrecy energy efficiency maximization in cognitive radio networks," *IEEE Access*, vol. 5, pp. 2641–2650, 2017.
[24] C. Belady, A. Rawson, J. Pfleuger, and T. Cader, "Green grid data center power efficiency metrics: Pue and dcie," Technical report, Green Grid, Tech. Rep., 2008.
[25] R. C. Zoie, R. D. Mihaela, and S. Alexandru, "An analysis of the power usage effectiveness metric in data centers," in *2017 5th International Symposium on Electrical and Electronics Engineering (ISEEE)*. IEEE, 2017, pp. 1–6.
[26] H. Hamdoun, P. Loskot, T. OâĂŹFarrell, and J. He, "Survey and applications of standardized energy metrics to mobile networks," *Annals of telecommunications-annales des télécommunications*, vol. 67, no. 3-4, pp. 113–123, 2012.
[27] X. Wang, A. V. Vasilakos, M. Chen, Y. Liu, and T. T. Kwon, "A survey of green mobile networks: Opportunities and challenges," *Mobile Networks and Applications*, vol. 17, no. 1, pp. 4–20, 2012.
[28] A. A. Ibrahim, K. P. Kpochi, and E. J. Smith, "Energy consumption assessment of mobile cellular networks."
[29] P. Gandotra, R. K. Jha, and S. Jain, "Green communication in next generation cellular networks: A survey," *IEEE Access*, vol. 5, pp. 11 727–11 758, 2017.
[30] Y. Wu, Y. Chen, J. Tang, D. K. So, Z. Xu, I. Chih-Lin, P. Ferrand, J.-M. Gorce, C.-H. Tang, P.-R. Li *et al.*, "Green transmission technologies for balancing the energy efficiency and spectrum efficiency trade-off," *IEEE Communications Magazine*, vol. 52, no. 11, pp. 112–120, 2014.
[31] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green cellular networks: A survey, some research issues and challenges," *arXiv preprint arXiv:1108.5493*, 2011.
[32] Y. Zhang, Y. Xu, Y. Sun, Q. Wu, and K. Yao, "Energy efficiency of small cell networks: metrics, methods and market," *IEEE Access*, vol. 5, pp. 5965–5971, 2017.
[33] M. Parker and S. Walker, "Roadmapping ict: An absolute energy efficiency metric," *IEEE/OSA Journal of Optical communications and Networking*, vol. 3, no. 8, pp. A49–A58, 2011.
[34] J. Wu, Y. Zhang, M. Zukerman, and E. K.-N. Yung, "Energy-efficient base-stations sleep-mode techniques in green cellular networks: A survey," *IEEE communications surveys & tutorials*, vol. 17, no. 2, pp. 803–826, 2015.
[35] V. Foteinos, D. Kelaidonis, G. Poulios, P. Vlacheas, V. Stavroulaki, and P. Demestichas, "Cognitive management for the internet of things: A framework for enabling autonomous applications," *IEEE vehicular technology magazine*, vol. 8, no. 4, pp. 90–99, 2013.
[36] F. H. Raab, "Efficiency of doherty rf power-amplifier systems," *IEEE transactions on Broadcasting*, no. 3, pp. 77–83, 1987.
[37] H. Claussen, L. T. Ho, and F. Pivit, "Effects of joint macrocell and residential picocell deployment on the network energy efficiency," in *2008 IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*. IEEE, 2008, pp. 1–6.
[38] F. Ghannouchi, M. Ebrahimi, and M. Helaoui, "Inverse class f power amplifier for wimax applications with 74% efficiency at 2.45 ghz," in *2009 IEEE International Conference on Communications Workshops*. IEEE, 2009, pp. 1–5.
[39] B. Berglund, J. Johansson, and T. Lejon, "High efficiency power amplifiers," *Ericsson Review*, vol. 83, no. 3, pp. 92–96, 2006.
[40] M. H. Alsharif, R. Nordin, and M. Ismail, "A review on intelligent base stations cooperation management techniques for greener lte cellular networks," *Journal of Communications*, vol. 9, no. 12, pp. 937–945, 2014.
[41] A. D. Gandhi and M. E. Newbury, "Evaluation of the energy efficiency metrics for wireless networks," *Bell Labs Technical Journal*, vol. 16, no. 1, pp. 207–215, 2011.
[42] D. Rapone, D. Sabella, and M. Fodrini, "Energy efficiency solutions for the mobile network evolution towards 5g: an operator perspective," in *2015 Sustainable Internet and ICT for Sustainability (SustainIT)*. IEEE, 2015, pp. 1–9.
[43] E. Chavarria-Reyes, I. F. Akyildiz, and E. Fadel, "Energy consumption analysis and minimization in multi-layer heterogeneous wireless systems," *IEEE Transactions on Mobile Computing*, vol. 14, no. 12, pp. 2474–2487, 2015.
[44] Y. Lv, H. Zhang, and S. Xueming, "Analysis of base stations deployment on power saving for heterogeneous network," in *2017 IEEE 17th International Conference on Communication Technology (ICCT)*. IEEE, 2017, pp. 1439–1444.
[45] I. Ashraf, L. T. Ho, and H. Claussen, "Improving energy efficiency of femtocell base stations via user activity detection," in *2010 IEEE Wireless Communication and Networking Conference*. IEEE, 2010, pp. 1–5.
[46] F. Cao and Z. Fan, "The tradeoff between energy efficiency and system performance of femtocell deployment," in *2010 7th International Symposium on Wireless Communication Systems*. IEEE, 2010, pp. 315–319.
[47] F. Richter, A. J. Fehske, and G. P. Fettweis, "Energy efficiency aspects of base station deployment strategies for cellular networks," in *2009 IEEE 70th Vehicular Technology Conference Fall*. IEEE, 2009, pp. 1–5.
[48] F. Richter and G. Fettweis, "Cellular mobile network densification utilizing micro base stations," in *2010 IEEE International Conference on Communications*. IEEE, 2010, pp. 1–6.
[49] F. Richter, A. J. Fehske, P. Marsch, and G. P. Fettweis, "Traffic demand and energy efficiency in heterogeneous cellular mobile radio networks,"

Journal Pre-proof

in *2010 IEEE 71st Vehicular Technology Conference*. IEEE, 2010, pp. 1–6.
[50] X. Xu, C. Yuan, W. Chen, X. Tao, and Y. Sun, "Adaptive cell zooming and sleeping for green heterogeneous ultradense networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 2, pp. 1612–1621, 2018.
[51] I. G. Karnila, H. Mahmudah, O. Puspitorini, A. Wijayanti, and N. A. Siswandari, "Analysis of cell zooming technique using adaptive traffic load balancing method on macrocell base station," in *2018 International Electronics Symposium on Engineering Technology and Applications (IES-ETA)*. IEEE, 2018, pp. 157–162.
[52] R. Balasubramaniam, "Cell zooming techniques for power efficient base station operation," Ph.D. dissertation, San Diego State University, 2012.
[53] R. I. Yuanita, O. Puspitorini, and H. Mahmudah, "Analysis technique of cell zooming using lowerto-higher scheme on centralized algorithm towards green cellular network in surabaya," in *2016 International Electronics Symposium (IES)*. IEEE, 2016, pp. 132–137.
[54] Z. Niu, "Tango: Traffic-aware network planning and green operation," *IEEE Wireless Communications*, vol. 18, no. 5, pp. 25–29, 2011.
[55] J. Lorincz, A. Capone, and D. Begusic, "Impact of service rates and base station switching granularity on energy consumption of cellular networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, no. 1, p. 342, 2012.
[56] X. Guo, Z. Niu, S. Zhou, and P. Kumar, "Delay-constrained energy-optimal base station sleeping control," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1073–1085, 2016.
[57] J. Gong, J. S. Thompson, S. Zhou, and Z. Niu, "Base station sleeping and resource allocation in renewable energy powered cellular networks," *IEEE Transactions on Communications*, vol. 62, no. 11, pp. 3801–3813, 2014.
[58] C. Liu, B. Natarajan, and H. Xia, "Small cell base station sleep strategies for energy efficiency," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 3, pp. 1652–1661, 2016.
[59] M. F. Hossain, K. S. Munasinghe, and A. Jamalipour, "A protocooperation-based sleep-wake architecture for next generation green cellular access networks," in *2010 4th International Conference on Signal Processing and Communication Systems*. IEEE, 2010, pp. 1–8.
[60] A. Bousia, A. Antonopoulos, L. Alonso, and C. Verikoukis, ""green" distance-aware base station sleeping algorithm in lte-advanced," in *2012 IEEE International Conference on Communications (ICC)*. IEEE, 2012, pp. 1347–1351.
[61] M. A. Marsan, L. Chiaraviglio, D. Ciullo, and M. Meo, "Switch-off transients in cellular access networks with sleep modes," in *2011 IEEE international conference on communications workshops (ICC)*. IEEE, 2011, pp. 1–6.
[62] M. H. Alsharif, R. Nordin, and M. Ismail, "Classification, recent advances and research challenges in energy efficient cellular networks," *Wireless Personal Communications*, vol. 77, no. 2, pp. 1249–1269, 2014.
[63] T. Chen, Y. Yang, H. Zhang, H. Kim, and K. Horneman, "Network energy saving technologies for green wireless access networks," *IEEE Wireless Communications*, vol. 18, no. 5, pp. 30–38, 2011.
[64] (16-11-2018) Extended cell dtx. [Online]. Available: http://www.3gpp.org
[65] M. H. Alsharif, R. Nordin, and M. Ismail, "Survey of green radio communications networks: techniques and recent advances," *Journal of computer networks and communications*, vol. 2013, 2013.
[66] A. Ghandri, Y. Boujelben, and M. B. Jemaa, "Dynamic mbsfn subframe allocation algorithm for bursty video traffic in lte-advanced network," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2018, pp. 1–6.
[67] K. Kanwal, G. A. Safdar, M. Ur-Rehman, and X. Yang, "Energy management in lte networks," *IEEE Access*, vol. 5, pp. 4264–4284, 2017.
[68] D. Migliorini, G. Stea, M. Caretti, and D. Sabella, "Power-aware allocation of mbsfn subframes using discontinuous cell transmission in lte systems," in *2013 IEEE 78th Vehicular Technology Conference (VTC Fall)*. IEEE, 2013, pp. 1–5.
[69] (16-11-2018) Energy saving techniques for lte. [Online]. Available: http://www.3gpp.org
[70] I. F. Akyildiz, D. M. Gutierrez-Estevez, R. Balakrishnan, and E. Chavarria-Reyes, "Lte-advanced and the evolution to beyond 4g (b4g) systems," *Physical Communication*, vol. 10, pp. 31–60, 2014.
[71] (16-11-2018) Radio regulations, international telecommunication union std. [Online]. Available: http://www.itu.int
[72] S. Rostami and K. Arshad, "A novel spectrum-aggregating technique in cognitive radio networks," in *International conference on Advance Computing and Communication Systems, Decemmber*, 2013.
[73] B. Barakat and K. Arshad, "Energy efficient carrier aggregation for lte-advanced," in *2015 IEEE 8th GCC Conference & Exhibition*. IEEE, 2015, pp. 1–5.
[74] Z. Shen, A. Papasakellariou, J. Montojo, D. Gerstenberger, and F. Xu, "Overview of 3gpp lte-advanced carrier aggregation for 4g wireless communications," *IEEE Communications Magazine*, vol. 50, no. 2, pp. 122–130, 2012.
[75] H. A. H. Hassan, L. Nuaymi, and A. Pelov, "Renewable energy in cellular networks: A survey," in *2013 IEEE online conference on green communications (OnlineGreenComm)*. IEEE, 2013, pp. 1–7.
[76] F. Ahmed, M. Naeem, W. Ejaz, M. Iqbal, A. Anpalagan, and H. Kim, "Renewable energy assisted traffic aware cellular base station energy cooperation," *Energies*, vol. 11, no. 1, p. 99, 2018.
[77] M. Alsharif and J. Kim, "Hybrid off-grid spv/wtg power system for remote cellular base stations towards green and sustainable cellular networks in south korea," *Energies*, vol. 10, no. 1, p. 9, 2017.
[78] Y. Zhang, M. Meo, R. Gerboni, and M. A. Marsan, "Minimum cost solar power systems for lte macro base stations," *Computer Networks*, vol. 112, pp. 12–23, 2017.
[79] M. Alsharif, J. Kim, and J. Kim, "Green and sustainable cellular base stations: An overview and future research directions," *Energies*, vol. 10, no. 5, p. 587, 2017.
[80] D. Niyato, X. Lu, and P. Wang, "Adaptive power management for wireless base stations in a smart grid environment," *IEEE Wireless Communications*, vol. 19, no. 6, pp. 44–51, 2012.
[81] H. A. H. Hassan, A. Pelov, and L. Nuaymi, "Integrating cellular networks, smart grid, and renewable energy: Analysis, architecture, and challenges," *IEEE access*, vol. 3, pp. 2755–2770, 2015.
[82] X. Wang, Y. Zhang, T. Chen, and G. B. Giannakis, "Dynamic energy management for smart-grid-powered coordinated multipoint systems," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1348–1359, 2016.
[83] T.-Y. Wu and T. Chang, "Interference reduction by millimeter wave technology for 5g-based green communications," *IEEE Access*, vol. 4, pp. 10 228–10 234, 2016.
[84] W. Liu, X. Li, and M. Chen, "Energy efficiency of mimo transmissions in wireless sensor networks with diversity and multiplexing gains," in *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 4. IEEE, 2005, pp. iv–897.
[85] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5g cellular: It will work!" *IEEE access*, vol. 1, pp. 335–349, 2013.
[86] Y. Gai, L. Zhang, and X. Shan, "Energy efficiency of cooperative mimo with data aggregation in wireless sensor networks," in *2007 IEEE Wireless Communications and Networking Conference*. IEEE, 2007, pp. 791–796.
[87] M. Di Renzo, "Energy efficiency metrics and performance tradeoffs of green wireless networks," *Green Communications: Principles, Concepts and Practice*, pp. 43–54, 2015.
[88] B. C. Banister, M. Brehler, P. Gaal, M. Kitazoe, and K. Bhattad, "Energy saving mode with maintained number of advertised transmit antennas," Aug. 15 2017, uS Patent 9,736,707.
[89] K. Kanwal and G. A. Safdar, "Reduced early handover for energy saving in lte networks," *IEEE Communications Letters*, vol. 20, no. 1, pp. 153–156, 2016.
[90] T. Chen, H. Zhang, Z. Zhao, and X. Chen, "Towards green wireless access networks," in *2010 5th International ICST Conference on Communications and Networking in China*. IEEE, 2010, pp. 1–6.
[91] W. Ejaz, S. K. Sharma, S. Saadat, M. Naeem, A. Anpalagan, and N. Chughtai, "A comprehensive survey on resource allocation for cran in 5g and beyond networks," *Journal of Network and Computer Applications*, p. 102638, 2020.
[92] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser mimo systems," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1436–1449, 2013.
[93] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user mimo systems: Is massive mimo the answer?" *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3059–3075, 2015.
[94] H. Q. Ngo, L.-N. Tran, T. Q. Duong, M. Matthaiou, and E. G. Larsson, "On the total energy efficiency of cell-free massive mimo," *IEEE*

Journal Pre-proof

Transactions on Green Communications and Networking, vol. 2, no. 1, pp. 25–39, 2018.

[95] N. Lu, Y. Jiang, F. Zheng, and X. You, "Energy efficient power control for the two-tier networks with small cells and massive mimo," in *2016 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*. IEEE, 2016, pp. 127–132.

[96] J. Chen, X. Chen, T. Liu, and L. Lei, "Toward green and secure communications over massive mimo relay networks: Joint source and relay power allocation," *IEEE Access*, vol. 5, pp. 869–880, 2017.

[97] Y. Xin, D. Wang, J. Li, H. Zhu, J. Wang, and X. You, "Area spectral efficiency and area energy efficiency of massive mimo cellular systems," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 3243–3254, 2016.

[98] W. Xu, S. Li, S. Wang, Z. Feng, J. Lin, and A. V. Vasilakos, "Joint parameter selection for massive mimo: An energy-efficient perspective," *IEEE Access*, vol. 4, pp. 3719–3731, 2016.

[99] M. Alonzo, S. Buzzi, and A. Zappone, "Energy-efficient downlink power control in mmwave cell-free and user-centric massive mimo," in *2018 IEEE 5G World Forum (5GWF)*. IEEE, 2018, pp. 493–496.

[100] O. Saatlou, M. O. Ahmad, and M. Swamy, "Optimal power allocation for massive mu-mimo downlink tdd systems," in *2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*. IEEE, 2016, pp. 1–4.

[101] L. Chen, F. R. Yu, H. Ji, B. Rong, X. Li, and V. C. Leung, "Green full-duplex self-backhaul and energy harvesting small cell networks with massive mimo," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3709–3724, 2016.

[102] (30-01-2019) Effect of electromagnetic radiation on human-health. [Online]. Available: https://www.airestech.com

[103] (01-11-2018) Health effects from radio frequency electromagnetic fields. [Online]. Available: http://www.hpa.org.uk

[104] K. Agarwal and Y.-X. Guo, "Interaction of electromagnetic waves with humans in wearable and biomedical implant antennas," in *2015 Asia-Pacific Symposium on Electromagnetic Compatibility (APEMC)*. IEEE, 2015, pp. 154–157.

[105] M. Younis, T. Rommel, F. Bordoni, G. Krieger, and A. Moreira, "On the pulse extension loss in digital beamforming sar," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 7, pp. 1436–1440, 2015.

[106] B. M. Hochwald and D. J. Love, "Minimizing exposure to electromagnetic radiation in portable devices," in *2012 Information Theory and Applications Workshop*. IEEE, 2012, pp. 255–261.

[107] W. Ni, P.-H. Chiang, and S. Dey, "Energy efficient hybrid beamforming in massive mu-mimo systems via eigenmode selection," in *2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*. IEEE, 2017, pp. 400–406.

[108] O. Tervo, L.-N. Tran, H. Pennanen, S. Chatzinotas, M. Juntti, and B. Ottersten, "Energy-efficient coordinated multi-cell multi-group multicast beamforming with antenna selection," in *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2017, pp. 1209–1214.

[109] L. Zhou, Z. Xu, W. Jiang, and W. Luo, "Joint multicast beamforming and user scheduling in large-scale antenna systems," *IET Communications*, vol. 12, no. 11, pp. 1307–1314, 2018.

[110] S. Tariq, D. Psychoudakis, O. Eliezer, and F. Khan, "A new approach to antenna beamforming for millimeter-wave fifth generation (5g) systems," in *2018 Texas Symposium on Wireless and Microwave Circuits and Systems (WMCS)*. IEEE, 2018, pp. 1–5.

[111] M. Scarfi *et al.*, "International and national expert group evaluations: biological/health effects of radiofrequency fields," *International journal of environmental research and public health*, vol. 11, no. 9, pp. 9376–9408, 2014.

[112] R. Gómez-Villanueva, H. Jardón-Aguilar, and R. L. y Miranda, "State of the art methods for low sar antenna implementation," in *Proceedings of the Fourth European Conference on Antennas and Propagation*. IEEE, 2010, pp. 1–4.

[113] L. Ragha and M. Bhatia, "Evaluation of sar reduction for mobile phones using rf shields," *International Journal of Computer Applications*, vol. 1, no. 13, pp. 80–85, 2010.

[114] D. Ying, D. J. Love, and B. M. Hochwald, "Beamformer optimization with a constraint on user electromagnetic radiation exposure," in *2013 47th Annual Conference on Information Sciences and Systems (CISS)*. IEEE, 2013, pp. 1–6.

[115] Y. Yang, Y. Zhang, L. Dai, J. Li, S. Mumtaz, and J. Rodriguez, "Transmission capacity analysis of relay-assisted device-to-device overlay/underlay communication," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 1, pp. 380–389, 2017.

[116] M. Jung, K. Hwang, and S. Choi, "Joint mode selection and power allocation scheme for power-efficient device-to-device (d2d) communication," in *2012 IEEE 75th vehicular technology conference (VTC Spring)*. IEEE, 2012, pp. 1–5.

[117] X. Tao, X. Xiao, and J. Lu, "A qos-aware power optimization scheme in ofdma systems with integrated device-to-device (d2d) communications," *Engine*, 2012.

[118] M. Belleschi, G. Fodor, and A. Abrardo, "Performance analysis of a distributed resource allocation scheme for d2d communications," in *2011 IEEE GLOBECOM Workshops (GC Wkshps)*. IEEE, 2011, pp. 358–362.

[119] P. Mach, Z. Becvar, and T. Vanek, "In-band device-to-device communication in ofdma cellular networks: A survey and challenges," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 1885–1922, 2015.

[120] G. Fodor, E. Dahlman, G. Mildh, S. Parkvall, N. Reider, G. Miklós, and Z. Turányi, "Design aspects of network assisted device-to-device communications," *IEEE Communications Magazine*, vol. 50, no. 3, pp. 170–177, 2012.

[121] P. Li and S. Guo, "Incentive mechanisms for device-to-device communications," *IEEE Network*, vol. 29, no. 4, pp. 75–79, 2015.

[122] S. Mumtaz, H. Lundqvist, K. M. S. Huq, J. Rodriguez, and A. Radwan, "Smart direct-lte communication: An energy saving perspective," *Ad Hoc Networks*, vol. 13, pp. 296–311, 2014.

[123] Z. Zhou, M. Dong, K. Ota, J. Wu, and T. Sato, "Energy efficiency and spectral efficiency tradeoff in device-to-device (d2d) communications," *IEEE Wireless Communications Letters*, vol. 3, no. 5, pp. 485–488, 2014.

[124] Y. Xu and S. Wang, "Mode selection for energy efficient content delivery in cellular networks," *IEEE Communications Letters*, vol. 20, no. 4, pp. 728–731, 2016.

[125] C. Kai, H. Li, L. Xu, Y. Li, and T. Jiang, "Energy-efficient device-to-device communications for green smart cities," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 4, pp. 1542–1551, 2018.

[126] P. Gandotra and R. K. Jha, "Device-to-device communication in cellular networks: A survey," *Journal of Network and Computer Applications*, vol. 71, pp. 99–117, 2016.

[127] A. Ramezani-Kebrya, M. Dong, B. Liang, G. Boudreau, and S. H. Seyedmehdi, "Joint power optimization for device-to-device communication in cellular networks with interference control," *IEEE Transactions on Wireless Communications*, vol. 16, pp. 5131–5146, 2017.

[128] A. Shahid, K. S. Kim, E. De Poorter, and I. Moerman, "Self-organized energy-efficient cross-layer optimization for device to device communication in heterogeneous cellular networks," *IEEE Access*, vol. 5, pp. 1117–1128, 2017.

[129] Y. Jiang, Q. Liu, F. Zheng, X. Gao, and X. You, "Energy-efficient joint resource allocation and power control for d2d communications," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6119–6127, 2016.

[130] A. Habbal, S. I. Goudar, and S. Hassan, "A context-aware radio access technology selection mechanism in 5g mobile network for smart city applications," *Journal of Network and Computer Applications*, vol. 135, pp. 97–107, 2019.

[131] A. M. Hamed and R. K. Rao, "Evaluation of capacity and power efficiency in millimeter-wave bands," in *2016 International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS)*. IEEE, 2016, pp. 1–6.

[132] T. S. Rappaport, Y. Xing, G. R. MacCartney, A. F. Molisch, E. Mellios, and J. Zhang, "Overview of millimeter wave communications for fifth-generation (5g) wireless networks with a focus on propagation models," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 12, pp. 6213–6230, 2017.

[133] S. A. R. Naqvi, S. A. Hassan, H. Pervaiz, Q. Ni, and L. Musavian, "Self-adaptive power control mechanism in d2d enabled hybrid cellular network with mmwave small cells: An optimization approach," in *2016 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2016, pp. 1–6.

[134] Z. Wei, X. Zhu, S. Sun, Y. Huang, A. Al-Tahmeesschi, and Y. Jiang, "Energy-efficiency of millimeter-wave full-duplex relaying systems: Challenges and solutions," *IEEE Access*, vol. 4, pp. 4848–4860, 2016.

[135] M. Hussain and N. Michelusi, "Energy efficient beam-alignment in millimeter wave networks," in *2017 51st Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2017, pp. 1219–1223.

[136] M. Feng, S. Mao, and T. Jiang, "Dynamic base station sleep control and rf chain activation for energy-efficient millimeter-wave cellular

systems," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 10, pp. 9911–9921, 2018.
[137] Z. Chen, X. Ma, B. Zhang, Y. Zhang, Z. Niu, N. Kuang, W. Chen, L. Li, and S. Li, "A survey on terahertz communications," *China Communications*, vol. 16, no. 2, pp. 1–35, 2019.
[138] K. Tekbıyık, A. R. Ekti, G. K. Kurt, and A. Görçin, "Terahertz band communication systems: Challenges, novelties and standardization efforts," *Physical Communication*, vol. 35, p. 100700, 2019.
[139] R. Arshad, S. Zahoor, M. A. Shah, A. Wahid, and H. Yu, "Green iot: An investigation on energy saving practices for 2020 and beyond," *IEEE Access*, vol. 5, pp. 15 667–15 681, 2017.
[140] A. Srivastava, M. S. Gupta, and G. Kaur, "Green smart cities," *Green and Smart Technologies for Smart Cities*, pp. 1–18, 2019.
[141] L. Da Xu, W. He, and S. Li, "Internet of things in industries: A survey," *IEEE Transactions on industrial informatics*, vol. 10, no. 4, pp. 2233–2243, 2014.
[142] M. Maksimovic, "Greening the future: Green internet of things (g-iot) as a key technological enabler of sustainable development," in *Internet of Things and Big Data Analytics Toward Next-Generation Intelligence*. Springer, 2018, pp. 283–313.
[143] A. Jain, M. Mishra, S. K. Peddoju, and N. Jain, "Energy efficient computing-green cloud computing," in *2013 International Conference on Energy Efficient Technologies for Sustainability*. IEEE, 2013, pp. 978–982.
[144] S. Li, L. Da Xu, and X. Wang, "Compressed sensing signal and data acquisition in wireless sensor networks and internet of things," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 4, pp. 2177–2186, 2013.
[145] Y. Cho, M. Kim, and S. Woo, "Energy efficient iot based on wireless sensor networks," in *2018 20th International Conference on Advanced Communication Technology (ICACT)*. IEEE, 2018, pp. 294–299.
[146] R. Sukjaimuk, Q. Nguyen, and T. Sato, "A smart congestion control mechanism for the green iot sensor-enabled information-centric networking," *Sensors*, vol. 18, no. 9, p. 2889, 2018.
[147] M. A. Zamora-Izquierdo, J. Santa, and A. F. Gómez-Skarmeta, "An integral and networked home automation solution for indoor ambient intelligence," *IEEE Pervasive Computing*, vol. 9, no. 4, pp. 66–77, 2010.
[148] Y. H. Lee and I. Marsic, "A sensing platform for a uhf rfid system," *International Journal of Scientific & Technology Research*, vol. 4, no. 8, pp. 114–121, 2015.
[149] M. Razzaque and S. Dobson, "Energy-efficient sensing in wireless sensor networks using compressed sensing," *Sensors*, vol. 14, no. 2, pp. 2822–2859, 2014.
[150] P. Sathyamoorthy and E. Ngai, "Energy efficiency as a orchestration service for the internet of things," in *Proc. 11th Swedish Nat. Comput. Netw. Workshop*, 2015, pp. 41–44.
[151] M. Ghamkhari and H. Mohsenian-Rad, "Optimal integration of renewable energy resources in data centers with behind-the-meter renewable generator," in *2012 IEEE International Conference on Communications (ICC)*. IEEE, 2012, pp. 3340–3344.
[152] D. Bol, J. De Vos, F. Botman, G. Streel, b. Bernard, D. Flandre, and J.-D. Legat, "Green socs for a sustainable internet-of-things," in *2013 IEEE Faible Tension Faible Consommation*. IEEE, 2013, pp. 1–4.
[153] J. Li, Y. Liu, Z. Zhang, J. Ren, and N. Zhao, "Towards green iot networking: Performance optimization of network coding based communication and reliable storage," *IEEE Access*, vol. 5, pp. 8780–8791, 2017.
[154] M. Moreno, B. Úbeda, A. Skarmeta, and M. Zamora, "How can we tackle energy efficiency in iot basedsmart buildings?" *Sensors*, vol. 14, no. 6, pp. 9582–9614, 2014.
[155] V. Cisco, "Cisco visual networking index: Forecast and trends, 2017–2022," *White Paper*, 2018.
[156] E. Fcc, "Docket no 03-222 notice of proposed rule making and order," 2003.
[157] F. Hu, B. Chen, and K. Zhu, "Full spectrum sharing in cognitive radio networks toward 5g: A survey," *IEEE Access*, vol. 6, pp. 15 754–15 776, 2018.
[158] H. Eltom, S. Kandeepan, R. J. Evans, Y. C. Liang, and B. Ristic, "Statistical spectrum occupancy prediction for dynamic spectrum access: a classification," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, p. 29, 2018.
[159] M. Gupta and K. Kumar, "Progression on spectrum sensing for cognitive radio networks: A survey, classification, challenges and future research issues," *Journal of Network and Computer Applications*, 2019.
[160] J. Zhu and Y. Zou, "Cognitive network cooperation for green cellular networks," *IEEE Access*, vol. 4, pp. 849–857, 2016.
[161] S. R. Sabuj and M. Hamamura, "Energy efficiency analysis of cognitive radio network using stochastic geometry," in *2015 IEEE Conference on Standards for Communications and Networking (CSCN)*. IEEE, 2015, pp. 245–251.
[162] P. S. Tripathi and R. Prasad, "Energy efficiency in cognitive radio network," in *Wireless VITAE 2013*. IEEE, 2013, pp. 1–5.
[163] Z. Shi, T. Tan, K. C. Teh, and K. H. Li, "Energy efficient cognitive radio network based on multiband sensing and spectrum sharing," *IET Communications*, vol. 8, no. 9, pp. 1499–1507, 2014.
[164] F. Zhou, N. C. Beaulieu, Z. Li, J. Si, and P. Qi, "Energy-efficient optimal power allocation for fading cognitive radio channels: Ergodic capacity, outage capacity, and minimum-rate capacity," *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2741–2755, 2016.

Journal Pre-proof

Journal Pre-proof

AUTHOR DECLARATION

Akanksha Srivastava, Mani Shekhar Gupta & Gurjit Kaur
Department of E&CE,
Delhi Technological University,
Shahbad Daulatpur, Main Bawana Road,
Delhi, India, PIN- 110042.
April 22, 2020

Subject: Regarding Conflict of Interest for manuscript entitled "Energy Efficient Transmission Trends Towards Future Green Cognitive Radio Networks (5G): Progress, Taxonomy and Open Challenges".

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been financial support for this work that could not have influenced its outcome.
We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

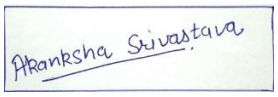
We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). He/she is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs.

We confirm that we have provided a current, correct email address which is accessible by the Corresponding Author and which has been configured to accept email from (akankshasrivastava30@gmail.com).

Journal Pre-proof

Regards



Akanksha Srivastava
(Corresponding author)



Mani Shekhar Gupta
(Co-author)



Dr. Gurjit Kaur
(Co-author)

Journal Pre-proof

Energy Efficient Transmission Towards Future Green Cognitive Radio Networks (5G): Progress, Taxonomy and Open Challenges

Akanksha Srivastava[1], Mani Shekhar Gupta[2], Gurjit Kaur[3]
[1, 2, 3]Department of Electronics and Communication Engineering
[1, 3]Delhi Technological University, Delhi, India
[2]National Institute of Technology, Hamirpur, Himachal Pradesh, India
Email id: [1]akankshasrivastava30@gmail.com, [2]mi_sr87@yahoo.com, [3]gurjitkaur@dtu.ac.in

Authors' detail:



Akanksha Srivastava has received her B. Tech. degree in Electronics and Communication Engineering from Uttar Pradesh Technical University (now Dr. A.P.J. Abdul Kalam Technical University), Lucknow, India in 2010, and her M. Tech. degree in Electronics and Communication Engineering from Dr. A.P.J. Abdul Kalam Technical University, Lucknow, India in 2014. She is pursuing her Ph.D. degree in Electronics and Communication Engineering from Delhi Technological University, Delhi, India. She is also working as a Principal Investigator in a project under Women Scientists Scheme-A (WOS-A) funded by Department of Science and Technology, Government of India (vide File No. SR/WOS-A/ET-154/2017) at the department of electronics and communication engineering, Delhi Technological University, Delhi. From 2014 to 2018, she was an Assistant Professor in the Department of Electronics and Communication Engineering at Pranveer Singh Institute of Technology, Kanpur, India. She has various publications in conferences, journals and book chapters, etc. Her current research includes Green Communication, Energy Efficient Wireless Networks, Cognitive Radio Networks, Wireless Communication, and Heterogeneous Networking.



Mani Shekhar Gupta has received his B. Tech. degree in Electronics and Communication Engineering from Uttar Pradesh Technical University (now Dr. A.P.J. Abdul Kalam Technical University), Lucknow, India in 2009 and his M. Tech. degree in Electronics and Communication Engineering from National Institute of Technology, Hamirpur, India in 2011. He is currently pursuing his Ph.D. degree in Electronics and Communication Engineering from National Institute of Technology, Hamirpur, India. He worked as an Assistant Professor in Department of Electronics and Communication Engineering at Pranveer Singh Institute of Technology, Kanpur, India, from 2011- 2017. He has qualified national level Graduate Aptitude Test in Engineering (GATE) and was a recipient of the GATE Financial Scholarship from 2009 to 2011. Currently, he is a recipient of research fellowship from Ministry of Human Resource Development (MHRD), Government of India. He has various publications in conferences,

journals and book chapters etc. His current research includes Cognitive Radio Networks, Vehicular networks, Spectrum and mobility management, Wireless communication, Next Generation Networks, Resource allocation and Heterogeneous networking.



Gurjit Kaur is an Associate Professor in the Department of Electronics & Communication Engineering at the Delhi Technological University (DTU), Delhi, India and having 16 years of research and teaching experience in well-reputed institutes. She was awarded a 'Gold Medal' by former President of India Dr. A P J Abdul Kalam for being the overall topper of the Punjab Technical University, Jalandhar in B.Tech program. She has completed her M. Tech degree in 2003 and Ph.D. in 2010 from PEC University of Technology, Chandigarh. She has also authored many books at the international and national level. She contributed various book chapters in well-renowned publishers' books i.e., IGI, Springer, CRC, etc. For her outstanding research work she received Prof. Indira Parikh 50 Women in Education Leader Award in World Education Congress, Mumbai, India in 2017. Due to her pioneer research, her name is also listed in Who's Who in Science and Engineering Directory, USA and her biography was also published by International Biographic Centre, Great Britain as 2000 Outstanding Scientists. She is also a reviewer of various reputed journals. Her research interests mainly include Optical CDMA, Wireless Communication system, high-speed interconnect and IoT.

Journal Pre-proof

Proceedings of the Fourth International Conference on Trends in Electronics and Informatics (ICOEI 2020)
IEEE Xplore Part Number: CFP20J32-ART; ISBN: 978-1-7281-5518-0

Ensemble learning-based approach for crack detection using CNN

Vibhu Kailkhura*, S. Aravindh†, Shashank Shekhar Jha‡, N. Jayanthi§
Department of Electronics & Communication Engineering
Delhi Technological University
*vibhu bt2k16@dtu.ac.in, †saravindh bt2k16@dtu.ac.in, ‡jhashashank012@gmail.com, §njayanthi@dce.ac.in

***Abstract*— Crack detection is of pivotal importance in civil engineering and other related applications. Traditional methods of human inspection are tedious and severely limited. Automated crack detection by conventional image processing techniques is challenging due to their inability to discriminate crack features from background noise. Inhomogeneous lighting, shadows, and surface finish hinder the performance of digital image processing methods. The use of convolutional neural networks has helped achieve remarkably better results in the field of computer vision. Ensemble learning is an approach to aggregate the results of a number of individual models for classification or regression. Ensemble learning for crack detection has been implemented using deep convolutional neural networks (DCNN) in this paper. The models are evaluated on a number of performance metrics, namely–(i) accuracy, (ii) precision, (iii) recall (iv) Matthews correlation coefficient (MCC), (v) AUROC, and (vi) F1 score. Experimental results show the robustness of the ensembling method and offer promising scope in crack detection. They outperform the current best performance on open source concrete crack dataset. The ensemble models achieved much better performance than their individual counterparts with the best ensemble achieving a validation accuracy of 99.67%.**

Keywords—crack detection, convolutional neural network (CNN), transfer learning, ensemble learning

I. INTRODUCTION

Concrete structures suffer from decay over time by various means. Cracks develop on the surface, which need to be properly detected. A need for Structural Health Monitoring(SHM) arises as the properties of concrete depends on many factors and are often hard to predict in practice. SHM is a process aimed at providing real-time and accurate information about structural conditions and performance on a proactive basis. SHM is very economical and provides good returns. The cost of installing and implementing an SHM system is only 0.5% to 3% of total construction and repair costs, respectively [1]. Having an SHM system in place can save a lot of money and human lives. Manual detection of cracks cannot set a uniform standard and is subject to an individual's expertise. Automated crack detection overcomes this problem by precise inspection without human error. The application of traditional algorithms of image processing for crack detection led to satisfactory results.

Yamaguchi et al., 2010 [2] used a percolation based image processing method for fast crack detection. A terminate and skip procedure was used where only the crack pixels were checked instead of the background. Circularity of percolated region was then found to set the threshold for classifying pixels. Salman et al., 2013 [3] proposed to use a Gabor filter of a given orientation, which was convolved with the input pre-processed image. After the completion of convolution, the real component of the response out of the kernel was threshold to generate a binary output image. Finally, the binary images resulting from the differently oriented filters were combined by logical OR operation to produce an output image that contained detected crack segments. Yiyang et al., 2014 [4] proposed a glass crack detection algorithm using image processing techniques. The crack information was obtained by making use of image pre-processing, segmentation, and feature extraction of the glass crack image. The pre-processing task involved image smoothing filtering and image sharpening. A twice threshold segmentation was used in [5], and road marking was removed by the Otsu threshold algorithm. Segmentation was done by the improved adaptive iterative threshold segmentation algorithm. Shan et al., 2016 [6] proposed a Canny-Zernike combination algorithm for image processing. SIFT (Scale Invariant Feature Transform) was used for feature extraction. The nearest neighbor algorithm was applied to match the feature points extracted by the SIFT algorithm. Inhomogeneous illumination of the surface and discriminative crack features being confused with background noise are a challenging task for image processing methods.

The image processing algorithms have now been largely outperformed by artificial neural network-based methods. Of the various neural networks in use, the most significant for computer vision is the convolutional neural network (CNN). It offers the advantage of lesser computation in terms of a lesser number of weights to train for a single kernel that can be applied repeatedly to all the portions of the image. It is also able to learn features automatically. The first use of CNNs for crack detection was implemented by Zhang et al., 2016 [7]. They applied CNN on images downsized from a road image of 3264 x 2248 and achieved substantial accuracy. Li et al. [8] used a modified version of GoogLeNet for concrete crack detection. A deep learning-based CNN architecture was proposed to detect cracks in concrete surfaces in [9]. A comparative study was conducted to show the performance difference between CNN and traditional image processing techniques for the purpose of crack detection [10]. The vision-based method using deep CNN performed quite better than the traditional Canny and Sobel edge detection technique. Li et al., 2019 [11] used AlexNet as the CNN architecture for crack detection, and the model was trained for optimum learning parameters.

Ensemble learning is the approach to use multiple models for accurate prediction. The results of the individual models are combined by various methods to achieve better accuracy. The technique of ensembling has been applied in various domains. Shahin at al. [12] discuss the improved accuracy of their approach using ensemble learning for the problem of skin lesion classification. Ensemble models yield promising results when the individual models have significant diversity among them [13].

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 05:00:11 UTC from IEEE Xplore. Restrictions apply.

In this paper, ensemble learning is applied with the motive of crack detection. The paper is structured in smaller sections for better understanding. Section II discusses the concept of Convolutional Neural Networks. Section III is based on ensemble learning. Section IV elaborates the methodology adopted in this paper for crack detection. Section V includes the specifications of the dataset used. Section VI lists the results. Analysis of results is done in section VII. Section VIII presents the conclusion.

II. CONVOLUTIONAL NEURAL NETWORKS

A Convolutional Neural Network (CNN) is a deep learning algorithm that takes an image as an input, learns the various features of that image, and then classifies them into their respective classes. The CNN architecture is similar to the pattern of the neurons in a human brain. It enables machines to perceive the world as humans do by applying relevant filters to capture the spatial and temporal dependencies of an image successfully. The convolution layer and pooling layer are used in feature learning. The Fully Connected layers and the softmax layer are used for the classification task. All the mentioned components are explained in detail as follows:

A. Convolution Layer

It is the core building block of a Convolutional Network and does most of the computational tasks. The input image is essentially an array of pixel values. The kernel or a filter is also an array of numbers (called weights or parameters). The depth of this filter is the same as that of the input. The filter starts convolving around the input by multiplying the pixel values element-wise. These multiplications are then summed up to a single number. This process is then repeated for every unique location. After sliding the filter over all locations, an activation or feature map is obtained. The objective of this operation is to extract high-level features such as edges from the input image.

B. Pooling Layer

The pooling layer reduces the images into a form that is easier to process, without losing critical features. It is responsible for reducing the spatial size of the feature map. Through dimensionality reduction, the computation cost to process the data is reduced. This layer also controls overfitting to some extent. There are two types of pooling available - max pooling and average pooling. Studies have shown that max pooling performs a lot better than average pooling. The convolution layers and pooling layers together form structural units which can be increased to suit the complexity of the image but at the cost of increased computational power.

C. Fully Connected Layer

The Fully Connected layer represents the typical layer of a neural network wherein the activation of a neuron in next layer is determined by its connections with all neurons in the present layer. The FC layers are attached to the end of the network. They take the output of the preceding layers as input and the last FC layer gives an N-dimensional vector output, where N is the number of classes. Each number in this vector represents the confidence of model for a particular class.

D. Softmax Layer

The softmax function limits the output between 0 to 1 and allows it to be directly interpreted as a probability. Since they are interpreted as a probability, they must sum to unity. A softmax layer is typically the final layer in a CNN.

III. ENSEMBLE LEARNING

Stochastic optimization is employed by deep convolutional neural networks (DCNNs) to learn. However, due to their sensitivity to small fluctuations in the training dataset, their performance gets limited as the variance in predictions becomes high. Random noise from training data also gets modeled into the learning process leading to overfitting of data. The objective is to reduce model variance and improve robustness and generalization. Training multiple, diverse models and pooling in their predictions is a practical approach to reduce variance. This is the basis of the ensemble learning process, which yields results that are superior to any individual model [14]. Ensemble learning is known to improve the performance of a model or reduce the probability of an unfortunate selection of a poor one [15]. However, merely using ensemble learning does not guarantee better results. Only smartly chosen ensemble models would provide a substantial improvement in the results.

Most of the ensemble learning algorithms work on different data distribution schemes to individual models and the different combination methods of constituent classifier results. In the proposed work, the individual models are trained by using the whole of the training data.

IV. METHODOLOGY

For the purpose of crack detection, the concept of ensemble learning is applied as it provides improved results. A flow diagram illustrating the steps involved is shown in Fig.1.

Various pre-trained CNNs are used as the base models which vary in architecture and performance. A fastai library [16] is used for deep learning, which is an open-source library built on top of Pytorch [17] and provides powerful learning techniques. The pre-trained CNN models with ImageNet [18] are used. These models are fine-tuned for the specific task of crack detection. The fastai library provides good default values for the hyperparameters involved and the option to train different layers of the network at different learning rates. Hence, the initial layers which recognize basic image features are trained at lowest learning rate and the final layers at the highest learning rate. All the models are trained on the training data and their performance is analyzed on the common validation set.

For building ensemble models, a combination that is expected to show improved results should be selected. The selection criteria has been used to choose the base model combination is based on the following two characteristics-

- The diversity in architecture of the individual base models
- Analyzing the confusion matrix of each model on the crack dataset. The critical parameters of false negatives and false positives were taken into account while deciding the combinations of ensemble models which could lead to improvement.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 05:00:11 UTC from IEEE Xplore. Restrictions apply.

The selected model combinations are then used to build ensemble models. The model combinations are indexed in table III. The ensembling methods used in this paper include the most popular techniques of averaging, voting and weighted voting [15]. Using these methods on all the model combinations, various ensemble models are built. The ensemble models are then validated on the same validation dataset used for individual base models and the performance is analyzed.

Let there be *M* number of models for classifying the sample into *C* classes. Then the various ensembling methods may be formulated as below.

1) Averaging

This method is based on taking the simple algebraic average of the predictions of the base models and then treating the average as the final prediction of the ensemble model. If the prediction of the i^{th} model for the j^{th} class to an image input *X* be given as $y_{i,j}(X)$ then the ensemble prediction for the j^{th} class will be expressed as

$$Y_{e,j}(X) = \frac{1}{M}\sum_{i=1}^{M} y_{i,j}(X) \quad (1)$$

where the subscript *e* denotes the prediction by the new ensemble model.

The final classification label of the ensemble model for image input *X* is generated as

$$\hat{Y}_e(X) = class(\max(Y_{e,1}(X), Y_{e,2}(X), \dots, Y_{e,C}(X))) \quad (2)$$

2) Majority Voting

In this method, the individual model predictions are first used to generate classification labels. For each output label $\hat{y}_{i,j}(X)$ generated by a base model corresponding to full support to j^{th} class, the vote $V_j(X)$ of the j^{th} class increases by 1. The class which gains the highest votes is taken to be the final predicted class by the ensemble model.

The final classification label of the majority voting ensemble model can be represented as

$$\hat{Y}_e(X) = class(\max(V_1(X), V_2(X), \dots, V_C(X))) \quad (3)$$

3) Weighted Voting

In this method, the votes of a model to a certain label are weighted in proportion to the performance of the model. Equivalently, the vote of a model with higher accuracy is given more preference.

The individual model predictions are first used to generate classification labels. For each output label $\hat{y}_{i,j}(X)$ generated by a base model corresponding to full support to j^{th} class, the vote $V_j(X)$ of the j^{th} class increases by w_i.

The symbol w_i denotes the weight factor assigned to the vote of i^{th} model and is given as

$$w_i = \frac{a_i}{\sum_{i=1}^{M} a_i} \quad (4)$$

The term a_i is the validation accuracy of the i^{th} model.

The final classification label of the weighted voting ensemble model can be represented as

$$\hat{Y}_e(X) = class(\max(V_1(X), V_2(X), \dots, V_C(X))) \quad (5)$$

V. DATASET AND TRAINING

For training and validation, concrete crack dataset [11] is used. This dataset [11] for training and validation was built out of 1250 images of bridge towers and anchor chambers of a suspension bridge in Dalian, Liaoning, China. The images were captured using a smartphone at a distance of 0.5-1.0m to the surface and were cropped into 256x256 pixel dimension images. Finally, a dataset of 60,000 images was prepared after manual classification.

The images include cracks in varied lighting conditions and situations for enhanced robustness of model. The dataset specifications are given in table I. Sample figures from the dataset are shown in Fig.2.

TABLE I

SPECIFICATION OF THE DATASET USED

Type	Training	Validation	Total
Positive	24k	6k	30k
Negative	24k	6k	30k
Total	48k	12k	60k

VI. EXPERIMENTS AND RESULTS

The models have been trained using the fastai library with Tesla P100-PCie GPU for acceleration on Google Colab, which is a cloud computing platform for machine learning.

The best ensemble model in our experiments outperforms the previous best results as shown in table II.

Let TP, TN, FP and FN denote the true positives, true negatives, false positives and false negatives for the model. Then the performance metrics may be expressed in these terms. The metrics used are listed below.

A. Accuracy

It is a measure of the number of correct predictions made over all kinds of predictions made.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (6)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 05:00:11 UTC from IEEE Xplore. Restrictions apply.

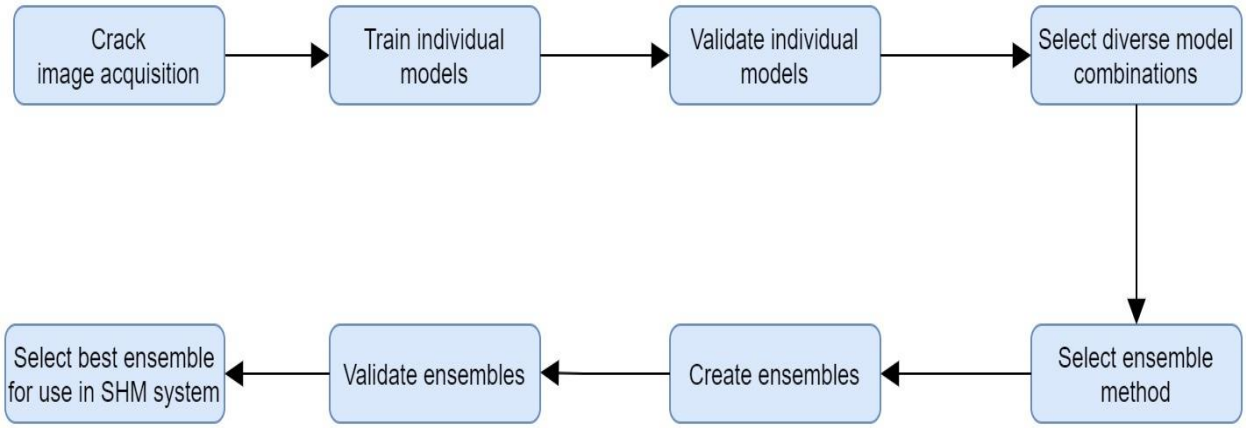


Fig. 1. Flow Diagram for the technique

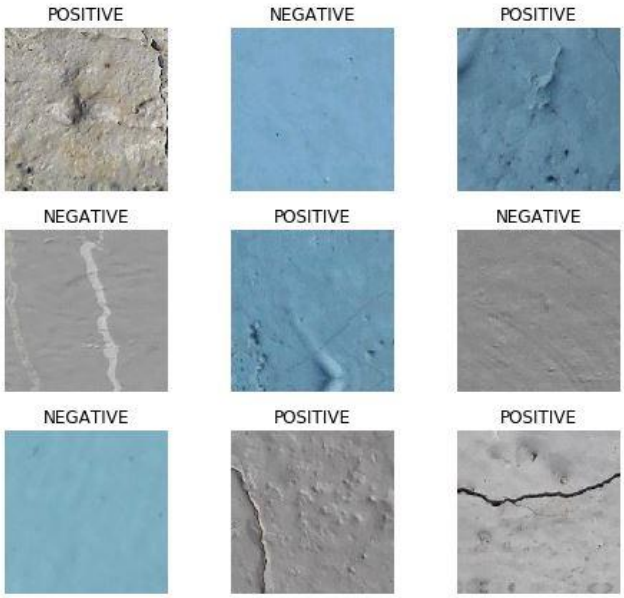


Fig. 2. Sample Images from the dataset

B. Precision

Precision is a measure of what proportion of predicted positives were truly positive.

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

C. Recall

Recall is a measure of what proportion of actual positives were classified correctly.

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

D. Matthews Correlation Coefficient (MCC)

MCC is a correlation between observed and predicted binary classifications.

$$MCC = \frac{(TP.TN) - (FP.FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{9}$$

E. AuROC

AuROC is a measure of how well the probabilities of the positive classes are separated from the negative classes.

F. F1 Score

F1 score is the harmonic mean of recall and precision. It maintains a balance between recall and precision.

$$F1\ Score = 2\,.\frac{Precision.Recall}{Precision + Recall} \tag{10}$$

TABLE II
PERFORMANCE COMPARISON ON DATASET WITH PREVIOUS WORK

Work	**Accuracy**
Li et al. [8]	99.39
Li et al. [11]	99.06
Proposed Ensemble Model	99.67

Results based on the ensembling method are shown for various models in table IV, V, VI and VII.

VII. DISCUSSION

The comparative studies in the form of tables provide a deep insight for analysis. The results are discussed based on the three ensemble methods used. The model combinations are indexed in table III. The models m1-m5 are the individual constituent models. The models m6-m11 are the combinations of two models. Similarly models m12-m14 represent the three-model combinations and m15-m16 are the four-model combinations. Model m17 is the total-ensemble model built using all of the models m1-m5. All ensemble combinations (m6-m17) have been implemented using the three ensembling techniques of averaging, majority voting and weighted voting. Table IV shows the performance of the base models. The results based on the ensembling methods are discussed below.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 05:00:11 UTC from IEEE Xplore. Restrictions apply.

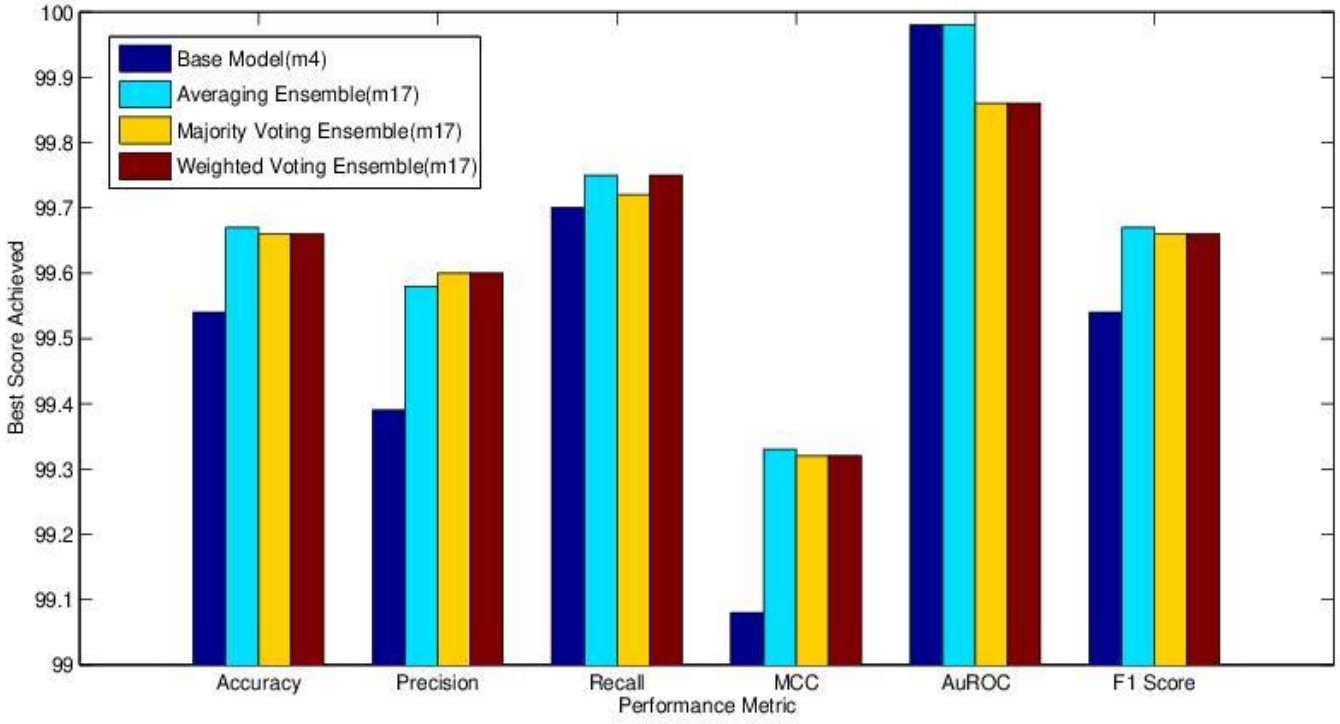


Fig. 3. Performance improvement of ensemble model m17 over the best single model m4

TABLE III

MODEL COMBINATIONS LIST

Model Index	Base Models
m1	ResNet 34
m2	VGG 19
m3	SqueezeNet
m4	DenseNet
m5	AlexNet
m6	DenseNet, SqueezeNet
m7	ResNet, DenseNet
m8	DenseNet, VGG
m9	VGG, ResNet
m10	ResNet, SqueezeNet
m11	SqueezeNet, VGG
m12	ResNet, VGG, DenseNet
m13	DenseNet, VGG, SqueezeNet
m14	SqueezeNet, ResNet, DenseNet
m15	DenseNet, ResNet, VGG, AlexNet
m16	SqueezeNet, AlexNet, VGG, ResNet
m17	Total-ensemble

A. Averaging Based Ensemble Models

The results on the concrete crack dataset can be assessed very well by accuracy since this dataset is balanced in terms of number of images of both classes. Accuracy helps to determine the degree of correct prediction of a model. This includes correct prediction efficiency for both classes-positive and negative. From table IV, it has been noted that, the maximum accuracy for any single model is 99.54 for m4 (DenseNet). However, the top accuracy of 99.67 is achieved using ensembling as shown in table V. Almost all averaging-based ensembles perform better than the best model in the combination.

The top precision of 99.60 is achieved by model m16 which is an improvement over the 99.52 precision value of m1 (ResNet). The improvement in precision value implies that the false alarms in the system are brought down. This helps to reduce the efforts of the visual inspection team which will be given lesser false alarms hence leading to better utilization of resources. Among the single models, the recall value of 99.73 is highest for m2 (VGG19) while ensemble models m8 and m13 have recall of 99.75. Since recall represents the ratio of predicted cracks out of total cracks, a high value of recall is very important. The F1 score presents a combined perspective of precision and recall, and hence higher F1 score is generated when model provides lesser false alarms and is able to detect most of the cracked images correctly. The highest value for F1 score among the single models is that by m4 (DenseNet) with a value of 99.54. This value is significantly improved to 99.67 in the total-ensemble model (m17). The AuROC metric helps to understand the quality of the predictions of the model and how well it can differentiate the two classes. The AuROC value of 99.98 of m4 is maintained in most of the ensemble models. This shows that the models are confident of the predictions they make. The MCC metric may be used when detection of both classes is to be given equal importance. It represents the correlation between the predicted and true classes and is very useful when the dataset is unbalanced. The model m4 provides the best value of 99.08 among the individual models, while the total-ensemble achieves the MCC score of 99.33.

B. Majority Voting Based Ensemble Models

Table VI shows the performance of ensemble models based on majority voting. In this method, the class label which obtains maximum votes from all the models is termed as the true label for the given data sample.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 05:00:11 UTC from IEEE Xplore. Restrictions apply.

TABLE IV

PERFORMANCE OF INDIVIDUAL MODELS

Model	Accuracy	Precision	Recall	MCC	AuROC	F1 score
m1	99.48	**99.52**	99.43	98.95	99.97	99.47
m2	99.42	99.12	**99.73**	98.85	99.97	99.43
m3	99.23	99.15	99.30	98.45	99.95	99.23
m4	**99.54**	99.39	99.70	**99.08**	**99.98**	**99.54**
m5	99.32	99.40	99.23	98.63	99.95	99.32

TABLE V

PERFORMANCE OF ENSEMBLE MODELS BASED ON AVERAGING

Model	Accuracy	Precision	Recall	MCC	AuROC	F1 score
m6	99.58	99.50	99.67	99.17	99.97	99.58
m7	99.57	99.52	99.62	99.13	99.98	99.57
m8	99.54	99.34	**99.75**	99.08	99.98	99.54
m9	99.55	99.48	99.62	99.10	99.98	99.55
m10	99.58	99.55	99.60	99.15	99.97	99.58
m11	99.52	99.34	99.70	99.03	99.97	99.52
m12	99.59	99.48	99.70	99.18	99.98	99.59
m13	99.60	99.45	**99.75**	99.20	99.98	99.60
m14	99.63	99.57	99.70	99.27	99.98	99.63
m15	99.63	99.53	99.73	99.27	99.98	99.63
m16	**99.67**	**99.60**	99.73	**99.33**	99.98	**99.67**
m17	**99.67**	99.58	**99.75**	**99.33**	99.98	**99.67**

TABLE VI

PERFORMANCE OF ENSEMBLE MODELS BASED ON MAJORITY VOTING

Model	Accuracy	Precision	Recall	MCC	AuROC	F1 score
m6	99.58	99.50	99.67	99.17	99.71	99.58
m7	99.57	99.52	99.62	99.13	99.76	99.57
m8	99.54	99.34	**99.75**	99.08	99.75	99.54
m9	99.55	99.48	99.62	99.10	99.73	99.55
m10	99.58	99.55	99.60	99.15	99.75	99.58
m11	99.52	99.34	99.70	99.03	99.73	99.52
m12	99.61	99.48	99.73	99.22	99.81	99.61
m13	99.57	99.39	**99.75**	99.13	99.81	99.57
m14	99.64	99.57	99.72	99.28	99.79	99.64
m15	99.63	99.53	99.73	99.27	**99.86**	99.63
m16	99.65	99.57	99.73	99.30	99.84	99.65
m17	**99.66**	**99.60**	99.72	**99.32**	**99.86**	**99.66**

TABLE VII

PERFORMANCE OF ENSEMBLE MODELS BASED ON WEIGHTED VOTING

Model	Accuracy	Precision	Recall	MCC	AuROC	F1 score
m6	99.54	99.39	99.70	99.08	99.71	99.54
m7	99.54	99.39	99.70	99.08	99.76	99.54
m8	99.54	99.39	99.70	99.08	99.75	99.54
m9	99.48	99.52	99.43	98.95	99.73	99.47
m10	99.48	99.52	99.43	98.95	99.75	99.47
m11	99.42	99.12	99.73	98.85	99.73	99.43
m12	99.61	99.48	99.73	99.22	99.81	99.61
m13	99.57	99.39	**99.75**	99.13	99.81	99.57
m14	99.64	99.57	99.72	99.28	99.79	99.64
m15	99.61	99.48	99.73	99.22	**99.86**	99.61
m16	99.61	99.53	99.68	99.22	99.84	99.61
m17	**99.66**	**99.60**	99.72	**99.32**	**99.86**	**99.66**

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 05:00:11 UTC from IEEE Xplore. Restrictions apply.

TABLE VIII

TABULAR COMPARISON FOR PERFORMANCE IMPROVEMENT IN METRICS

Model Type	Accuracy	Precision	Recall	MCC	AuROC	F1 score
Single Model	99.54(m4)	99.52(m1)	99.73(m2)	99.08(m4)	99.98(m4)	99.54(m4)
Averaging Ensemble	99.67(m16,m17)	99.60(m16)	99.75(m8,m13,m17)	99.33(m16,m17)	99.98(most models)	99.67(m16,m17)
Voting Ensemble	99.66(m17)	99.60(m17)	99.75(m8,m13)	99.32(m17)	99.86(m15,m17)	99.66(m17)
Weighted Voting Ensemble	99.66(m17)	99.60(m17)	99.75(m13)	99.32(m17)	99.86(m15,m17)	99.66(m7)

It can be observed that for even number of models in an ensemble, there is a chance of a tie for the number of votes that each label gets. Hence, the confidence of prediction to the given label is considered to resolve such a tie. No such corrective measure is required for voting ensembles with an odd number of models. As can be seen, almost every ensemble model outperforms the individual best model in the combination.

As can be seen from table VI, the maximum performance is again achieved by the total-ensemble model (m17) on all the parameters except the recall metric. The individual maximum recall of 99.73 for model m2 is outperformed by voting based m8 and m13 models with 99.75 recall score. The recall value for m17 is 99.72 which is almost the same as that of m2. The AuROC value of the best voting ensemble at 99.86 (for m15 and m17) is lower than the individual best of 99.98(for m4). However, the value is still very high and can be termed as a confident distinction ability between the two classes.

The voting ensembles improve the accuracy of 99.54 (for m4) to 99.66 (for m17), the precision of 99.52 (for m1) to 99.60 (for m17), the recall value of 99.73 (for m2) to 99.75 (for m8 and m13), f1 score of 99.54 (for m4) to 99.66 (for m17) and the MCC value of 99.08 (for m4) to 99.32 (for m17).

C. Weighted Voting Based Ensemble Models

The weighted voting method provides higher vote value to model with better accuracy. The results are tabulated in table VII. As expected, the weighted voting results of ensembles with 2 constituent models (m6-m11) are the same as that of the better performing model among the two in the ensemble. For three or more model combinations(m12-m17), the weighted voting ensemble outperforms the best model in the combination.

The weighted voting ensembles improve the accuracy of 99.54 (for m4) to 99.66 (for m17), the precision of 99.52 (for m1) to 99.60 (for m17), the recall value of 99.73 (for m2) to 99.75 (for m13), f1 score of 99.54 (for m4) to 99.66 (for m17) and the MCC value of 99.08 (for m4) to 99.32 (for m17). It is observed that the best AuROC value of 99.86 (for m15 and m17) is lower than the individual best of 99.98 (for m4). However, the value is still very high and demonstrates a good confidence in the predictions.

Hence, it can be concluded that the ensemble models markedly improve the performance than the individual models and the total-ensemble model (m17) shows best results for most of the performance metrics.

D. Comparative analysis of ensemble methods for crack detection

Since the concrete crack dataset is a balanced one, the accuracy provides a good direct estimate of performance. The performance improvement in the metrics by various ensemble methods is compared in table VIII. Among the constituent models, DenseNet (m4) is the best performing, while the total-ensemble model (m17) is the best among the ensembles. The accuracy improvement is from 99.54 to 99.67. The recall improvement is from 99.73 to 99.75 in the m13 model based on all the three methods of ensemble. The precision value is improved by all the three ensemble methods from 99.52(for m1) to 99.60 (for m17). The f1 score is again improved best by averaging method from 99.54 (for m4) to 99.67 (for m16, m17). The MCC metric which gives equal focus to both class labels without any bias, is improved from 99.08 (for m4) to 99.33 (for m16, m17) in the averaging ensemble. The AuROC value is already high and maintained in averaging ensembles which denotes the quality of the classifying decision.

It is observed that the best performing single base model on the given dataset is DenseNet (m4). The total-ensemble model(m17) achieves improvement in almost all the metrics. The performance improvement of the total-ensemble model (m17) over the best single model m4 is shown in Fig.3.

It is important to note that roughly all the parameters are boosted the most by the averaging method of ensemble. This is intuitively expected as the algebraic aggregation of predictions is done and then converted to class labels. It is observed that m16 and m17 based on averaging perform the same on most of the metrics except precision and recall. However, m17 outperforms m16 in other ensembling methods. This shows the strength of the averaging method which helps to achieve improved results. Owing to the simplicity and the algebraic nature of combination, for most of the cases, a simple averaging method may be used to build an optimal ensemble model with a number of diversified models.

VIII. CONCLUSION

In this paper, ensemble learning has been proposed for the problem of concrete crack detection. The individual CNN models are trained and then various ensemble models based on ensemble methods are built. The ensemble models have shown to boost the performance of the individual models. The best ensemble model gives benchmark results with an accuracy of 99.67% on 12k images used for validation. The results offer promising scope of application of more ensemble-based algorithms for crack detection.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 05:00:11 UTC from IEEE Xplore. Restrictions apply.

IX. REFERENCES

[1] P. Roshan, A. Kumar, D. Tewatia, and S. Pal, "Review paper on structural health monitoring: Its benefit and scope in india," Journal of Civil Engineering and Environmental Technology, vol. 2, no. 2, pp. 109–112, 2015.

[2] T. Yamaguchi and S. Hashimoto, "Fast crack detection method for large-size concrete surface images using percolation-based image processing," Machine Vision and Applications, vol. 21, no. 5, pp. 797–809, 2010.

[3] M. Salman, S. Mathavan, K. Kamal, and M. Rahman, "Pave-ment crack detection using the gabor filter," in 16th interna-tional IEEE conference on intelligent transportation systems (ITSC 2013). IEEE, 2013, pp. 2039–2044.

[4] Z. Yiyang, "The design of glass crack detection system based on image preprocessing technology," in 2014 IEEE 7th Joint International Information Technology and Artificial Intelligence Conference. IEEE, 2014, pp. 39–42.

[5] L. Peng, W. Chao, L. Shuangmiao, and F. Baocai, "Research on crack detection method of airport runway based on twice-threshold segmentation," in 2015 Fifth International Conference on Instrumentation and Measurement, Computer, Communica-tion and Control (IMCCC). IEEE, 2015, pp. 1716–1720.

[6] B. Shan, S. Zheng, and J. Ou, "A stereovision-based crack width detection approach for concrete surface assessment," KSCE Journal of Civil Engineering, vol. 20, no. 2, pp. 803–812, 2016.

[7] L. Zhang, F. Yang, Y. D. Zhang, and Y. J. Zhu, "Road crack detection using deep convolutional neural network," in 2016 IEEE international conference on image processing (ICIP). IEEE, 2016, pp. 3708–3712.

[8] S. Li and X. Zhao, "Convolutional neural networks-based crack detection for real concrete surface," in Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2018, vol. 10598. International Society for Optics and Photonics, 2018, p. 105983V.

[9] G. Yao, F.-J. Wei, J.-Y. Qian, and Z.-G. Wu, "Crack detection of concrete surface based on ƒnewline convolutional neural net-works," in 2018 International Conference on Machine Learning and Cybernetics (ICMLC), vol. 1. IEEE, 2018, pp. 246–250.

[10] Y.-J. Cha, W. Choi, and O. Buy¨uk¨ozt¨urk,¨ "Deep learning-based crack damage detection using convolutional neural networks," Computer-Aided Civil and Infrastructure Engineering, vol. 32, no. 5, pp. 361–378, 2017.

[11] S. Li and X. Zhao, "Image-based concrete crack detection using convolutional neural network and exhaustive search technique," Advances in Civil Engineering, vol. 2019, 2019.

[12] A. H. Shahin, A. Kamal, and M. A. Elattar, "Deep ensemble learning for skin lesion classification from dermoscopic im-ages," in 2018 9th Cairo International Biomedical Engineering Conference (CIBEC). IEEE, 2018, pp. 150–153.

[13] D. Opitz and R. Maclin, "Popular ensemble methods: An em-pirical study," Journal of artificial intelligence research, vol. 11, pp. 169–198, 1999.

[14] T. G. Dietterich, "Ensemble methods in machine learning," in International workshop on multiple classifier systems. Springer, 2000, pp. 1–15.

[15] R. Polikar, "Ensemble learning. scholarpedia 4 (1): 2776," 2008.

[16] J. Howard and S. Gugger, "fastai: A layered api for deep learning," 2020.

[17] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf,¨ E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chil-amkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," 2019.

[18] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009, pp. 248–255.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 05:00:11 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Environment Monitoring System for Agricultural Application using IoT and Predicting Crop Yield using Various Data Mining Techniques

Gautam Gupta
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
gg0918914@gmail.com

Rashi Setia
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
rashika.setia8c@gmail.com

Archana Meena
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
pooranlal1973@gmail.com

Bhavnesh Jaint
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
bhavneshjaint@dtu.ac.in

Corresponding Author: Gautam Gupta, gg0918914@gmail.com

***Abstract*— Nowadays, Internet is facing an exponential increase in the number of electronic devices interfaced through it. Previously, there is a limit of connecting only mobile and PCs (personal computer) with the internet but now millions of devices can be connected using the concept of IoT (internet of things). Hence, IoT is able to communicate data between machine to machine and the data that is previously available to private server now easily available to the internet so that it can be accessed efficiently. This paper show the increase in the usage of IoT in monitoring like applications. Smart farming is the technique which intend to provide all the necessary resources for the specified amount of time. These resources required are light intensity, ambient temperature required, relative humidity, soil moisture content, pH reading of the soil. The central idea is to sense all these parameters one by one and take the final decision accordingly. A sensor node should be developed for sensing all the required resources parameters and subsequently send the data to the cloud for further processing. After getting all the required resources values our final aim is to predict the crop production using various data mining techniques. As the human lives are primarily depended on the food resources, the agricultural process need to be efficient and this efficiency can be enhanced if there is an accurate amount of resources. By getting the accurate amount of resources, Data mining techniques such as Random forest, KNN, SVM are used to analyze the crop production in advance such that farmer always have an upper hand on it and by comparing with the previous trend he will be able to detect which kind of parameter is accurate and which is not. A User defined site is designed to monitor all these values of agricultural process.**

***Keywords*- Internet of Things, Wi-Fi, ESP8266 Wi-Fi module, sensor node, Data Mining, Prediction, random Forest, Support Vector Machine (SVM), K-Nearest Neighbor (KNN).**

I. INTRODUCTION

Agriculture is the foremost only field industry where the recent improvement in technology has not been welcomed in huge number. One of the main reason regarding this issue is poor lethargic condition of the farmers in developing countries say India. The scarcity of agricultural products is also increasing day by day due to two main reasons, first one is overpopulation and the second one is urbanization. Considering the case of overpopulation, it increases the demand of agricultural resources and with the case of increasing population it emphasis the better growth of farming products but due to globalization the most of agricultural land is taken by big industries and they convert these land into non-agricultural land. Basically the rural area for farming is decreasing day by day and with that decrease the agricultural resources are also decreasing. Therefore, the need of an hour is to take a look at the situation carefully and emphasis is given to improve the crop production by consuming the resources conservatively without wasting the existing resources. Smart farming is one such methodology to do it. In this type of farming the crop is fed up with the sufficient quantity of resources required for it to grow and by taking the calculated amount of time. In normal Traditional farming method, there is a time based irrigation system i.e. farmer know exactly that after certain limited time (few days) he needs to irrigate his crop. But due to the lack of knowledge farmer is not capable of irrigating his/her field at the exact duration of time which significantly lead to a wastage of water. Hence to overcome such issue smart farming method is employed. To overcome the issue of traditional farming, a soil moisture detection sensor is implanted at the observation site to get the soil moisture content and with this the crop is irrigated accordingly. Similarly, soil moisture sensor, temperature sensor, Humidity sensor and pH sensor are used around the plant to get the desired value, which helps in reducing the exploitation of extra resources used initially in traditional farming. Soil moisture is important in determining the physical

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:57:29 UTC from IEEE Xplore. Restrictions apply.

structure of the plant whereas all the other sensors play a major role in photosynthesis. This paper basically suggest an idea to generate the sensor node which is able to send the data to the central node i.e. to the thing speak cloud. This paper consist of methodology used in smart farming and after visualizing the exact percentage of resources required for crop to appropriately predict the crop yield using various data mining techniques. Data Mining [1] is a procedure to retrieve useful information from very enormous datasets. This technique is used to predict the crop production in advance so that the farmers could take an immediate measures accordingly as the crop demands. It also helps government to get the value of exact MSP (minimum selling price) for farmers. It aims to predict the crop prediction by using various regression and classification methods like multiple linear regression, SVM, Random forest, KNN etc. With these techniques of data mining, the effect of various parameters (sensor node parameters in general) on the prediction of crop production has been studied.

II. LITERATURE SURVEY

In paper [2] author tries to explains the requirement of WSN in agricultural field so as to increase the crop productivity. This paper [2] also give emphasis on the need of smart farming techniques in today's scenario which ultimately analyze the architecture of proposed system by monitoring the environment parameters. The paper [3] suggest the usage of IoT in the field of agriculture. It basically shows how the intermediate chains of agriculture are connected and how IoT can be useful in doing it. It also tell us about all the technologies that are designed by using the concept of IoT. In paper [4] author explained the system using the feature of GPS to monitor the agricultural field. The prime aim of GPS is to look at the data sensed at sensor node and transfer this data using RS232 that are obtained from sensor node to the central node. The proposed system in paper [5] shows the data of light intensity factor around the crop and saving it in a excel sheet. The data stored in database is further used to compare and analyze it later on. In this system only the light intensity parameter is targeted as their only parameter to control the crop production. The author in paper [6] tried to sort out the problem of unequal rain distribution in crop production by controlling the parameters like soil moisture sensor and pH sensor. Arduino is used as a controlling unit which is further also used to communicate the values to the cloud part. The paper [7] shows the problem of food scarcity in Africa. It significantly put forward a schematic way which would predict total crop production needed to meet the African native citizens' needs. The model of Artificial Neural Network is built and subsequently the prediction is done. It helps in predicting the annual amount of crops needed by 2021. The paper [8] defined the comparison algorithms which one can use to make a prediction of kharif crop in suburban zones of India. The algorithms is being executed on an open source platform named WEKA. Afterwards the various type of error is calculated of all the algorithms and being compared individually. In paper [9] author proposed smart agriculture based on IoT and cloud computing methods therefore the agriculture information cloud is constructed with different dynamic resources so as to achieve load balancing. Large amount of data obtained through RFID, and wireless communication are supported through the agriculture information cloud schematic model. In paper [10] the limitation of IoT is discussed. Actually there is communication delay between the cloud server and the end user application. So, paper proposes the model of AI-enabled edge computing to work for IoT applications. It improves the speed and range for the IoT applications. In paper [12] author tries to convey the various data mining techniques one can use while dealing with concept of IoT techniques and provide an overview of what is data mining and different techniques one must use while dealing with it having maximum efficiency.

III. PROPOSED WORK

Our proposed work is divided mainly into two broad categories. First is getting the sensor data from senor node used at the data transmission end and uploading each senor value to thingspeak so as to monitor it for further evaluation. Secondly, after getting sensor data values utilizing it for analyzing the crop yield production will have an upper hand on the future prediction of crop production.

A. Working with sensor node and thingspeak

The main central idea of the work is explained in Fig.1 given below which contains various components namely Relative humidity & ambient air temperature sensor, Soil Moisture content sensor, ph sensor, CO2 sensor, Microcontroller Unit (MCU) along with Wi-Fi module, Wi-Fi router, and thingspeak cloud. Finally the user defined website to monitor the details. Moisture content of the soil is measured with the use of electrode fabricated in it. The main responsible factor through which the soil moisture content changes is the resistance between the two terminals of the electrode. Whenever the soil moisture changes the resistance between the terminal changes. The change in the value of resistance and is the amount of soil moisture content percentage in the soil. A chip is also fabricated inside it which is used to convert the change in resistance to some analog voltage. It is fed with 3.3v input to give an output voltage in the range 0-3.3v. Output voltage signal is fed to MCU unit respectively. Soil moisture sensor have working voltage less than 5V DC and current limit is < 20mA. DHT11 is multivariable sensor which shows the reading of the relative humidity content in percentage and temperature of the atmosphere. It is capable of measuring both the humidity and temperatures single handedly. It senses the relative humidity using capacitors and temperature using the temperature dependent resistors through which change in the output voltage of the resistance gives the relative change in temperature from the previous value. The output waveform of this sensor is digital in nature hence noise must be removed

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:57:29 UTC from IEEE Xplore. Restrictions apply.

subsequently to protect the actual output wave from getting distorted. The digital nature of output pulse force the connection of the sensor to one of the input pins of MCU which is digital in nature. As the resulting output wave is of digital in nature, hence the data is updated after every 2 seconds timeframe. Humidity range of this sensor is 20-80% with 5% accuracy and temperature range of DHT11 is 0-50 degree Celsius with 2 degree accuracy. The sampling rate of DHT11 is 1 Hz and current limit of sensor is 2.5mA. MQ-2 sensor is used to sense the amount of CO2 in surrounding atmosphere. It is metal oxide semiconductor type gas sensor also known as chemiresistors as the sensing is based upon the change in the value of resistance as the gas comes directly in contact with the sensor material. Using the voltage divider rule, the concentration of gas can be found later on in voltage terms. It can detect gas concentration at 200pm-1000pm. It is fed with 5V dc input.

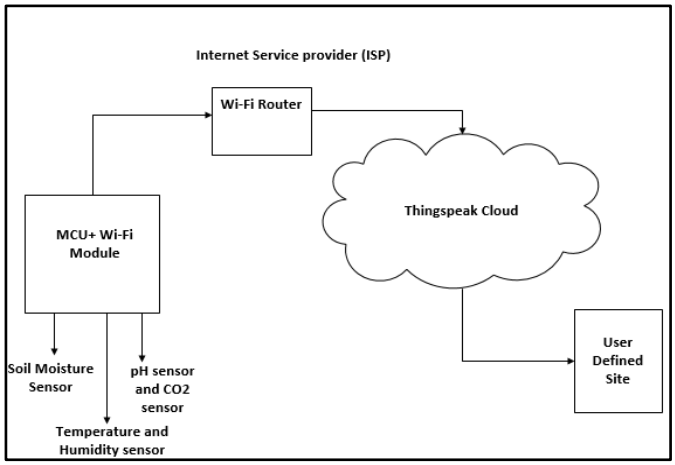


Fig 1. Block Diagram of proposed System

NodeMCU is microcontroller unit here. It is a board which functions on ESP8266 Wi-Fi module. Apart from General pins input-output it also come with additional feature i.e. Pulse Width Modulation capability and analog-digital converter (ADC) all these components are embedded in a single piece of board. NodeMCU basically collects the data from all the sensors that are mentioned above at its GPIOs pins. Data from humidity and temperature sensor is given at GPIO 0 while analog data from the soil moisture sensor is given at the analog input pin of the NodeMCU which having 10 bit of inbuilt ADC. As 10 bit gives a maximum count of 1023 units so it basically mapped the 0-3.3v value to 0-1023 count. It is based on Tensilica 32-bit RSIC microcontroller. The clock speed of the NodeMCU is 80MHz. The SRAM capacity of the device is 80KB. To code the MCU unit Arduino IDE software is used. To connect the NodeMCU board with Wi-Fi, hotspot of the cell phone could also be used which worked as router. Id and password are entered corresponding to the hotspot so that it can connected to the active internet connection.

Thingspeak is used as a cloud source in the above mentioned system. It is basically a platform service which allows you to visualize and analyze live data in the cloud. Certain number of channels are created in this platform where it analyzes the raw data obtained from sensors. Initially, 5 fields are created to analyze the data. The field names are soil moisture content, surrounding temperature, relative humidity content, CO2 and ph. correspondingly, API key is required to code the board that is used in our system. But with the help of thingspeak they are generate automatically. All the data are send in the form analog values so that the graphical representation of the sensor data can be plotted effectively.

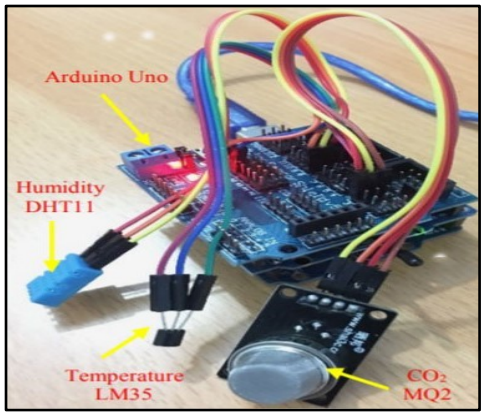


Fig 2. Interfacing of sensors

B. Implementation of Data mining

After sensing the values of all the sensors used above, the measurements of each parameters are stored in the form of datasets. After storing the value, various data mining techniques are used to predict the exact crop production in future. To predict the Crop production three major steps have to be done. These are datasets details, data preprocessing and building prediction model.

1. Datasets used

All the datasets were collected from the previously sensed value from the sensors used at observation site. Temperature, humidity, soil moisture, CO2 percentage and pH value are the common datasets used in the analysis. In addition to these values monthly rainfall dataset is taken from the publicly available records issued by Indian government. Area needed for crop production is also available at the government records. In addition to all these values yield of the crop is also taken under consideration so as to compare it with predicted value with per year yield (in tonnes/hectare).

2. Datasets preprocessing

Datasets preprocessing was done to collect all the datasets of the sensors and additional parameters in Microsoft excel. For every crop it consists of various columns namely: temperature, pH, Soil moisture, relative humidity, CO2 percentage, monthly

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:57:29 UTC from IEEE Xplore. Restrictions apply.

rainfall. By dividing the value of production by area under production, a value is obtained as somehow close to yield. Therefore, by removing the production column and adding yield column or vice-versa and finally saving the file in .csv extension.

3. *Data mining techniques*

Regression is the process performing a statistical modelling through which one can estimate the dependence of dependent variable over one or more than one independent variable. Predictant is the name given to dependent variable and predictors to independent variable. Considering production as dependent (target) variable and all others parameters whether it is area, temperature, soil moisture etc. as independent variables. Now, regression is somewhat similar to classification techniques. Algorithms used in this research paper are Random forest classification, KNN (K-Nearest neighbor) classification and SVM (Support Vector machine) classification.

4. *Constructing the prediction model using sci-kit learn libraries*

This paragraph discuss the building of prediction model of various data mining techniques namely multiple linear regression, K-nearest neighbor (KNN), Support vector machine (SVM) and random forest regression as discussed above using sci-kit learn and py-earth libraries. Sci-kit learn is an open source platform tool for data mining and data analysis via interfacing with python [10]. It provides classification, clustering and regression algorithms. Fig 3. Shows how to construct a prediction model for various data mining techniques as discussed above using sci-kit learn.

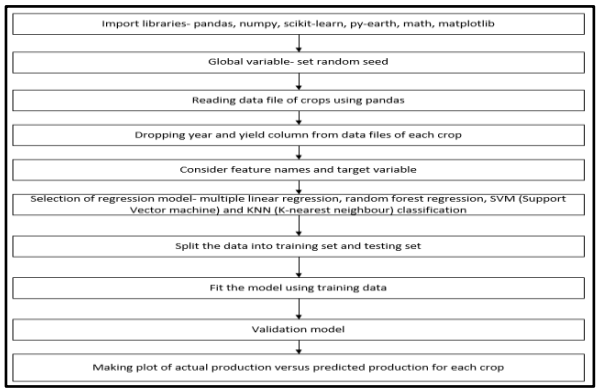


Fig 3. Procedure to construct the prediction model

In prediction model first of all the libraries are imported and then the global variable is created by random seeding method. As our dataset contain year and yield column but it also has production column which is Yield/Area. Hence, both these column are dropped. Now the code of above three classification is added and the data is divided inti training and testing data. Now training data is considered for evaluation and then respective algorithm is added to the data values and plot of actual vs predicted plot is created at last.

IV. RESULTS AND OBSERVATIONS

In the above mentioned work the output of the above methodology depicted in part A of the proposed work is depicted as the variation of values vs observation time and is monitored in thingspeak cloud server. The matlab representation of data sensed at the agricultural field is shown in the form of fields in channel. In fig 4, Temperature is plotted at thingspeak server, similarly in subsequent figures from 5-8 all remaining sensors plot have been obtained.

The plot of ambient temperature of surrounding and time interval is plotted and the values of temperature is recorded in Fahrenheit.

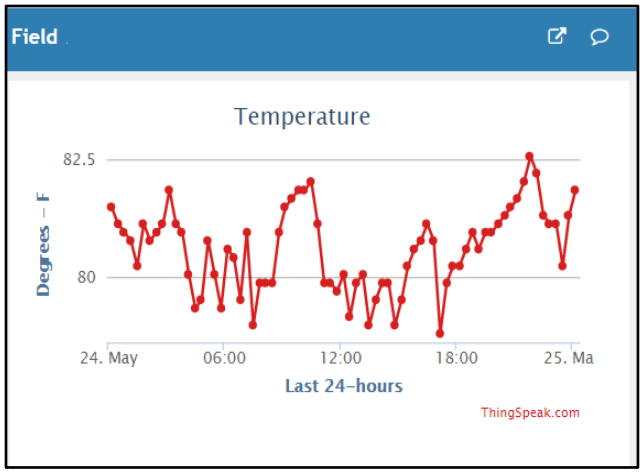


Fig 4. Surrounding air temperature

The plot of relative humidity and time interval is plotted and it is observed that maximum value recorded is 40 percent in humidity.

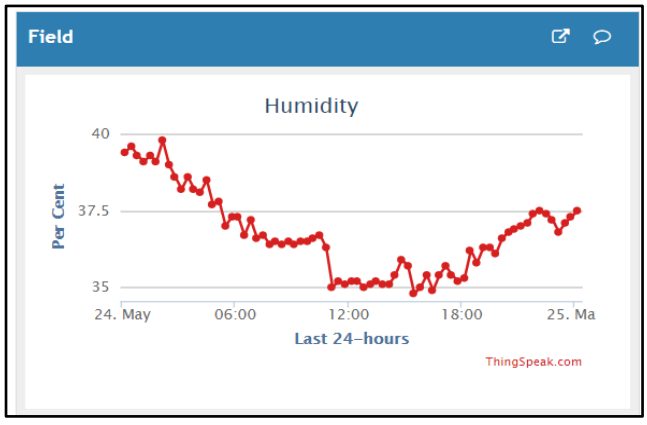


Fig 5. Surrounding air relative humidity

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:57:29 UTC from IEEE Xplore. Restrictions apply.

The plot of soil moisture content vs time interval is plotted and it is noted that soil moisture content is also in percentage.

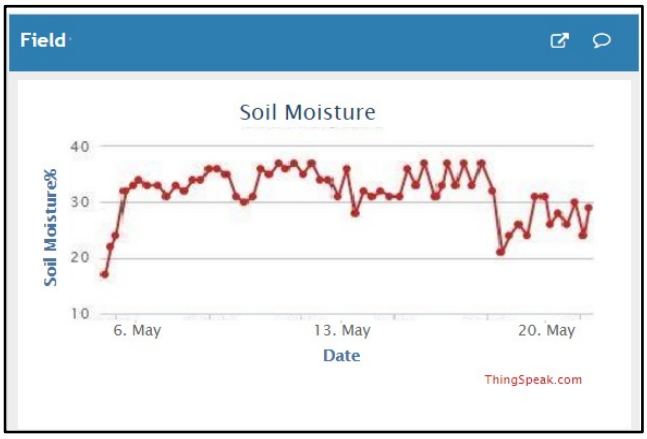


Fig 6. Soil moisture reading

The plot of CO2 vs. time interval is plotted and it is observed that maximum concentration of gas reaches at midway between 3-4 pm.

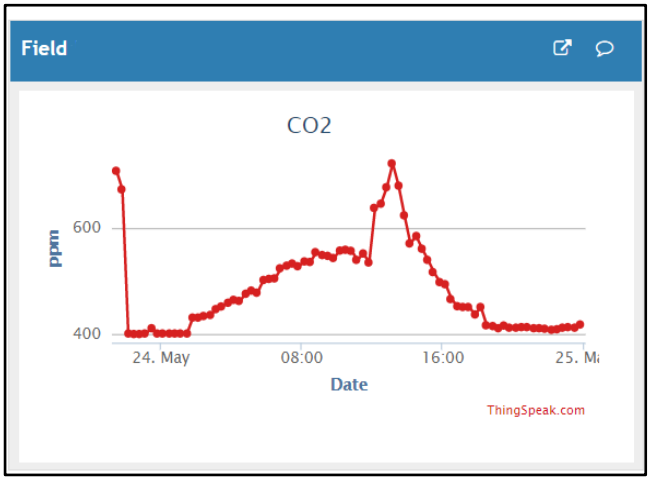


Fig 7. CO2 content in surrounding air

The plot of pH values vs. time interval is plotted and it is concluded that soil is alkaline in nature coz majority of time the value crosses 7.

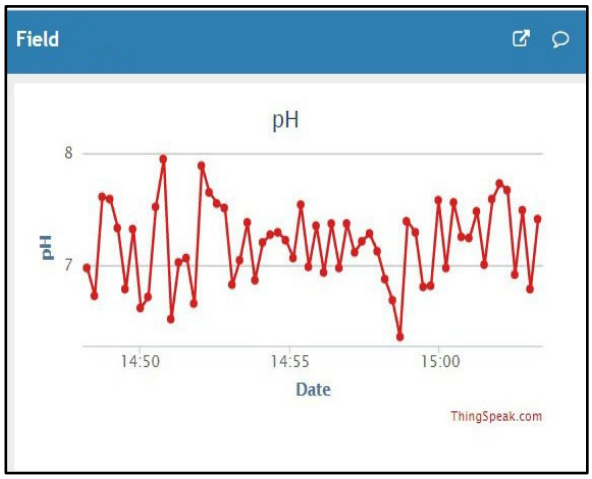


Fig 8. pH of soil

The result of the proposed work done in part-B i.e. visualization of predicted crop production using various data mining techniques mentioned above is shown in the form of plot between actual vs. predicted crop production. Each algorithms are applied on the given datasets and corresponding comparison plot is drawn as shown below.

The plot correspond to SVM is drawn and as SVM gives accurate data for more than 10000 data values peak is observed at this stage.

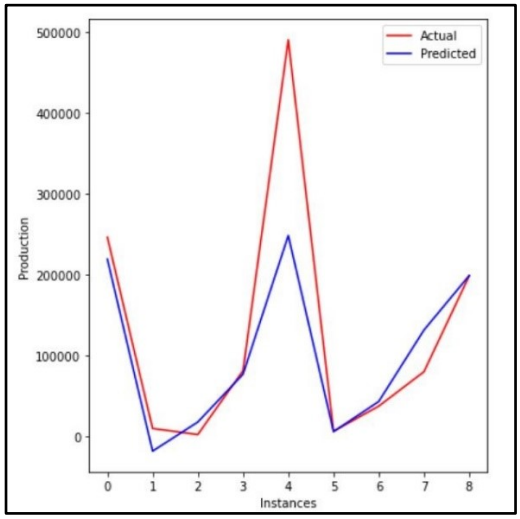


Fig 9. SVM classification- comparison between predicted and actual value of crop production

The plot correspond to KNN is drawn between the actual and predicted crop production. At small value instances the actual and predicted graph is exact replica but it fail to do so at higher values.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:57:29 UTC from IEEE Xplore. Restrictions apply.

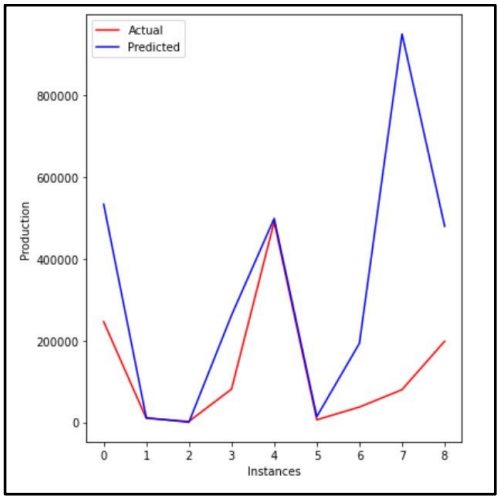


Fig 10. KNN classification- comparison between predicted and actual value of crop production

The plot correspond to Random forest classification which from graph it is concluded that it gives better accurate results on the small data values due to its accuracy in measurement of lower order data values as compared to higher values in actual vs. predicted crop production.

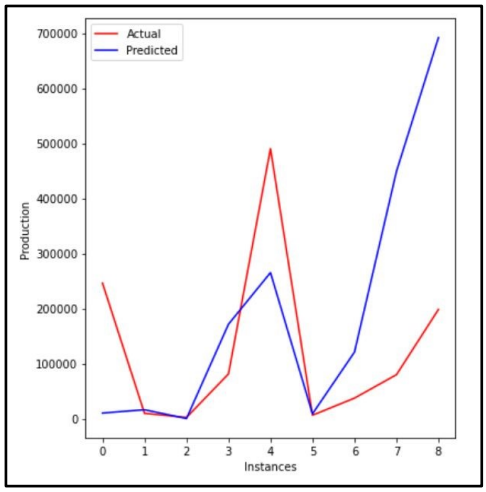


Fig 11. Random Forest- comparison between predicted and actual value of crop production

Based on the results of all three techniques an observation is done showing the accuracy percentage of calculated values with that of theoretical values. The result of observation is thus noted of various data values of dataset and are compared. A total of 11000 data values records are taken for observation.

TABLE I. ACCURACY COMPARISON OF RANDOM FOREST, SVM, KNN

NO OF DATA VALUES USED	Percentage of Accuracy of shown algorithms		
	Random Forest	SVM	KNN
0	0	0	0
2000	70.18	67.13	66.01
3000	70.56	65.4	67.49
4000	70.88	58.34	49.9
5000	68.46	61.58	56.6
6000	55.97	62.92	58.9
7000	63.89	63.01	62.0
8000	58.79	64.79	60.0
9000	60.25	65.25	78.69
10000	59.54	66.23	81.3
11000	56.2	68.8	87.9

In this table it is inferred that on the small number of data the random forest classification is the best to apply for prediction and for large no of data values KNN is recommended.

At the end a custom made webpage is designed having login page and crop prediction page. In crop prediction page after entering all the values of parameters and selecting the type of technique will get the content of Nitrogen, Potassium and phosphorus in soil related to that specific technique.

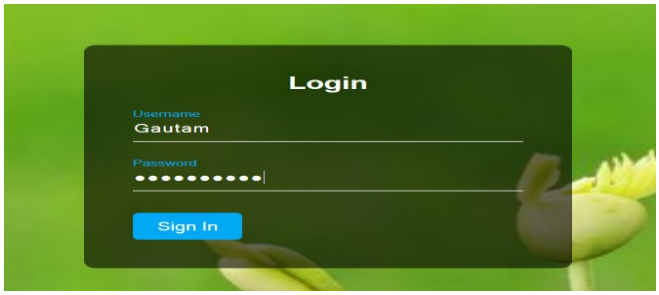


Fig 12. Login page of user defined site

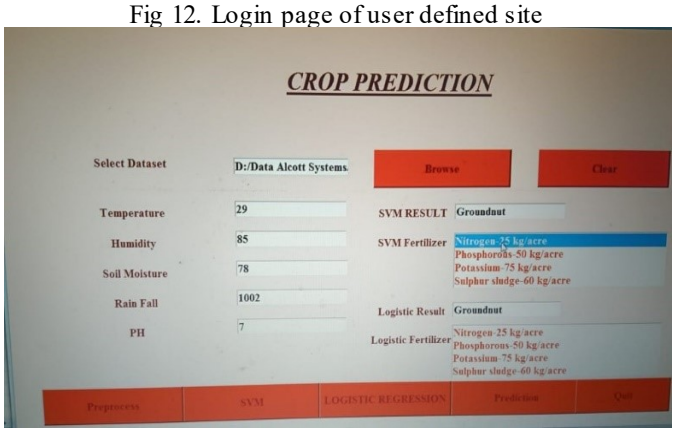


Fig 13. Crop prediction web page

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:57:29 UTC from IEEE Xplore. Restrictions apply.

V. CONCLUSION AND FUTURE SCOPE

This paper showed the transformation of system which is capable of collecting the data from senor node with the help of IoT in the field of agriculture. This system successfully able to sense the data and sending the data locally to thingspeak cloud which is further access by user in his/her custom made website. Data mining techniques are used to predict the crop production in advance so that farmer always got to know the extent of accuracy of their farming techniques. Various methods of regression and classification to get the output and their plot are plotted successfully. In future whether showing the data in custom made website can show the data in mobile app. Further, it detects the disease of the crop by using the concept of image detection. With that suitable type of fertilizer and pesticides is used to eliminate the disease. Artificial Intelligence and neural network could also be used as a regression method that promises to give more precise results.

ACKNOWLEDGMENT

We would like to thank Mrs. Bhavnesh Jaint, for his constant motivation and guidance during the project. Without his timely help and support, the completion of this paper wouldn't have been possible. We would also like to thank the publisher of the research papers from where we have taken references.

REFERENCES

[1] W. Klosgen, J.M. Zytkow, Handbook of data mining and knowledge discovery, Oxford University Press, 2002.

[2] Kiruthika M, Shweta T, Mritunjay O, Kavita S, "Parameter monitoring for the precision agriculture", International Journal Of The Scientific Research and Innovation 2015.

[3] O.P. Uma Maheshwari, M Savitha, "A Study of IoT (Internet of Things) in agriculture", International Journal of Innovative research in computer and communication Engineering.

[4] Dr. N. Suma, S. R. Samson, S. Saranya, G. Shanmugapriya, R. Subhashri " IoT Based Smart Agriculture Monitoring System", International Journal on Recent and Innovation Trends in computing and communication.

[5] Sandip Khot, M. Gaikwad, "Development of cloud based Light Intensity Monitoring system for Green house Using Raspberry Pi" , IEEE International conference on computing communication control and automation (ICCUBEA) Aug 2016

[6] Sheetal V, A Bakshi, Tanvi T, "Green House by using IoT and Cloud computing", IEEE International conference on Recent trends in Electronics, information & communication technology May 2016

[7] Aymen E Khedr, Mona Kadry, Ghada Walid(2015), "Proposed framework for Implementing Data Mining Techniques to enhance Decisions in Agriculture sector Applied case on Food Security Information Center Ministry of Agriculture, Africa", international conference on Communication, management and Information technology (ICCM).

[8] Kuljit Kaur, Kalnwalpreet Singh Atwal "Effect of Temperature and Rainfall on Paddy Yield using data mining ", 2017 7th international conference on Cloud Computing, Data Science and Engineering Confluence.

[9] F. TongKe, "Smart Agriculture Based on Cloud Computing and IoT", in Journal of Convergence Information Technology, Volume 8, Number 2, 2013.

[10] Sivaganesan, D. (2019). DESIGN AND DEVELOPMENT AI-ENABLED EDGE COMPUTING FOR INTELLIGENT-IOT APPLICATIONS. Journal of trends in Computer Science and Smart technology (TCSST), 1(02), 84-94.

[11] Joseph, S. I. T. (2019). SURVEY OF DATA MINING ALGORITHM'S FOR INTELLIGENT COMPUTING SYSTEM. Journal of trends in Computer Science and Smart technology (TCSST), 1(01), 14-24.

[12] Scikit-learn, Machine learning in python retrieved from http: //scikitlearn. Org.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:57:29 UTC from IEEE Xplore. Restrictions apply.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/342877873

Estimation of COVID-19 Prevalence under Closed Population using Capture-Recapture approach

Article *in* International Journal of Science and Research (IJSR) · January 2018

DOI: 10.21275/SR20629043404

CITATIONS
0

READS
18

2 authors, including:



Rakesh Kumar
Delhi Technological University
7 PUBLICATIONS **0** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Performance Measurement View project

Population Statistics View project

All content following this page was uploaded by Rakesh Kumar on 12 July 2020.

The user has requested enhancement of the downloaded file.

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2019): 7.583

Estimation of COVID-19 Prevalence under Closed Population using Capture-Recapture approach

Rakesh Kumar[1], R.S.Mishra[2]

Delhi Technological University, Delhi, India

Abstract: ***The nation-wide lockdown due to COVID-19 created the situation of closed population. This paper presents a general theory for estimating the size of a closed population. The theory is based on a log-linear model, with model parameters representing the natural logarithm of subpopulation size and the natural logarithm of subpopulation capture probabilities. Lastly the paper contains illustration of use of log-linear to closed population.***

Keywords: COVID-19, log-linear, capture-recapture

1. Introduction

Estimating population sizes has long been of interest, from the estimation of the human or ecological population size within regions or countries to the hidden number of corona patients in the state. Total enumeration of the population, for example, via a census, is often infeasible or simply impractical. However, a series of partial enumerations or observations of the population is often possible. This has led to the ideas of capture-recapture methods, which have been extensively used within ecology to estimate the size of wildlife populations, with an associated measure of uncertainty, and are most effectively applied when there are multiple capture occasions. The capture-recapture technique originally developed for animal studies and later has been applied by practitioners and researchers to human populations under the term `multiple-record system' for an arbitrary number of lists and under the term `dual-system especially for two lists. The earliest references to such applications can be traced back to Sekar & Deming (1949) for two samples, Wittes & Sidel (1968) for three samples, Wittes (1974) for four samples and Fienberg (1972) for five samples.

Since state circumstances will vary in terms of disease activity and capacity to respond, caution should be exercised in directly linking severity assessment at a global to national level to actions at the state level. It is likely that information will be limited early in the pandemic while the demand for information simultaneously escalates. If pandemic surveillance is to provide sufficient information and data to assess severity, country/state need to review their existing surveillance procedure and capacity to address the weaknesses to be prepared for pandemic surveillance. Essential components of an effective pandemic corona surveillance system will include: early detection and investigation; comprehensive assessment; and Monitoring. Delhi government launched 5T plan to stop the spread of the novel corona virus. The 5T plan includes Testing, Tracing, Treatment, Teamwork and Tracking based on the South Korean strategy Test, Trace and Isolate. But the step testing requires basis. India's apex medical body Indian Council of Medical Research (ICMR) intensified the random sampling of people who display flu-like symptoms but don't have any history of travel to outbreak zones to determine whether community transmission is taking place. Later ICMR experts aren't convinced about following South Korea's strategy, which has reported more recoveries from Covid-19 than new infections through aggressive testing without lockdowns. Indian experts said they will revise the testing protocol depending on the situation. The ICMR in its study highlighted that Covid-19 containment activities need to be targeted in districts reporting positive cases among Severe Acute Respiratory Infections (SARI) patients and stated that intensifying sentinel surveillance for Covid-19 among SARI patients may be an efficient tool to effectively use resources towards containment and mitigation. Testing: Until now testing was very limited. Only symptomatic people who had history of travel were tested. But now Delhi government plans to aggressively start, testing drives in Delhi areas which are being seen as Corona virus hot spot. Also there is going to be random testing of one lakh people in the next few days. Delhi is the first state to carry out random testing. Second T - Tracing as testing ramps up every single contact of positive cases through cellphone triangulation will be traced. Twenty four hour tabs will be kept on those who are in quarantine to ensure it's being followed. Such focused actions will help in staying at least two steps ahead of virus.
ICMR (Indian Council of Medical Research) conducted a test in few states of country and its research shown that there are many cases of corona virus where it is difficult to find the source i.e. they have no travel history and not get in contact with any such person who has travel history. The Ministry of Home Affairs (MHA) of India stated that the government has adopted a special strategy to fight coronavirus, with specific instructions for the district and state authorities. The strategic approach for COVID-19 containment broadly involves defining the area of operation, applying perimeter control, delineating containment and buffer zones. Besides, there is active search for cases, contact tracing, quarantine, clinical management and awareness generation among public.

Without discussing the methodology, It is estimated by Indian professor Ramanan Laxminarayan, Director and Senior Fellow, The Center for Disease Dynamics, Economics & Policy that by July 2020, 300 to 500 million Indians will get infected and a couple of million Indians can die of corona virus cases and fatalities. A data journalist Rukmini S. Has also presented her observation on projections of corona virus cases concluded that if the rate of COVID-19 spread continues, 30,000 deaths can occur in

Licensed Under Creative Commons Attribution CC BY

India by May 2020 and also mentioned that various variables like lockdowns in India and the enforcement of them are great challenges that statisticians cannot deal with.

Random Sampling

Random sampling is basic to many computing applications in science, statistics, and engineering. Random sampling is used on those occasions when processing the entire dataset is unnecessary and too expensive in terms of response time or resource usage. The savings generated by sampling may arise either from reductions in the cost of retrieving the data from the database or from subsequent "post-processing" of the sample. Sampling is useful for applications which are attempting to estimate some aggregate property of a set of records, such as the total number of records which satisfy some predicate. Thus random sampling is typically used to support statistical analysis of a dataset, either to estimate parameters of interest or for hypothesis testing. Random sampling has three common justifications: reduced database retrieval costs (or time), reduced "post-processing" computational costs, or reduced cost of physical inspections. So, random sampling is a basic sampling technique through which each individual is chosen entirely by chance and each member of the population has an equal probability of being included in the sample. It is the only way to be sure that the sample is not biased and is representative of the population as a whole, and even more complex sampling schemes such as systematic and stratified sampling are based on this basic scheme. ICMR's testing of a small random sample may not be able to capture it. As India is divided its 755 districts into green, orange and red zones. It will entail making sure that green zones are indeed green by deploying testing strategies. Delhi government announced a five-point action plan of "five Ts -- testing, tracing, treatment, team-work and tracking -- and monitoring" to contain the virus spread and said one lakh random rapid anti-body blood tests for COVID-19 tests will be conducted in the city's hotspot areas, the similar plan is followed by other states also.

2. Literature Review

The past literature is reviewed keeping in view for this study is concerned as well as COVID-19. Pandemic is not new for the world, but the lockdown concept is new for country like India experience in year 2020. Estimating population sizes has long been of interest, from the estimation of the human or ecological population size within regions or countries to the hidden number of corona patients in the state. Total enumeration of the population, for example, via a census, is often infeasible or simply impractical. However, a series of partial enumerations or observations of the population is often possible. This has led to the ideas of capture-recapture methods, which have been extensively used within ecology to estimate the size of wildlife populations, with an associated measure of uncertainty, and are most effectively applied when there are multiple capture occasions.

Capture-recapture methods are used in various areas, from epidemiology (Coeli et al., 2000) to zoological studies of populations (Fernandez, 1995) or even in the analysis of the variation of measurement intervals of closed communities of diverse magnitudes (Cormack, 1992). The capture-recapture techniques are used for closed populations, where the effects of birth rate or mortality and migration are not taken into account, in this case it is considered no change in population size. They can also be used for open populations, where population changes caused by birth rate, mortality and migration are taken into account (Cormack, 1989). Sekar and Deming (1949) proposed their research concerned with estimating birth and death rates in an area near Calcutta, India. The research based report of Robles, S. C. et al. (1988) described about usefulness of capture-recapture methods used to estimate completeness at the Ontario Cancer Registry. The method was applied in two fashions; first, using three data sources in a modeling approach and second, using two data sources and standard, simple capture-recapture methods. The modeling approach is more flexible, since several variables that influence cancer registration can be considered and can be used to identify reporting patterns of different data sources. Wittes, J. T (1974) in his paper considered the problem of estimating the total size of a target population from which a study sample has been obtained by merging names from several routinely collected lists. Each list is assumed to be a simple random sample from the target population and all lists are assumed independent. Hook, E. B. (1980) presented his work with goal to compare various methods for estimating the live-birth prevalence of spina bifida (cystica) using Bernoulli census analysis, allowing for under ascertainment in the sources used, and to determine if a stable estimate or at least a useful confidence boundary on the population live-birth prevalence rate could be derived from prenatal diagnosis of spina bifida undertaken as part of a cost-benefit analysis. Brenner, H. (1994) highlighted the drawback of routine systems of disease monitoring, such as morbidity registries, are often prone to specific threats of validity, such as in imperfect diagnoses or imperfect record linkage. Through quantitative assessment of the performance of the two-source capture-recapture method for disease monitoring in the presence of imperfect record linkage. International Working Group for Disease Monitoring and Forecasting (1995) evaluates the potential epidemiologic uses of capture-recapture, which include the primary area of determining disease frequency. Capture-recapture may be a means to effectively "count" new cases (incidence) or count existing cases (prevalence). In addition to generating an estimate of population size, another application of capture-recapture is to assess the costs of ascertainment relative to the degree of accuracy. Fisher N. et al. (1994) worked on routinely collected data from local agencies to provide a reliable estimate of the size and psychiatric morbidity of the homeless population of a given central London by using capture-recapture analysis. A multiple sample, log-linear capture-recapture method was applied to a defined geographical area during 6 months. The method calculates the total homeless population from the sum of the population actually observed and an estimate of the unobserved population. Sheila M. Bird and Ruth King (2018) with changed name of Capture-Recapture Estimation with Multiple Systems Estimation (MSE) advocated that the MSE approach has been preferred when estimating "capture-shy" or hard to-reach populations, including those caught up in the criminal justice system; or homeless; or trafficked; or civilian casualties of war. The paper focuses on whether the uncertainty which qualified their estimates was sufficiently narrow to orient public policy or not, if not, what options

Licensed Under Creative Commons Attribution CC BY

were available and/or taken to reduce the uncertainty or to seek external validation. *Jones H.E.et al.(2016) by using* Capture–recapture analysis recommended for estimating the prevalence of problem drug use or people who inject drugs .The work aim to demonstrate how naive application of Capture–recapture can lead to highly misleading results, and to suggest how the problems might be overcome. The method further validated by a case study of estimating the prevalence of people who inject drugs in Bristol, UK, using Capture–recapture. *Leclerc, P. (2014) already published similar type of work as Jones H.E.et al.(2016) carried but study area was* Island of Montréal, Canada, Using a Six-Source Capture-Recapture Model.

Even after worthiness of this method Stephen N Jarvis et al. (2000) stated with prior proof in his research paper with illustration that there is also serious drawbacks in the use of the mark/recapture technique when applied to injuries. Through example of two sample mark/recapture estimation in an epidemiological setting it is shown to violate virtually all the requirements for valid use of the technique. Further mentioned that very little can be deduced accurately about the scale or characteristics of an unobserved group by the use of mark/recapture applied to two overlapping health event registers. Begon, M. (1983) as a critic to the methods of Jolly he mentioned that there has also been a consistent interest in the method on the part of statisticians, a sporadic interest in field-based critiques of the method, and a considerable tendency, especially since the mid-seventies, to refer to Jolly's paper but make no use of its contents.

3. Capture-Recapture Method and Log-Linear Model

Pierre Simon Laplace (1749-1827) used similar to capture-recapture approach to estimate the size of human population of France. Further Lincoln-Peterson develop a technique with the name capture-recapture

Closed Population: A closed population is one in which the total number of individuals is not changing through births, deaths, immigration or emigration. But in practice, most real populations are not closed, it's a chance the in complete lockdown, the population is definitely closed. An open population is one that is (or could be) changing during the course of study, because of any combination of births, deaths, immigration or emigration. The 'closed captures' data types all assume the population of interest is closed during the sampling period (White *et al*. 1982). Meaning, the models assume that no births or deaths occur and no immigration or emigration occurs. Typically, we refer to a closed population as one that is free of unknown changes in abundance, as we can usually account for known changes.

The assumption is in context of human being is differ from animals in that way that- human being are self conscious and able to report about their illness and checked (tagged), human being always available whenever called for recounting. But this COVID-19 effected the mind and thought of people in such a way that people use to hide themselves from the society and that become a challenging task to the health department to do mapping to identify them from the huge crowd of population, even if it is closed due to lockdown. The 'closed captures' data types all assume the population of interest is closed during the sampling period (White *et al*. 1982). Meaning, the models assume that no births or deaths occur and no immigration or emigration occurs. Typically, we refer to a closed population as one that is free of unknown changes in abundance, as we can usually account for known changes.

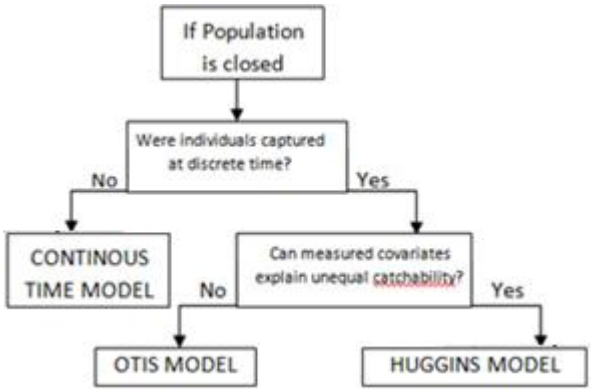


Figure 1: Flow Chart of the methods of CAPTURE-RECAPTURE

Consider a closed population with size N and the people are indexed by 1,2,. . . , N in any arbitrary fashion. In a usual t-sample capture-recapture experiment, people are captured and marked in each trapping sample (or trapping occasion), and we only notice in each sample whether each person was captured or not without recording the sequential order or the time when it was captured. Thus the capture history can be expressed as a N x t matrix (X_{ij}), where X_{ij} = I [the i^{th} person was caught in the j^{th} sample], i = 1, 2,. . . , N, j = 1, 2.. . , t. Extensive models have been discussed for the above discrete-time type data.

If the capture time and the sequential order for each individual can be recorded, then a continuous-time model is needed. Anderson and Wilson (1989) have done some simulation to compare the existing estimators for continuous-time models. Becker (1984), Becker and Heyde (1990) proposed an estimation procedure using a martingale approach.

In the complete lockdown situation the whole population is considered as closed population and the new birth in this period is neglected because the record is unavailable for this period. A closed population is one in which the total number of individuals is not changing through births, deaths, immigrations and emigrations and all these were not in assumption in corona period. In practice, most real populations are not closed. Sometimes, however, the changes over the time period of interest are small enough that the assumption of closure is a reasonable approximation, and the effects of violating that assumption are minimal. For that reason, the analysis of capture-recapture data from closed populations continues to be a topic of interest to researchers.
However closure is subdivided into two components: (i) geographic and (ii) demographic. The geographic closure of population by a boundary, which limits the population, and demographic is about birth, death, immigration and emigration.

Licensed Under Creative Commons Attribution CC BY

The models of Otis et al.(1978) denoted M_0, M_t, M_b, M_h, M_{bh}, M_{tb}, M_{th}, and M_{tbh}.

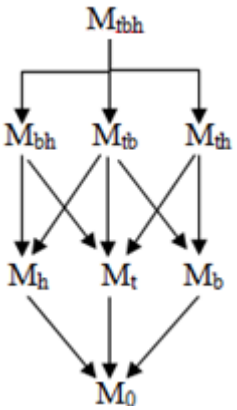
Model M_0 assumes a single common probability of capture. The subscripts t, b, and h correspond to temporal variation in common probability of capture, behavioral response to common probability of capture, and individual response to capture, respectively. Multiple subscripts indicate a combination of capture responses.

The above mentioned eight models summarised by symbol, source of variation in capture probabilities , and the associated estimators, if any , program CAPTURE uses theses names for estimators.

Table 1: Discrete-Time Models and Associated Appropriate Estimator with Log Linear form

Model	Source of variation in capture probabilities	Appropriate estimator	Log Linear form
M_0	None	Null (All estimators given below are valid)	$log\ (P_{ij}) = \alpha$
M_t	Time	Darroch (1958)	$log\ (P_{ij}) = \mu_j$
M_b	Behaviour	Zippin (1956)	$log\ (P_{ij}) = \alpha + \gamma\ Y_{ij}\ (\alpha_i = \alpha)$
M_h	Heterogeneity	Jackknife (1978)	$log\ (P_{ij}) = \alpha_i$
M_{bh}	behaviour heterogeneity	Generalized Removal (Otis et al. 1978)	$log\ (P_{ij}) = \alpha_i + \gamma\ Y_{ij}$
M_{tb}	time behaviour	None	$log\ (P_{ij}) = \mu_j + \gamma\ Y_{ij}$
M_{th}	time heterogeneity	None	$log\ (P_{ij}) = \alpha_i + \mu_j$
M_{tbh}	time behaviour heterogeneity	None	$log\ (P_{ij}) = \alpha_i + \mu_j + \gamma\ Y_{ij}$

Relation of the model



Each arrow points from one model to an immediate special case of the model. Test of assumptions are based partly on these relations among the model

Figure 2: Relation of the Model

There are also capture-recapture models available for the open population like Jolly-Seber model, Modern open population model, Multi State Model, Target Recovery Model and Joint modeling of live and dead recovery, with different assumptions. These models are widely used by researchers to count the animal populations as well as human also.

Log-Linear Models

The log-linear model approach was originally proposed by Fienberg (1972) and has been widely used in many applications (Cormack 1989; Agresti 1994; Evans, Bonett, and McDonald 1994; Coull and Agresti 1999; IWGDMF 1995). In this approach, the data are regarded as a form of an incomplete 2^t contingency table (t is the number of lists) for which the cell corresponding to those individuals uncaptured by any of the samples is missing. Then various log-linear models are fitted to the observed cells and the chosen model is projected onto the unobserved cell by assuming that there is no t-sample interaction. We use the three-sample case as an illustration. There are seven observed cells: z_{111}, z_{110}, z_{101}, z_{100}, z_{011}, z_{010}, and z_{001}, where z_{100} is the number of individuals captured in sample 1 only and z_{011} is the number of individuals caught in samples 2 and 3 but not in sample 1. A similar interpretation pertains to other capture histories. The missing cell z_{000} denotes the number of missed individuals. The log-linear approach models the logarithm of the expected value of each observable category. The most general log-linear model for the three-sample case is

$$log\ E(z_{ijk}) = u + u_1 I(i = 1) + u_2 I(j = 1) + \underline{u}_3 I(k = 1) + u_{12} I(i = j = 1) + u_{13} I(i = k = 1) + u_{23} I(j = k = 1) + u_{123} I(i = j = k = 1); \quad \text{----------(1)}$$

where I(A) is an indicator function for event A. The above is a reparameterization of the eight expected values. We have seven observed cells, whereas there are eight parameters in the right-hand side of equation (1). Therefore, a commonly adopted assumption is that there is no three-sample interaction term, i.e., $u_{123} = 0$, which allows an extrapolation formula for the missing cases using the fitted values of the observable cells. The independent model with main effect terms is equivalent to model M_t in Table 1. The log-linear models corresponding to the multiplicative models M_b and M_{tb} are discussed in Cormack (1989), who also showed how to t various models using the GLIM package (GLIM is a statistical package(marketed by NAG) which analyses generalised linear models by iterated reweighted least squares). Many authors stated that random sampling is required in capture studies. This assumption stems from ball and urn experiments, in which marked and unmarked balls are shaken completely, and a random sample is taken at the end of each sampling occasion. Traditional sampling methods include procedures for drawing random samples. Uses of the procedures require knowledge of the sampling probabilities (for finite populations). Deliberate control over the elements to be sampled is required. It is unrealistic to think that if any COVID patient is captured (marked) in one corner of the trapping grid may be captured subsequently in the opposite corner of the boundary. The concept of random sampling does not apply to situations assumed in Models M_b, M_h, M_{bh}, M_{tb}, M_{th}, and M_{tbh}. The goal of these models is to provide an analysis of the sampled data in the face of behavioural response and heterogeneity, both of which are contrary to the traditional role of random sampling.

4. Data Source For Illustration

According to ICMR the data of the following people should be considered for COVID-19 testing.

1) Symptomatic (fever, sore throat, running nose, dyspnea, etc) individuals who have travelled to affected countries

Licensed Under Creative Commons Attribution CC BY

(China, Hong Kong, Japan, South Korea, Singapore, Iran, Italy etc.)
2) Close contacts of confirmed positive cases of COVID-19 infection.
3) All individuals evacuated and quarantined from Wuhan, China and Diamond Princess ship, Japan.

It includes information on COVID-19 service activities by private and public hospitals and community health services from 25th March 2020 to 31st May 2020 (Lock-down period in India). This study may be use to estimate the number of people who were infected by the corona virus (COVI-19) in an outbreak in whole India and for more accuracy study area may be carried out at state wise. For example if the study area is Delhi selected with approximately 16787941 population in 11 different districts. Our data are restricted to census records of government. In total, 271 cases were reported from the following three sources. (1) A-list: records based on Symptomatic (fever, sore throat, running nose, dyspnea, etc) individuals who have travelled to affected countries (China, Hong Kong, Japan, South Korea, Singapore, Iran, Italy etc.) . There were 135 ascertained cases. (2) B-List: records provided by Close contacts of confirmed positive cases of COVID-19 infection. The records included 122 cases reported by doctors in local hospitals. (3) C-list: records based on All individuals evacuated and quarantined from Wuhan, China and Diamond Princess ship, Japan.. A total of 126 cases were in the C-list.

Table 2: Data on corona virus infection

A	B	C	Data
1	1	1	Z_{111}= 28
1	1	0	Z_{110}= 21
1	0	1	Z_{101}= 17
1	0	0	Z_{100}= 69
0	1	1	Z_{011}= 18
0	1	0	Z_{010}= 55
0	0	1	Z_{001}= 63
0	0	0	-----

The data are given in Table 2 Several log-linear models were fitted to these data. Except for the saturated model, the log-linear models, which do not take into account heterogeneity (i.e. models (AC, B), (BC, A), (AB, C), (AB, BC), (AB, AC) and (BC, AC)), do not fit the data well. The quasi-symmetric and three partial quasi-symmetric models (see Darroch *et al*., 1993), which incorporate heterogeneity effects, fit the data well. The four adequate models produce approximately the same estimates, 1300, with an approximate estimated s.e. of 520. This relatively large estimated i.e. shows the data are actually insufficient to model the heterogeneity effect.

Estimation Procedure

$$\hat{C} = 1 - \frac{1}{t} \sum_{k=1}^{t} \frac{S_k}{n_k}$$

Equation for estimator of sample coverage ----------(2)

The singletons cannot contain any overlap information. Therefore, the sample coverage can be Interpreted as a measure of overlap. When neither dependence exists, it follows from the definition of C that $C = D/N$, where D denotes the average of overlapped cases, i.e.

$$D = \frac{1}{t} \sum_{k=1}^{t} \{ \sum_{k=1}^{N} I [\sum_{j \neq k} Xij > 0] = M - \frac{1}{t} \sum_{k=1}^{t} S_k$$

----------(3)

Here, M denotes the number of different cases identified in at least one of the lists. Hence, in the independent case, a valid estimator is

$$N_0 = D/\hat{C} \qquad \text{----------(4)}$$

the sample coverage is estimated to be $\hat{C} = 51.27\%$ and the average of the overlapped cases is D= 208.67. If we incorrectly assume that the three samples are independent, then from (4) an estimate would give $N_0 = D/\hat{C} = 407$, which is slightly larger than the estimate of 388 based on an independent log-linear model. We feel these data, with estimated sample coverage of 51%, do not contain enough information to correct for undercount. Thus, this example presents a very valuable data set with the advantage of a known true parameter.

5. Discussion

We discuss first the extent to which this study provides evidence for or against the use of mark/recapture techniques in the epidemiological setting. We then consider the relationship of these findings to other similar studies. To what extent then does the present study meet the "requirements" for the mark/ recapture method? As far as possible, the study has attempted to define a "closed" population—Otis condition "a". Note here that "closed" does not infer that the COVID-19 patient population of Delhi is unchanged—rather that the reference population for individual cases is the same for both lists. In this study the reference population is COVID-19 patients from Delhi addresses with following people should be considered for COVID-19 testing.
1) Symptomatic (fever, sore throat, running nose, dyspnea, etc) individuals who have travelled to affected countries (China, Hong Kong, Japan, South Korea, Singapore, Iran, Italy etc.)
2) Close contacts of confirmed positive cases of COVID-19 infection.
3) All individuals evacuated and quarantined from Wuhan, China and Diamond Princess ship, Japan.

Indian Council of Medical Research (ICMR) affirms, while there is no readily available test kit to detect coronavirus infection, doctors and researchers go by a series of tests to confirm the infection i.e. Swab test, Nasal aspirate and tracheal aspirate test. Indian Council of Medical Research (ICMR) suggests that anyone with a recent travel history to COVID-19 affected countries or someone who has been in close contact with infected patients is also advised screening. Despite the extended time over which our lists are compiled, an individual's availability for both lists is almost concurrent. The crucial assumption, that an individual in the population from which one list is a sample, is also in the population from which the other list is a sample, is thus

Licensed Under Creative Commons Attribution CC BY

satisfied. In any area of epidemiology or, in fact, any human monitoring system involving population estimates of the number of people who have a certain characteristic, capture-recapture should be considered a complementary or an alternate approach to traditional counting technology. One application of capture-recapture, in addition to generating an estimate of population size, is to assess the costs of ascertainment relative to the degree of accuracy obtained. A primary goal of monitoring is the ability to obtain accurate and timely incidence rates at a reasonable cost. The more expensive it is to accurately estimate the counts of disease, the less likely it is to assess that disease and the smaller the geographic coverage. When setting out to determine incidence rates, therefore, one should first determine the amount of precision desired, as greater degrees of precision are likely to be more expensive. The appropriate degree of precision required for estimating incidence rates will vary by the use(s) to which the data are put. There are no right or wrong decisions as to how many samples should be examined, or the degree of ascertainment needed, as this is a function of both the acceptable degree of precision and the amount of resources available to carry it out.

We strongly believe that with any disease monitoring system, whether communicable or non-communicable, there should be the means to formally evaluate the degree of undercount. With capture-recapture, we have this technology. All disease monitoring systems should be able to estimate the proportion of the diseases diagnosed in the community, region, or nation that were identified by the monitoring system. Without this information, it is difficult, if not impossible, to determine the degree to which the rates of disease are compromised by missing cases. Capture-recapture is not a panacea. However, it offers public health officials the opportunity to get closer to the truth if the methods are used appropriately. There is much to be learned in the development of capture-recapture models for use in epidemiology. The assumptions need to be assessed and the models tested. Epidemiologists and biostatisticians working with human disease need to continue the dialogue with those who have developed the approaches using capture-recapture. Much remains to be learned, and there is great potential. As indicated by the IWGDMF (1995), the log-linear model approach has many advantages as follows: (a) all models are under a unified framework; (b) model selection can be easily implemented and carried out in a flexible fashion, based on data and guided by prior information; (c) tests are available for comparing models; (d) dependence can be incorporated by adding proper interactions; and (e) all inference is within the mainstream of statistical data analysis. An untestable assumption involved is that the highest-order interaction does not exist. The IWGDMF commented that the existence of heterogeneity in three-list data might result in the lack of a reliable estimate. Another concern is that two equally fitted models might produce quite different estimates. As the number of lists increases, the number of adequate models increases rapidly and thus model selection causes further problems.

References

[1] Anderson, D. R., and Wilson, K. R., 1989: Capture-recapture population estimation using continuous time models. Progress Report.
[2] Becker, N. G., 1984: Estimating population size from capture-recapture experiments in continuous time. Austral. J. Statist. 26, 1-7.
[3] Becker, N. G., and Heyde, C. C., 1990: Estimating population size from multiple recapture experiments. Stochastic Process Appl. 36, 77-83.
[4] Bird SM, King R., "Multiple Systems Estimation (or Capture-Recapture Estimation) to Inform Public Policy", *Annual review of statistics and its application. 2018 Mar; 5: 95-118*, doi: 10.1146/annurev- statistics-031017-100641
[5] Begon, M. (1983). Abuses of Mathematical Techniques in Ecology: Applications of Jolly's Capture- Recapture Method. Oikos, 40(1), 155. doi:10.2307/3544213
[6] Coeli, C.M., Veras, R.P. and Coutinho, E.S.F., 2000. Metodologia de captura-recaptura: uma opção para a vigilância das doenças não transmissíveis na população idosa. Cadernos de Saúde Pública, vol. 16, no. 4, pp. 1071-1082. PMid:11175530. http://dx.doi.org/10.1590/S0102-311X2000000400025.
[7] Cormack, R.M., 1992. Interval estimation for mark-recapture studies of closed populations. Biometrics, vol. 48, no. 2, pp. 567-576. PMid:1637979. http://dx.doi.org/10.2307/2532310.
[8] Cormack, R.M., 1989. Log-linear models for capture-recapture. Biometrics, vol. 45, no. 2, pp. 395-413. http://dx.doi. org/10.2307/2531485.
[9] Fernandez, F.A.S., 1995. Métodos para estimativas de parâmetros populacionais por captura, marcação e recaptura. In: P.R. PERES-NETO, J.L. Valentin and F.A.S. Fernandez, eds. Programa de pós-graduação em ecologia. Rio de Janeiro: Instituto de Biologia, UFRJ, pp. 1-26. http://dx.doi.org/10.4257/ oeco.1995.0201.01.
[10] Fienberg, S. E. (1972) The multiple recapture census for closed populations and incomplete 2^k contingency tables, Biometrika, 59, pp. 591-603.
[11] Hook, E. B., Albright, S. G., & Cross, P. K. (1980). Use of bernoulli census and log-linear methods for estimating the prevalence of spina bifida in livebirths and the completeness of vital record reports in New York state. American Journal of Epidemiology, *112(6), 750– 758.* doi:10.1093/oxfordjournals.aje.a113047
[12] H. Brenner (1994). Application of Capture-Recapture Methods for Disease Monitoring: Potential Effects of Imperfect Record Linkage. Methods Inf Med 1994; 33(05): 502-506. DOI: 10.1055/s-0038-1635053.
[13] International Working Group for Disease Monitoring and Forecasting (1995). Capture-Recapture and Multiple-Record Systems Estimation II: Applications in Human Diseases. American Journal of Epidemiology, 142(10), 1059–1068. doi:10.1093/oxfordjournals.aje.a117559
[14] Jones H.E., Welton NJ, Ades AE, Pierce M, Davies W, Coleman B, Millar T, Hickman M. Problem drug use prevalence estimation revisited: heterogeneity in capture–recapture and the role of external evidence.

Licensed Under Creative Commons Attribution CC BY

Addiction (Abingdon, England). 2015 Dec 28; 111(3): 438-447. doi: 10.1111/add.13222
[15] Jarvis SN, Lowe PJ, Avery A, *et al.*Children are not goldfish—mark/recapture techniques and their application to injury data Injury Prevention 2000;**6:**46-50.
[16] Leclerc P, Vandal AC, Fall A, Bruneau J, Roy É, Brissette S, Archibald C, Arruda N, Morissette C., "Estimating the Size of the Population of Persons Who Inject Drugs in the Island of Montréal, Canada, Using a Six-Source Capture-Recapture Model", Drug Alcohol Depend. 2014 Sep 1; 142:174-80. Epub 2014 Jun 24. doi: 10.1016/j.drugalcdep.2014.06.022.
[17] Nigel Fisher, Stuart W Turner, Robert Pugh, Colin Taylor, Estimating numbers of homeless and homeless mentally ill people in north east Westminster by using capture-recapture analysis , BMJ. 1994 Jan 1; 308(6920): 27–30.doi: 10.1136/bmj.308.6920.27
[18] Robles, S. C., Marrett, L. D., Aileen Clarke, E., & Risch, H. A. (1988). An application of capture-recapture methods to the estimation of completeness of cancer registration. Journal of Clinical Epidemiology, 41(5), 495–501. doi:10.1016/0895-4356(88)90052-2
[19] Sekar, C. & Deming W. E. (1949) On a method of estimating birth and death rates and the extent of registration, Journal of the American Statistical Association, 44, pp. 101-115.
[20] Wittes, J. T. & Sidel, V. W. (1968) A generalization of the simple capture-recapture model with applications to epidemiological research, Journal of Chronic Diseases, 21, pp. 287-301.
[21] Wittes, J. T. (1974) Applications of a multinomial capture-recapture method to epidemiological data, Journal of the American Statistical Association, 69, pp. 93-97.
[22] White, G.C., Anderson, D. R., Burnham, K. P., and Otis, D. L. (1982) Capture-recapture and removal methods for sampling closed populations. Los Alamos National Laboratory Publication LA-8787-NERP. Los Alamos, NM.
[23] White, G.C., Anderson, D. R., Burnham, K. P., and Otis, D. L. (1982) Capture-recapture and removal methods for sampling closed populations. Los Alamos National Laboratory Publication LA-8787- NERP. Los Alamos, NM.
[24] Wittes, J. T., Colton, T., & Sidel, V. W. (1974). Capture-recapture methods for assessing the completeness of case ascertainment when using multiple information sources. *Journal of Chronic Diseases, 27(1-2), 25–36.* doi:10.1016/0021-9681(74)90005-8

Author Profile



Rakesh Kumar (rakeshtech1@gmail.com) Graduated in Industrial Engineering and Master's in Mechanical Engineering with specialization in Computer Integrated Manufacturing from India. He also obtained Post Graduate Diploma in IPR. He is currently an Assistant Professor and research scholar at Delhi Technological University in Department of Mechanical. His research interest includes Total Quality Management, Reliability Engineering, Operations and Supply chain management, Fuzzy Logic, AI and Decision Science and Intellectual Property Rights. http://orcid.org/0000-0001-5060-1259



R.S. Mishra (Professor) (rsmishra@dce.ac.in) received the Ph.D. degree in Mechanical engineering from IIT Delhi, Indiaand currently at Delhi Technological University serving since 1997. He having 150 research papers in peer reviewed in International journals and 75 research papers in proceedings of International and National conferences. Prior to joining DTU, he served 09 years at different engineering colleges of India, viz, REC Hamirpur, PAU Ludhiana (Punjab)and HAU, Hisar (Haryana.). His research area includes Green Technology, Refrigeration & Air Conditioning, Metal Cutting Technology, Total Quality Management, Solar Energy Technology, Bio Fuels, power Plant engineering.

Licensed Under Creative Commons Attribution CC BY

Materials Today: Proceedings xxx (xxxx) xxx

Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

Experimental studies of vapour compression refrigeration system with eco-Friendly primary refrigerant and brine mixed with nano particles as secondary refrigerant

Uma Shankar Prasad [a,*], R.S. Mishra [b], R.K. Das [a]

[a] IIT(ISM), Dhanbad, India
[b] Delhi Technological University (DTU), Delhi, India

ARTICLE INFO

Article history:
Received 3 March 2020
Received in revised form 3 June 2020
Accepted 4 June 2020
Available online xxxx

Keywords:
Alternative refrigerants
Eco friendly refrigerants
Nano-refrigerant
Energy exergy analysis
Irreversibility analysis

ABSTRACT

This paper is to study the thermal analysis of vapor compression refrigeration system using R134a in primary circuit and Al_2O3-Water based nano-fluids in secondary circuit. The study provides the information of the secondary fluids input conditions, Performance parameters of vapour compression refrigeration system using eco-friendly refrigerants and without nano materials, Effect of overall evaporator heat transfer coefficient using different nano particles mixed with R718 in the secondary evaporator circuit and R134a in primary circuit, the compressor power consumption and the system overall energy performance. Such an analysis with the help of engineering equation solver software (EES) tool can be conveniently useful to compare the thermal performance of different nano-particles (i.e., Al_2O3, TiO_2) based nano fluid as a secondary fluid in a vapor compression refrigeration system.
© 2020 Elsevier Ltd. All rights reserved.
Selection and Peer-review under responsibility of the scientific committee of the 2nd International Conference on Nanoscience and Nanotechnology.

1. Introduction

Nanofluids are a new class of fluids engineered by dispersing nanometer-sized materials (nanoparticles, nanofibers, nanotubes, nanowires, nanorods, nanosheet, or droplets) in base fluids. In other words, nanofluids are nanoscale colloidal suspensions containing condensed nanomaterials. Nanofluids have been found to possess enhanced thermophysical properties such as thermal conductivity, thermal diffusivity, viscosity, and convective heat transfer coefficients compared to those of base fluids like oil or water. It has demonstrated great potential applications in many fields. For a two-phase system, there are some important issues we have to face. One of the most important issues is the stability of nanofluids, and it remains a big challenge to achieve the desired stability of nanofluids (Fig. 1).

1.1. Brief literature review

Lots of research work has been done for replacing "old" refrigerants with "new" refrigerants Ahamed et al. [1] emphasized on use of hydrocarbons and mixture compressor shows much higher exergy destruction as compared to rest of components in the vapour compression refrigeration system and this exergy destruction can be minimized by using of nanofluid and nanolubricants in compressor. Anand [2] did detailed exergy analysis of 2 ton of refrigeration capacity window air conditioning test rig with R22 as working fluid and reached to the conclusions , that irreversibility in system components will be highest when the system is 100% charged and lowest when 25% charged and irreversibility in compressor is highest among system. Arora et al. [3] developed numerical model of actual vapour compression refrigeration system with liquid vapour heat exchanger and did energy and exergy analysis on the same in the specific temperature range of evaporator and condenser and concluded that R502 is the best refrigerant compared to R404A and R507A and compressor is the worst component and liquid vapour heat exchanger is best component of the system in case of exergy transfer. Bolaji et al. [4] had done experimentally comparative analysis of R32, R152a and R134a refrigerants in vapour compression refrigerator and concluded that R32 shows lowest performance whereas R134a and R152a showing nearly same performance but best performance was obtained of system using R152a. Chopra et.al [5] carried out analysis of energetic and exergetic based comparison Multiple Evaporators with

* Corresponding author.
E-mail address: ushankar@piemr.edu.in (U. Shankar Prasad).

https://doi.org/10.1016/j.matpr.2020.06.083
2214-7853/© 2020 Elsevier Ltd. All rights reserved.
Selection and Peer-review under responsibility of the scientific committee of the 2nd International Conference on Nanoscience and Nanotechnology.

Please cite this article as: U. Shankar Prasad, R. S. Mishra and R. K. Das, Experimental studies of vapour compression refrigeration system with eco-Friendly primary refrigerant and brine mixed with nano particles as secondary refrigerant, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2020.06.083



Fig. 1. (a) VCRS setup.

compound compression and flash intercooler with individual and multiple throttle valves and find out the effect of flash intercooler on second law performances. Han et al. [6] Under different working conditions experimental results revealed that there could be replacement of R407C in vapour compression refrigeration system having rotor compressor with mixture of R32/R125/R161 showing higher COP, less pressure ratio and slightly high discharge compressor temperature without any modification in the same system. Mishra et al. [7–9] performed numerical analysis of vapour compression refrigeration system using R134a, R143a, R152a, R404A, R410A, R502 and R507A, and discussed the effect of evaporator temperature, degree of subcooling at condenser outlet, superheating of evaporator outlet, vapour liquid heat exchanger effectiveness and degree of condenser temperature on COP and exergetic efficiency. They reported that evaporator and condenser temperature have significant effect on both COP and exergetic efficiency and also found that R134a has the better performance while R404a has poor performance in all respect (Table 1).

1.2. Problem statement

From the above literature review following points are left by different authors-

- Different arrangements & combinations of refrigerants with nano particles are still left for research.
- Al_2O_3 , TiO_2 & CuO response is better and they may improve the COP of refrigerator by introducing a secondary circuit in the evaporator unit.

But still there is vast area is to explore for performance improvement in VCRS at low cost and low input (Table 2).

Table 1
Performance parameters of vapour compression refrigeration system using eco-friendly refrigerants and without nano materials.

Refrigerants	COP	ED_Ratio	ETA_Second
R134a	2.973	2.447	0.2901
R404a	2.705	3.158	0.2717
R407a	2.754	2.333	0.30

Table 2
Effect of overall evaporator heat transfer coefficient using different nano particles mixed with R718 in the secondary evaporator circuit and R134a in primary circuit.

Nano particles	Overall heat transfer coefficient (U_e)	Difference in water temperature (°C)	Difference in brine temperature (°C)
Copper oxide	1372.14	15.71	11.58
Copper	1367.18	14.55	11.44
Al_2O_3	1318.0	14.62	11.5
Ti_2O_3	1242.43	14.43	11.32
Without nano particles	691.04	12.24	9.29

1.3. Materials and methodology

To improve thermal performance of vapour compression air conditioning systems by improving:

- First law efficiency- According to first law of thermodynamic energetic efficiency /COP is defined as the ratio of net air conditioning effect to the per unit power consumed. First law analysis restricted to calculate only coefficient of performance of the systems and,
- Second law efficiency- The concept of exergy was given by second law of thermodynamics. Second law efficiency is the exergy of the heat abstracted in to the evaporators from the space to be cooled and exergy of fuel is actual compressor work input.

In this paper our main focus on performance of three suspended nano particles i.e., Al_2O_3, TiO_2, & CuO in water based refrigerant in secondary circuit and in primary circuit we are using eco-friendly refrigerants (Table 3).

Table 3
Effect of overall evaporator heat transfer coefficient using different nano particles mixed with R718 in the secondary evaporator circuit and R134a in primary circuit.

Nano particles	Overall heat transfer coefficient (U_e)	% improvement
Copper oxide	1325.7	91.41
Copper	1324.25	91.4099
Al_2O_3	1318.70	90.6077
Ti_2O_3	1242.43	79.6
Without nano particles	691.04	—

Please cite this article as: U. Shankar Prasad, R. S. Mishra and R. K. Das, Experimental studies of vapour compression refrigeration system with eco-Friendly primary refrigerant and brine mixed with nano particles as secondary refrigerant, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2020.06.083

1.4. Results

In this paper, first law and second law analysis of vapour compression refrigeration systems using multiple evaporators and single compressor and single expansion valve with eco-friendly refrigerants in the system and secondary refrigerant (water used in secondary circuit with and without nano-particles mixed with water used as refrigerant) have been presented with the help of EES software results.

1.5. Conclusion

Various nano-particles with R134a refrigerant can be used as an excellent refrigerant to improve the heat transfer characteristics in a refrigeration system. A successful model has been designed and the basic theoretical heat transfer analysis of the refrigeration system has been done.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

I Mr. Uma Shankar Prasad, Research scholor, IIT(ISM), Dhanbad 149 and Assistant Professor at PIEMR, Indore gratefully acknowledge my Research guide Prof. R.K das, Mechanical Department IIT (ISM), Dhanbad and Prof. R.S Mishra, Mechanical Department DTU, New Delhi for providing funds for the research to review this article.

References

[1] U. Ahamed, R. Saidur, H.H. Masjuki, A review on exergy analysis of vapor compression refrigeration system, Int J Renewable and sustainable energy reviews 15 (2011) 1593–1600.
[2] S. Anand, S.K. Tyagi, Exergy analysis and experimental study of a vapour compression refrigeration cycle, Int J Therm Anal Calorim, 2012, 110:961-971.
[3] AkhileshArora, S.C. Kaushik, Theoretical analysis of a vapour compression refrigeration system with R502, R404A and R507A, Int J Refrigeration. 2008 (31) (2008) 998–1005.
[4] B.O. Bolaji, M.A. Akintunde, T.O. Falade, Comparative analysis of performance of three ozone-friends HFC refrigerants in a vapor compression refrigerator, Int J Sustainable Energy & Environment, Vol-2, pp.61-64, 2011.
[5] Chopra et.al, Energetic and Exergetic Based Comparison Multiple Evaporators with Compound Compression and Flash Intercooler with Individual or Multiple Throttle Valves International Journal of Advance Research & Innovations (1), 2013, 73-81.
[6] Han, et.al, Cycle performance study on R32/ R125/ R161 as an alternative refrigerant to R407C. Int J Applied Thermal Engineering, 2007, 27:2559-2565.
[7] Mishra et.al, Methods for improving thermal performances of vapour compression Refrigeration system using eleven ecofriendly refrigerants, ISTE conference on "Technological Universities and Institutions in New Knowledge Age: Future Perspectives and Action plan, ISTE, 149, 9, Delhi Technological University, 2013.
[8] Mishra, Irriversibility analysis of multi-evaporators vapour compression refrigeration systems using new and refrigerants: R134a, R290, R600, R600a, R1234yf,R502, R404a and R152a and R12, R502, Int. J. Adv. Res. Innovations 1 (2013) 180–193.
[9] Mishra, Thermodynamic performance evaluation of multiple evaporators, single compressor, single expansion valve and liquid vapour heat exchanger in vapour compression refrigeration systems using thirteen ecofriendly refrigerants for reducing global warming and ozone depletion, Int. J. Adv. Res. Innovations 1 (2014) 163–171.

Please cite this article as: U. Shankar Prasad, R. S. Mishra and R. K. Das, Experimental studies of vapour compression refrigeration system with eco-Friendly primary refrigerant and brine mixed with nano particles as secondary refrigerant, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2020.06.083

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/335459256

Experimental Validation of Fully Informed Particle Swarm Optimization Tuned Multi-Loop L-PID Controllers for Stabilization of Gantry Crane System

Chapter · January 2020

DOI: 10.1007/978-981-13-8618-3_61

CITATION
1

READS
32

6 authors, including:



Sudarshan K Valluru
Delhi College of Engineering (Now Delhi Technological University) Govt.of NCT Delhi
37 PUBLICATIONS **78** CITATIONS
SEE PROFILE



Daksh Dobhal
Delhi Technological University
4 PUBLICATIONS **1** CITATION
SEE PROFILE



Kumar Kartikeya
Delhi Technological University
4 PUBLICATIONS **1** CITATION
SEE PROFILE



Manpreet Kaur
Bow Valley College
9 PUBLICATIONS **1** CITATION
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Design of Multi-Loop L-PID and NL-PID Controllers: An Experimental Validation View project

meta algorithms application to nonlinear dynamical systems View project

All content following this page was uploaded by Sudarshan K Valluru on 14 July 2020.

The user has requested enhancement of the downloaded file.

Experimental Validation of Fully Informed Particle Swarm Optimization Tuned Multi-Loop L-PID Controllers for Stabilization of Gantry Crane System



Sudarshan K. Valluru, Madhusudan Singh, Daksh Dobhal, Kumar Kartikeya, Manpreet Kaur and Arnav Goel

Abstract Linear PID controllers are commonly used as an electrical component to decrease the error between anticipated set value and actual measured values for control of various benchmarked systems. The Multi-Loop Linear PID (ML-PID) controller, gives a robust and efficient performance in most of the situations. This paper presents the implementation of linear PID controller to stabilize and control the Gantry Crane System. Optimal performance is obtained for a few specific combinations of the proportional, integral, and derivative gains, which makes it essential to tune these values through Optimization techniques. The Fully Informed Particle Swarm Optimizer (FIPSO) is used to tune the gain values of the ML-PID controlled Gantry Crane System. These values are validated experimentally, and obtained results prove that the FIPSO tuned multi-loop linear PID controller quickly stabilizes the system subjected to external disturbances.

Keywords Multi-Loop linear PID (ML-PID) · Fully Informed Particle Swarm Optimization (FIPSO) · Gantry crane system

S. K. Valluru (✉) · M. Singh · D. Dobhal · K. Kartikeya · M. Kaur · A. Goel
Department of Electrical Engineering, Incubation Center for Control Dynamical Systems and Computation, Delhi Technological University, New Delhi 110042, Delhi, India
e-mail: sudarshan_valluru@dce.ac.in

M. Singh
e-mail: madhusudan@dce.ac.in

D. Dobhal
e-mail: dakshdobhal@gmail.com

K. Kartikeya
e-mail: kkartikeya1998@gmail.com

M. Kaur
e-mail: manpreetdtu.26@gmail.com

A. Goel
e-mail: arnav_bt2k16@dtu.ac.in

© Springer Nature Singapore Pte Ltd. 2020

S. Choudhury et al. (eds.), *Intelligent Communication, Control and Devices*,
Advances in Intelligent Systems and Computing 989,
https://doi.org/10.1007/978-981-13-8618-3_61

1 Introduction

The Linear PID is one of the most widely used electrical component in the systems industry [1, 5]. It provides efficient solutions while being computationally inexpensive and simple to implement and simulate. The Linear PID can control and modulate responses of benchmarked dynamical systems [11, 12]. Though, if the system has nonlinearities, the controller may produce unsatisfactory response including high peak overshoot and large settling time, but still, it has limitations like optimality and tuning [10]. Thus, the difficulty of tuning is to be challenging particularly if the multi PID controllers have more than three objectives. The three terms in the controller namely Proportional, Integral, and Derivative can be varied to get an optimal response which is calculated according to the external disturbances by evolutionary computation [8].

In recent years, Particle Swarm Optimization has gained much attention of researchers as a bio-inspired algorithm for the optimization techniques especially for the multimodal and multi-output functions [6, 13]. Although the standard PSO showed better results and faster convergence time to reach global optima for unimodal objective functions, it suffers from falling of local minima problem for nonlinear, multimodal objective functions [6]. This paper addresses the simultaneous tuning of ML-PID controllers using a new variant of FIPSO [2, 4], and the obtained gain values are tested to control and stabilize the Gantry Crane System [3].

The rest of the paper is organized as following sections. Section 2 describes the design of a Multi-Loop Linear PID Controller. Mathematical modeling of the Gantry Crane system is presented in Sect. 3. The tuning of the Controller using FIPSO is presented in Sect. 4. Experimental results and discussions are given in Sect. 5, and finally, Sect. 6 outlines the conclusion of the paper.

2 Design of Multi-Loop L-PID Controllers

The linear proportional integral derivative controller operates upon the error term $\varepsilon(t)$, the mismatch between the real time output and reference input signals to generate the correction signal $\alpha(t)$. The correction signal is obtained as the sum of the error, derivative of the error, and integral of the error multiplied by their respective coefficients. Usually, the three coefficients are denoted by P, I, and D respectively. Hence the expression for the correction signal is obtained as (1):

$$\alpha(t) = K_p \varepsilon(t) + K_i \int_{t1}^{t2} \varepsilon(t) \cdot dt + K_d \frac{d\varepsilon(t)}{dt} \tag{1}$$

The primary task is determining the optimal combinations of the coefficients which yield the best performance. The term proportional to the error reduces the system response time as it is increased till a maximum allowed value above which

sustained oscillations are introduced in the system response. In reducing the steady state error, the major part is played by the integral term. The derivative term leads to a reduction of the corrected signal if the error is increasing.

3 Gantry Crane System

The crane system, in spite of its simple structure, is a complex system owing to its nonlinear nature [7]. Because of this, it is considered to be one of the under accutated benchmark of control systems. If it is assumed that the payload is a lumped-mass distribution rather than a distributed-mass system, the gantry crane can be modeled as a simple pendulum. Utilizing the assumptions of payload being a point mass distribution, massless cable, the cable of fixed length and not flexible and neglecting the damping coefficient of the cable, the free body diagram of the crane and functional diagrams are shown in Fig. 1 and Fig. 2 respectively. Consider M, m, l, b, and I to be the mass of the trolley, mass of the pendulum bob, length of the cable, the coefficient of friction of cart, and the MOI of the bob respectively.

The variables x and θ are the position of trolley along the belt and angular position of the pendulum. The equations of motion obtained for the system are (2) and (3):

$$F = (M+m)\cdot \acute{x} + b\acute{x} - mlsin\theta \cdot \acute{\theta}^2 + mlcos\theta \cdot \acute{\theta} \tag{2}$$

$$0 = \left(I + ml^2\right)\cdot \acute{\theta} - mglsin\theta + mlcos\theta \cdot \acute{x} \tag{3}$$

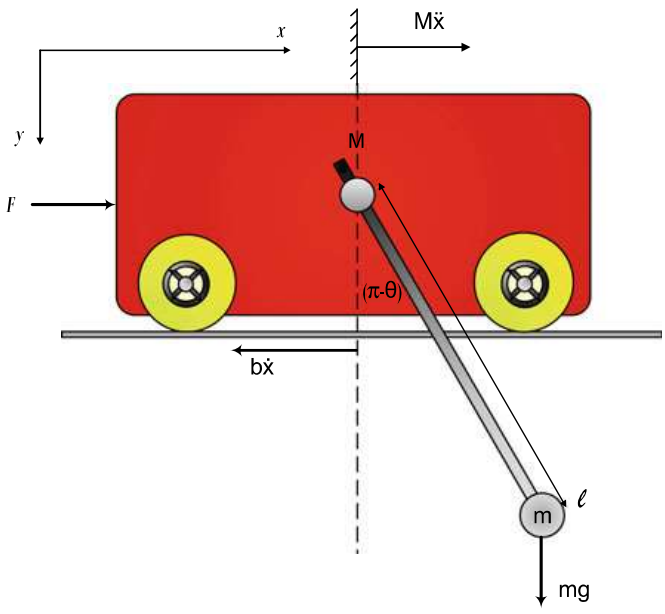


Fig. 1 FBD of gantry crane system

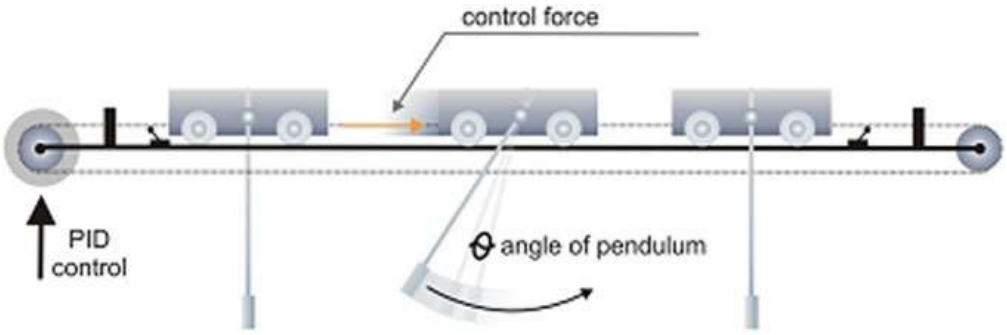


Fig. 2 Functional Diagram

Table 1 Gantry crane system parameters

S. No.	Parameters	Value
1	Mass of the cart (M)	2.4 kg
2	Mass of the pole (m)	0.23 kg
3	Cart friction coefficient (b)	0.05 Ns/m
4	Length of the pole (l)	0.32 m
5	Pendulum moment of inertia (I)	0.099 kg m^2
6	Gravity constant (g)	9.81 ms^{-2}
7	Range of operating voltage of D.C. servo motor (V)	±24 V

For $\theta \rightarrow 0$: $sin\theta \approx \theta$, $cos\theta \approx 1$, $\acute{\theta}^2 \approx 0$

For the simpler understanding of carrying out the stability analysis of the dynamics of the system, the linearized system Eqs. (4) and (5) as

$$F = (M + m) \cdot \acute{x} + b\acute{x} + ml \cdot \acute{\theta} \tag{4}$$

$$0 = \left(I + ml^2\right) \cdot \acute{\theta} - mgl\theta + ml \cdot \acute{x} \tag{5}$$

The system parameters as specified in Eqs. (4) and (5) are given in Table 1.

4 Tuning of ML-PID by FIPSO Algorithm

The PSO algorithm mimics the flocking of birds and the schooling of fishes to find the optimal value of the given target function [9]. The Fully Informed Particle Swarm Optimization takes into account the influence on each particle in a swarm by the mean of its nearest neighbor. Therefore, it has the advantage of fast convergence, escaping

the tendency to fall in local minima and ability to optimize complex functions when compared to Standard PSO algorithm.

The important variables associated with particles in the swarm in FIPSO are x_i (position), v_i (instant velocity), Pb_i (Present Personal best), and Gb_i (Global Position best). Each particle is updated on the basis of above parameters in the above-defined pop size (iteration) for which cost Function (F_{cost}) is to be minimized. For the Gantry Crane Benchmark, x_i can be comprehended as (6)

$$x_i = [K_{P1}, K_{I1}, K_{D1}, K_{P2}, K_{I2}, K_{D2}] \tag{6}$$

where x is the position of the particle, K_{P1}, K_{I1}, K_{D1} are PID constants for the cart's position controller and K_{P2}, K_{I2}, K_{D2} are for the angle controller of the gantry crane system. Particles in a swarm were initialized using the Eq. (7)

$$x_i = unifrnd(PosLL_i, PosUL_i, npop, 1) \tag{7}$$

where $unifrnd$ function generates uniform random numbers in the range ($PosLL_j$,$PosUL_j$). $PosLL$ And $PosUL$ are the lower and upper decision boundary variables respectively, $npop$ is the population (swarm) size and jth represents one of the PID constants shown in above equation.

The updated velocity and position of each particle in an iteration can be calculated as in Eqs. (8) and (9)

$$v_{i+1} = w * v_i + c_1 * (Pb_i - x_i) + c_2 * (Gb_i - x_i) + chi * (x_i - \mathrm{NN}_i) \tag{8}$$

$$x_{i+1} = x_i + v_{i+1} \tag{9}$$

where NN_i is the Nearest Neighbor, defined as the mean of the personal best position of the nearest five particles w is the inertial weight, c_1, c_2 are personal and global learning coefficients respectively and chi is the constriction coefficient. Finally, the innovative Global Best and Personal Best are updated as (10) and (11):

$$Pb_{i+1} = \begin{matrix} \{Pb_i, & Fcost(x_i) > Fcost(Pb_i) \\ x_i, & otherwise\} \end{matrix} \tag{10}$$

$$Gb_{i+1} = \min\{Fcost(Pb_{i+1})\} \tag{11}$$

4.1 Implementation of FIPSO for ML-PID Tuning

The overall FIPSO pseudo code to tune the ML-PID controller for gantry system is summarized as follows:

```
  FCost- call the function whose variables are to be
  tuned
  nVar- no. of variables to be tuned
  NoVar- row vector defining size of the decision
  variable matrix
  PosLL- minimum values of decision variables
  PosUL- maximum value of decision variables
  VelUL-maximum value of velocity
  VelLL- minimum value of velocity
  Setting required PSO parameters:
  Gen- maximum no of iterations to be undergone by the
  code
  PopSize- size of the swarm
  Km- No. of nearest neighbours considered
  NN- Best position closest of neighbours
  Initialisation of swarm parameters
  global best- array of size NoVar containing zeros
  fori=1: PopSize
      initialize position of particle assigning it random
      value;
      set velocity of particle zero;
      evaluate cost function for particle;
      set personal best position and cost equal to
      initial position and cost for each particle;
       if personal best cost of particle i> global best
       cost
          global best=personal best of i;
      end
  end

  Main Loop of PSO
  for it=1: Gen
      fori=1: PopSize
          update velocity using FIPSO innovative vector
          formula;
          apply velocity limits so that velocity remains
          between VelLL and VelUL;
          update position;
      apply position limits so that position remains
      betweenPosLl and PosUL;
          calculate value of cost function;
          if cost of i< personal best cost of i
              personal best position of i= position of i;
              personal best cost of i= cost of i;
          end
          if global best cost < best cost of i
              global best= personal best of i;
          end
      end
end
  apply position limits so that position remains between
  VarMin and VarMax;
  calculate value of cost function;
  if cost of i< personal best cost of i
  personal best position of i= position of i;
  personal best cost of i= cost of i;
  end
  if global best cost < best cost of i
  global best= personal best of i;
  end
  end
end
```

Table 2 ML-PID gain parameters

S. No	Gain co-efficient	Optimal value
1	K_{P1}	4.8590
2	K_{I1}	0.4918
3	K_{D1}	4.0352
4	K_{P2}	4.9101
5	K_{I2}	0.0973
6	K_{D2}	0.3000

The swarm size, iteration, w, c_1, c_2, chi values used in the experiment for the optimization of ML-PID controller for the gantry crane system are 50, 30, 0.7298, 1.4962, 1.4962, 0.7298 respectively. Two main objectives for control of Gantry Crane System, namely, swing minimization of pendulum and tracking of input voltage signal, were achieved using two separate PID controllers. The optimal gain values obtained at 30th iteration with minimum cost function is given in Table 2.

5 Experimental Results

The cart was required to track a sinusoidal trajectory of amplitude 0.3 V and a frequency of 0.1 Hz. A sample time of 100 s was chosen to operate the system. Small manual perturbations of increasing magnitude were given to pendulum for every 20s intervals. Cart Position Response, Pendulum Angle Response, and Control Signal depicting the performance of ML-PID were obtained and validated extensively for stabilizing the Gantry Crane System as shown in Fig. 3. The measure of performance

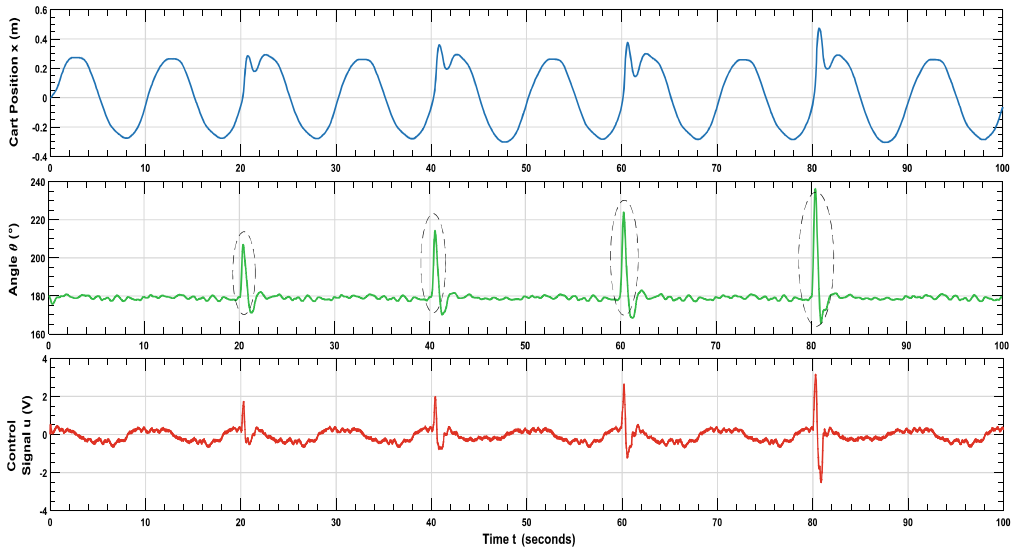


Fig. 3 Performance of Multi-Loop L-PID controller

Table 3 Pendulum angle stabilization response

S. No.	Instant of disturbance initiation (s)	Instant of stabilization (s)	Amplitude of disturbance (°)	Time taken to stabilize (s)
1	20.11	21.57	26.9	1.46
2	40.4	41.7	34.3	1.3
3	59.96	61.58	43.9	1.72
4	80.06	81.67	56.3	1.61

was time taken for stabilization after an external disturbance was given. The instant of stabilization was determined according to the 3% settling time stability condition.

The ML-PID responds to stabilization of the pendulum angle against the external noise/disturbance is given in Table 3. It is observed that as the magnitude of disturbance increases, the time taken for stabilization also increases.

6 Conclusions

In this paper, ML-PID controllers were designed, and the optimal gain parameters are obtained by using the FIPSO algorithm. The controllers were validated to stabilize the benchmarked gantry crane system. It is observed that the ML-PID controller brings the pendulum to the desired position under the perturbations in an average time of 1.522 s.

Acknowledgements The authors would like to thank the Govt. of India and Govt. NCT Delhi for providing funding through Delhi Technological University under TEQIP-II to procure Digital Pendulum System Experimental Setup.

References

1. Astrom, K.J., Hagglund, T.: Benchmark systems for PID control. In: IFAC Proceedings on Digital Control: Past, Present and Future of PlD Control, pp. 165–166 (2000)
2. Gao, Y., et al.: Multi-objective opposition-based learning fully informed particle swarm optimizer with favour ranking. In: Proceedings—2013 IEEE International Conference on Granular Computing, GrC 2013, pp. 114–119 (2013)
3. Jaafar, H.I., Mohamed, Z., Abidin, A.F.Z., Ab Ghani, Z.: PSO-tuned PID controller for a nonlinear gantry crane system. In: IEEE International Conference on Control System, Computing and Engineering, pp. 23–25, Penang, Malaysia (2012)
4. Mendes, R., et al.: The fully informed particle swarm: simpler, maybe better. IEEE Trans. Evol. Comput. **8**(3), 204–210 (2004)
5. Passino, K.M.: Intelligent control for autonomous systems. IEEE Spectr. **32**(6), 55–62 (1995)
6. Poli, R., et al.: Particle swarm optimization an overview. Swarm Intell. **1**(1), 33–57 (2007)

7. Ramli, L., et al.: Control strategies for crane systems: a comprehensive review. Mech. Syst. Signal Process. **95**, 1–23 (2017)
8. Reynoso Meza, G., et al.: Controller tuning by means of multi-objective optimization algorithms: a global tuning framework. IEEE Trans. Control Syst. Technol. **21**(2), 445–458 (2013)
9. Valluru, S.K., et al.: Comparative analysis of PID, NARMA L-2 and PSO tuned PID controllers for nonlinear dynamical system. J. Autom. Syst. Eng. **9**(2), 94–108 (2015)
10. Valluru, S.K.: Introduction to Neural Networks, Fuzzy Logic & Genetic Algorithms. Jaico Publishing House, Mumbai (2011)
11. Valluru, S.K., Singh, M.: Metaheuristic tuning of linear and nonlinear PID controllers to nonlinear mass spring damper system. Int. J. Appl. Eng. Res. **12**(10), 2320–2328 (2017)
12. Valluru, S.K., Singh, M.: Stabilization of nonlinear inverted pendulum system using MOGA and APSO tuned nonlinear PID controller. Cogent Eng. **4**(1), 1–15 (2017)
13. Zhao, J., et al.: Application of particle swarm optimization algorithm on robust PID controller tuning. In: Wang, L., Chen, K., Ong, Y.S. (eds.) Advances in Natural Computation, pp. 948–957. Springer, Berlin, Heidelberg (2005)

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Filter Based Scalable Blockchain for Domestic Internet of Things

Aditya Shyam Bazari
Department of Computer Engineering
Delhi Technological University
New Delhi, India
adityabazari_bt2k16@dtu.ac.in

Aditya Singh
Department of Computer Engineering
Delhi Technological University
New Delhi, India
adityasingh_bt2k16@dtu.ac.in

Abdul Ahad Khan
Department of Computer Engineering
Delhi Technological University
New Delhi, India
abdul_bt2k16@dtu.ac.in

Rajni Jindal
Department of Computer Engineering
Delhi Technological University
New Delhi, India
rajnijindal@dce.ac.in

***Abstract*—With the advancements in technology, the ease of interconnectedness among devices has increased manifold, leading to the widespread usage of Internet of Things. Internet of Things has also reached our homes, often referred to as domestic Internet of Things. However, the security aspect of domestic Internet of Things has largely been under question as the increase in inter-device communication renders the system more vulnerable to adversaries. Largely popular blockchain technology is being extensively researched for integration into the Internet of Things framework in order to improve the security aspect of the framework. Blockchain, being a cryptographically linked set of data, has a few barriers which prevent it from being successfully integrated to Internet of Things. One of the major barrier is the high computational requirements and time latency associated with it. This work tries to address this research gap and proposes a novel scalable blockchain optimization for domestic Internet of Things. The proposed blockchain model uses a flow based filtering technique as an added security layer to facilitate the scenario. This work then evaluates the performance of the proposed model in various scenarios and compares it with that of traditional blockchain. The work presents a largely encompassing evaluation, explanation and assessment of the proposed model.**

***Index Terms*—blockchain, internet of things, flow policies, IoT security, scalable blockchain**

I. Introduction

Blockchain, with the potential to revolutionize modern technological communication, is being rapidly amalgamated into a number of modern architectures such as smart-grid [1], healthcare [2], fintech [3] etc. By primarily providing transparency, security and decentralization to the host architecture, blockchain improves the overall functioning of the network. One such framework in which the applicability of Blockchain is being optimistically considered is the Internet of Things [4]. This paper addresses the scenario of Blockchain in domestically deployed Internet of Things and proposes a feasible novel blockchain communication methodology.

Traditionally, domestic Internet of Things have comprised of a simple permission-based security model, in which, after a user grants the IoT system permission to use a particular data, it can use the data in every possible way [5]. No further insights are usually provided into how the data is being used or the online sinks for data transfer. Such a scenario could possibly result in a massive security breach, compromising a user's private data, such as security camera footage, thermostat readings, schedule notes etc. Such a scenario necessitates additional security mechanisms to secure domestic Internet of Things [6].

Since the popularization of bitcoin and the conceptualization of blockchain technology, extensive research is underway to effectively integrate blockchain communication to Internet of Things [4]. This is due to the possibility that blockchain can potentially improve the communication nuances involved in Internet of Things [20]. Cryptographic links between blocks of data, in blockchain, impart enhanced integrity and immutability [7] to the system, thus adding on to the security layer. The distributed framework of blockchain provide decentralization [7] to the system, thus avoiding single points of failure. Blockchain works in a transparent manner [8], with every transactional record present throughout the system, thus suitable for untrusted environments. However, there are certain hurdles to efficient deployment of blockchain in Internet of Things, which translate to the increased resource consumption and added latency in the system. Domestic IoT devices, often being compact, with certain specialized tasks, do not contain adequate computational power to sustain cryptographic mining, necessitated in traditional blockchain [9]. As the cryptographic hashes grow more and more complex, the temporal aspect of the transactional processes rise, thus causing a large delay and forcing the system to lag. Addressing these issues, this paper proposes a modification for blockchain, in context of domestic Internet of Things, which outperforms the existing blockchain systems in terms of resource consumption and latency, while preserving the security aspects of the system. A detailed analysis of the comparison of performance is also

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:47:47 UTC from IEEE Xplore. Restrictions apply.

presented to illustrate the improvement achieved.

The remainder of the paper is organized as follows: Section II explains the background, Section III presents the proposed approach and Section IV reports the results. Conclusions are presented in Section V.

II. Background

Internet of Things is an interconnected network of devices communicating with one another for the exchange of some sort of information. Various devices can be part of the network such as smartphones, sensors, hubs, etc. These devices exchange information with each other to complete a particular task. Information is transferred, processed and analysed among devices in the network. It is not a necessity that the Internet of Things is always a part of the Internet as the name suggests. A few of the major components which define the overall functioning of the infrastructure are:

- Devices: Connecting devices are the basic building blocks of Internet of Things. These devices can be sensors or actuators. Different devices such as mobile camera, accelerometer, proximity detector, pressure sensors, etc can be the part of IoT.
- Connection: The basic blocks, devices in Internet of Things are connected with each other and with the Internet via some means. They may be connected through standard Wi-Fi, Bluetooth, wired LAN (Local Area Network).
- Data Processing: Data Processing is done by the software. It can be identifying a person using a camera or measuring the temperature of the room. The software processes data and gives appropriate results to the user.
- User Interface: User Interface is responsible to show the results processed by the applications to the user in a very suitable manner. It can be showing continuous results over a mobile application or giving an alert via notification on mobile phone to the user.

The process begins with the sensors and chips that are embedded in the physical things that surround us, each of which transmits valuable data. All these devices having sensors are connected with one another. There are IoT platforms which store the relevant data transmitted by the connected devices. IoT platform analyses the data received from various devices. After analysis, important data is transferred to the real time applications. The application will then show the results of analysis to the user. In this way, IoT helps us to interact with machines and devices.

With sophistication of technology, security has become a cardinal issue in modern day Internet of Things [9]. The devices in IoT are connected with each other and the Internet. The data exchange taking place between these devices over the network must be safe and secure. The inter-connected nature of this setup makes it exceptionally prone to attacks such as:

- Cryptanalysis Attacks [10]: In cryptanalysis attacks, the focus is to find the encryption key and to break the encryption. It refers to the act of breaking ciphers. The attacker will study and analyse the ciphers to break the encryption and will break into the system.
- Software Attacks [11]: In software attacks, viruses, trojan horse programs or worms are used to inject malicious code into the system. The malicious code will then destroy the software. An attack on software of IoT device may lead the attacker to steal the user's personal credentials.
- Network Attacks [11]: IoT is vulnerable to network attacks because the transmission of data takes place in a broadcast manner. As soon as an attacker hacks a device in the network, he can easily break into other devices present in the same network. Hence, it is important to keep the network safe. Network attacks can be active or passive.
  - Passive Attack: In passive network attacks, attacker attacks the functioning components of the network. The attacker attacks the network in a concealed manner and try to gather important information of the user. Traffic analysis, node destruction, node malfunctioning, node outage can be some passive attacks.
  - Active Attack: In active network attacks, the attacker tries to make the network services complex and destroy the function and operation of the network. Jamming, flooding, Denial of Service (DoS), wormhole, blackhole, sinkhole are some active attacks.
- Denial-of-Service (DoS) Attack [12]: In DoS attack, the attacker makes the IoT devices and services unavailable. It is very difficult to detect DoS attacks. The attacker makes unnecessary requests to the services which reduce the capacity of the network. The attacker generally tries to flood the centralized resources with requests such that it is unavailable for the other nodes in the network. In Distributed Denial-of-Service (DDoS) Attack, multiple attackers from multiple sources try to flood the requests to a single resource. It can neither detect the attack beforehand nor can stop it by just blocking a single source as it is originated by multiple sources.

Blockchain is an inter-connected list of time-stamped series of secure and immutable records which is distributed in nature [14]. As highlighted in the previous section, integration of blockchain communication into IoT infrastructures effectively enhances the security aspect of IoT infrastructures and protects the system from the above-mentioned threats [4]. Each and every block in this list or chain uses cryptographic methods to stay connected. All these blocks together provide security to each other and also prevents any form of modification to the data. This chain has no central authority. Due to absence of any central power, this distributed ledger is open to each and everyone to see and add. It is by its very nature a transparent entity and every node involved in it is accountable for its actions. The blockchain is a simple yet ingenious way of passing information from system X to system Y in a highly safe and secure manner. One party creates an initialization

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:47:47 UTC from IEEE Xplore. Restrictions apply.

block, it is then verified by millions of other distributed systems [13]. After the verification, the block is added to the chain which consists of previously added blocks.

The three primary factors that impart this high security to the blockchain setup are:

- Decentralization [7]: The system follows a decentralized computational approach in which every involved entity performs some computation locally. This system eliminates a single point of failure and increases the efficiency.
- Transparency [8]: The system is transparent as the blockchain contains every record in an encrypted manner and a record of it is present in each node. This increases the accountability in the system and enhances the security.
- Immutability [7]: Immutability of this system ensures that the encapsulated data is not tampered with. The hashing ensures that any change in the underlying data renders the entire blockchain invalid.

Despite the benefits of amalgamating blockchain to IoT, the integration is not usually feasible in real world scenarios [19]. This is due to the high latency and computational constraints that accompany the setup [9]. This work, in the following section, attempts to address this research gap and proposes an optimized filtering technique to enhance blockchain performance in domestic Internet of Things.

III. Proposed Model

Modern day Internet of Things are prone to security breaches as the devices store a huge variety of crucial personal and institutional data [9]. Several attacks, such as DDoS (Denial of Service), Side Channel Attacks or Man in the Middle attack can cause immense harm by breach of data or rendering the devices useless. For instance, in 2016, Dyn, an American DNS provider faced a large scale DDoS attack, which originated from millions of IP addresses. Several of these IP addresses belonged to IoT devices such as sensors, printers, routers, monitors etc which were infected with Mirai Malware [15]. Similar circumstances necessitate the improvement of security mechanisms in these devices. Integration of blockchain technology to the Internet of Things may provide a solution, as explained in the earlier sections. However, a major obstacle in the adoption of blockchain by IoT devices is the high computational requirements of blockchain. IoT devices are usually small, with limited resources and computation at their disposal. This work tries to close this gap in the case of domestic Internet of Things by providing a case specific solution.

In traditional blockchain scenarios which have been integrated in the Internet of Things, every transaction initiated in the system, whether genuine or not, is being competed for by miners to validate [14]. The process of validation of a transaction is extremely bulky in nature and thus, utilizes a huge chunk of the computational power, thereby, increasing the processing time and resource usage [16]. This work designs a preliminary selective mechanism to identify the predicted authenticity of a transaction before the mining initiates. This mechanism eliminates a set of transactions before the mining process. The blockchain model uses the selective mechanism for the filtering of transactions, thus eliminating the need for validating a number of transactions which do not fit the selection criteria. This subsequently saves computational time of the system and thus, aids the scalability of blockchain in domestic Internet of Things.

The selectivity of transactions in the proposed blockchain model is based on the idea of flow policies [17] [18]. Flow Policies, in simple words are the underlying principles in a system that govern the flow of data. In domestic Internet of Things, the flow pattern of data files between devices are fairly static in nature, in the sense that any particular device has a predefined set of data flow paths. Under normal circumstances, the devices do not usually defy these predefined pathways and almost always confirm to them. For instance, a thermostat might send its data to a hub processor in the network, but it will never have to send the collected data to a bitmap image processor in the same network under normal circumstances. Similarly, a security camera might send a jpg file to a bitmap processor in the network, but might never have to send it directly to an untrusted online sink, while securely functioning. These predefined path structures in domestic Internet of Things have been researched previously and have also been successfully implemented in traditional security architectures. This paper leverages this attribute of domestic Internet of Things and amalgamates it with blockchain to improve the security architecture of the system, while maintaining the scalability of the associated blockchain framework in terms of time and resource consumption.

Every transaction necessitates the involvement of at least two devices. In the presented model, every device has an associated list of flow policies, defining which devices can send or receive data from them. In order to identify if the transaction is in accordance with the user defined flow policies, an intersection is performed between the associated flow policies of the involved devices and the resulting set of flow policies are used to evaluate the authenticity.

In the presented model, the user defines a set of flow rules for every device in the network, i.e. which devices can send in or receive data from that particular device. Every transaction, upon initiation by a sender is flooded into the blockchain network. In the network, every miner has the flow policies of the network. Upon receiving a transactional data packet, duly stamped with the sender's and receiver's IPs, the miner first checks the fulfilment of the associated flow policies. If the transaction is in accordance to the defined flow policies of both the sender and the receiver, the miner proceeds to the validation process. The miners then compete to validate the transaction and mine a block. This block is then added to the blockchain and the transaction is complete. If the transaction violates the flow policies, the data packet is dropped and the transaction cancelled. The flow policies in this model, thus, act like a primary security layer to filter out invalid transactions which do not follow the rudimentary flow patterns acceptable to the user. For a flow policy violation, no major computation is performed in this model, saving a certain

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:47:47 UTC from IEEE Xplore. Restrictions apply.

amount of computational power and time. In scenarios when the system is under attack, this filter thus considerably reduces the chances of the system being overwhelmed by adversaries generating fake requests in the system.

In real world scenarios, under normal circumstances, devices usually function without any interference, following those user-defined paths. During an attack by an adversary, a burst of faulty transactions of data are initiated with the intention of interrupting the functioning of the model. If processing those faulty transactions consume too much of the system resources or time, the system starts to fall apart and eventually, fails. The proposed model protects an IoT establishment from such scenarios in which a system is under constant attack by a particular device or application, initiating multiple bogus transactions to create bottlenecks in the system and slow it down, ultimately resulting in a system failure.

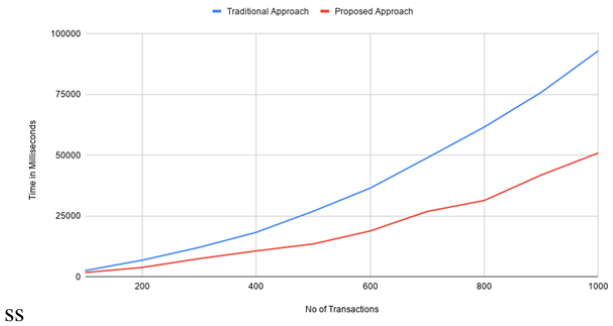
IV. Results

A simulation environment is developed to conduct these runs and only generalized average scenarios are taken into consideration in order to account for the usual trend observed. However, as pointed out earlier, in real world scenarios, the occurrence of fraudulent transactions are often concentrated in short periods of time in which a malicious entity attempts to impair the system. In usual circumstances, the system functions without the presence of any malicious entities without any fraudulent transactions. This work presents scenarios encompassing all possible scenarios.

The first set of simulations are performed on a randomized set of data transactions in the underlying IoT ecosystem in which the flow policies are defined in a probabilistic manner. The flow policy for every pair of device has a defined probability which determines if that particular transactional flow will be valid or not. In real world scenarios, the user directly interacts with the system to explicitly state the flow policies. Thus, the probabilistic method used in this work is merely an attempt to simulate the real world user interaction. The system then generates completely randomized data transactions in the IoT simulation. Blockchain Nodes are subsequently invoked to initiate these transactions. Depending upon the percentage of faulty transactions, the performance evaluation changes.

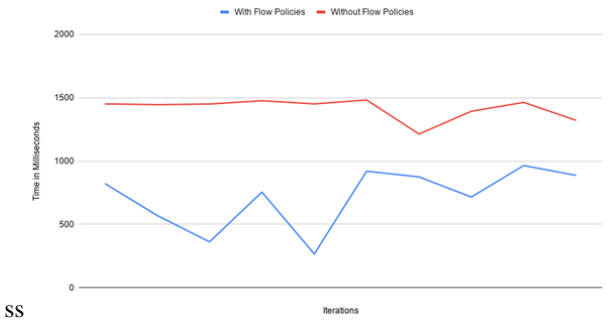
In the particular case of Fig. 1, the probability for every pair of node was set to be 0.5 and since the transactions initiated were randomized, the scenario translated to almost 50% of the transactions being faulty. As visible in the graph presented, the proposed method significantly outperforms the existing approach as almost half the transactions are invalidated by the flow policies, thus, saving the time otherwise required in mining. The presented graph represents an averaged out scenario of multiple iterations of simulation, with data values being reset at every iteration.

The “difficulty” in these blockchain simulations represent a quantized value of the string length of the initial cryptographic challenge which is required to be solved in order to block a mine and validate transactions. In simple words, “difficulty” of the blockchain system is representative of the difficulty of the



SS

Fig. 1. Comparison of traditional and proposed approach



SS

Fig. 2. Difficulty of encryption: 2

cryptographic mining. While for cryptocurrencies, this number is very large, in the case of domestic Internet of Things, this number is relatively smaller. This is due to the disparity in use case, security requirements and time constraints.

The simulations work on “difficulties” ranging from 2 to 5. In other words the length of the initial cryptographic challenge is varied from 2 to 5 in order to give an idea of how this factor affects the time considerations of the implementation. For various difficulty levels, graphs are presented portraying the changes in timing between traditional blockchain approach and the presented approach enhanced with flow policies. It can be seen in Fig. 2, Fig. 3, Fig. 4 and Fig. 5 that as the difficulty in the blockchain method is increased the associated time of transactions also increase. It is evident that the proposed approach provides a huge temporal benefit in optimizing the operation of the Internet of Things Infrastructure of domestic use case scenario. For difficulty levels of 2, 3, 4 and 5, the performance of all the iterations are clearly better for the proposed approach, as compared to the traditional blockchain approach.

This work also encompasses the evaluation of the proposed method in scenarios with different percentages of invalid transactions. Fig. 6 represents the performance of the model in the above-mentioned scenario. It can be observed that as the percentage of invalid transactions increases, the time consumption of the system reduces. This is due to the fact that invalid transactions are rejected in the filtering phase and are thus, not mined at all. This saves the computational time otherwise wasted in their validation. However, even when

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:47:47 UTC from IEEE Xplore. Restrictions apply.

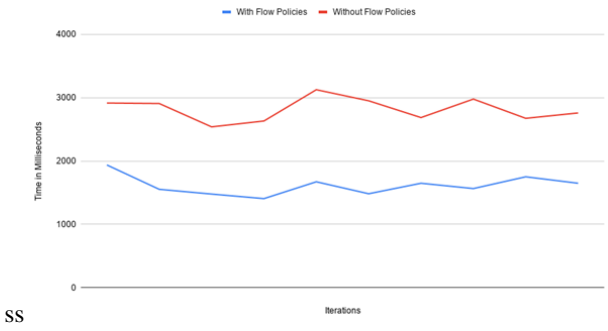


Fig. 3. Difficulty of encryption: 3

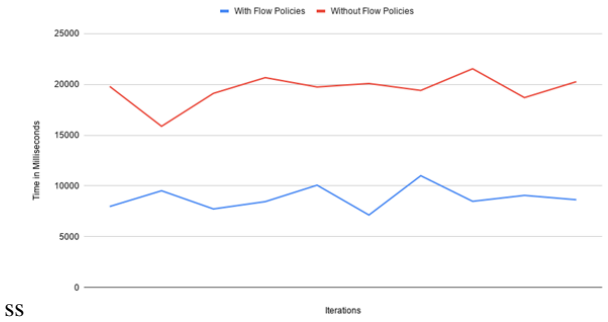


Fig. 4. Difficulty of encryption: 4

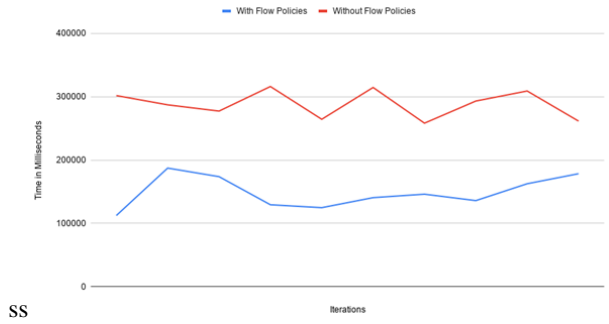


Fig. 5. Difficulty of encryption: 5

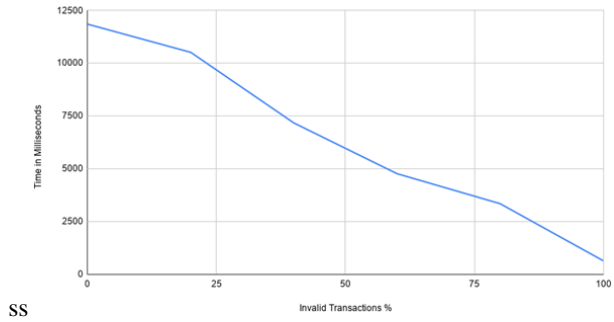


Fig. 6. Variability in Percentage of Invalid Transactions

100% of the transactions are invalid, The time does not reduce to 0. This is because some computation is still required by the system to run processes other than mining. When 0% of the transactions in the system are invalid, the system requires a lot of time as all the transactions are mined as usual and no apparent optimization is achieved.

V. Conclusion

This work navigates through the blockchain mechanism of operation and analyses the scope of integrability to domestic Internet of Things (IoTs). While navigating through the use case scenario of modern Blockchain Technology, the work considers as to how blockchain is suitable for domestic Internet of Things and the benefits that the integration of blockchain provides by enhancing the security and reliability of blockchain by imparting decentralization, immutability and transparency.

One of the major barriers in scaling blockchain into domestic IoTs is the immense latency which comes along with the blockchain technology. Primarily, the Proof of Work (PoW) consensus algorithm associated with blockchain is responsible for the latency caused as the cryptographic challenge presented before miners are time consuming. However, it is also crucial to impart security to the process of validation of transactions. Intensive and inclusive literature review and research led us to an optimizing technique, which when integrated to traditional blockchain led to a significant decline in time considerations. This technique requires installation of entity specific traffic flow policies which are defined for each entity in the IoT infrastructure. This optimization is very specific to the use case of domestic Internet of Things in which data flow paths are relatively static. In other scenarios, such as cryptocurrency, this is not viable owing to the variable data flow paths and the unpredictable nature of transactional packet flow.

This work then conducts an analytical simulation in order to assess the performance metrics by varying the user specific parameters. A few parameters such as percentage of malicious transactions, “difficulty” of the mining and overall number of transactions are varied and the corresponding data sets are used to obtain performance graphs. These graphs indicate the performance improvement achieved and discusses the implication of this improvement in real world scenarios.

References

[1] Mengelkamp, Esther, et al. ”A blockchain-based smart grid: towards sustainable local energy markets.” Computer Science-Research and Development 33.1-2 (2018): 207-214.
[2] Mettler, Matthias. ”Blockchain technology in healthcare: The revolution starts here.” 2016 IEEE 18th international conference on e-health networking, applications and services (Healthcom). IEEE, 2016.
[3] Eyal, Ittay. ”Blockchain technology: Transforming libertarian cryptocurrency dreams to finance and banking realities.” Computer 50.9 (2017): 38-49.
[4] Samaniego, Mayra, and Ralph Deters. ”Blockchain as a Service for IoT.” 2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). IEEE, 2016.
[5] Fernandes, Earlence, et al. ”Flowfence: Practical data protection for emerging iot application frameworks.” 25th USENIX Security Symposium (USENIX Security 16). 2016.
[6] Lin, Huichen, and Neil W. Bergmann. ”IoT privacy and security challenges for smart home environments.” Information 7.3 (2016): 44.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:47:47 UTC from IEEE Xplore. Restrictions apply.

[7] Puthal, Deepak, et al. "The blockchain as a decentralized security framework [future directions]." IEEE Consumer Electronics Magazine 7.2 (2018): 18-21.
[8] Jeppsson, André, and Oskar Olsson. "Blockchains as a solution for traceability and transparency." (2017).
[9] Khan, Minhaj Ahmad, and Khaled Salah. "IoT security: Review, blockchain solutions, and open challenges." Future Generation Computer Systems 82 (2018): 395-411.
[10] Andrea, Ioannis, Chrysostomos Chrysostomou, and George Hadjichristofi. "Internet of Things: Security vulnerabilities and challenges." 2015 IEEE Symposium on Computers and Communication (ISCC). IEEE, 2015.
[11] Deogirikar, Jyoti, and Amarsinh Vidhate. "Security attacks in IoT: A survey." 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC). IEEE, 2017.
[12] Kolias, Constantinos, et al. "DDoS in the IoT: Mirai and other botnets." Computer 50.7 (2017): 80-84.
[13] Zheng, Zibin, et al. "Blockchain challenges and opportunities: A survey." International Journal of Web and Grid Services 14.4 (2018): 352-375.
[14] Bazari, Aditya Shyam, et al. "Node Criticality Assessment in a Blockchain Network." Proceedings of the 2nd Workshop on Blockchain-enabled Networked Sensor. 2019.
[15] Antonakakis, Manos, et al. "Understanding the mirai botnet." 26th USENIX Security Symposium (USENIX Security 17). 2017.
[16] Dorri, Ali, Salil S. Kanhere, and Raja Jurdak. "Towards an optimized blockchain for IoT." 2017 IEEE/ACM Second International Conference on Internet-of-Things Design and Implementation (IoTDI). IEEE, 2017.
[17] Fernandes, Earlence, et al. "Flowfence: Practical data protection for emerging iot application frameworks." 25th USENIX Security Symposium (USENIX Security 16). 2016.
[18] Bull, Peter, et al. "Flow based security for IoT devices using an SDN gateway." 2016 IEEE 4th International Conference on Future Internet of Things and Cloud (FiCloud). IEEE, 2016.
[19] Samaniego, Mayra, and Ralph Deters. "Blockchain as a Service for IoT." 2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). IEEE, 2016.
[20] Suma V. SECURITY AND PRIVACY MECHANISM USING BLOCKCHAIN. Journal of Ubiquitous Computing and Communication Technologies (UCCT). 2019;1(01):45-54.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:47:47 UTC from IEEE Xplore. Restrictions apply.

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 2, ISSN No. 2455-2143, Pages 217-221
Published Online June 2020 in IJEAST (http://www.ijeast.com)



FLEXIBLE SOLAR POWER TREE - A CONCEPT

Mukesh Kumar
Physics Department SSNC
University of Delhi, INDIA

J P Kesari
Mechanical department
Delhi Technological University

Lokesh Rana
Physics & Astrophysics Department
University of Delhi, INDIA

Abstract— **Sunlight is always a great asset for human beings. Among various systems available to access solar power, the most well known innovation is the photovoltaic(PV) transformation of daylight into electricity. Regardless of its benefits, sunlight based PV innovation has issues such as bulky and less efficiency. The most precious process of photosynthesis, the sun catching nature of trees is an inventive thought of light catching cells. The idea of a sun oriented flexible tree is fit for tending to these issues adequately with polish. In this paper efforts have been made to audit the parts of the sun based tree and its energy harvesting capability. The different energy enhancement structures are additionally talked about alongside utilizations of the sun powered tree. The paper additionally addresses the difficulties engaged with this innovation and proposes future research heading.**

Keywords— **Photovoltaic, Solar tree, Solar Energy, Renewable Energy**

I. INTRODUCTION

The journey for green and supportable energy sources has gotten probably the greatest test for our time, because of exhaustion of fossil energy products, environmental challenges, an Earth-wide temperature boost and perpetually developing energy demands. Energy from the sun comprises light and warmth from the Sun, it is being utilized via different advancing technologies. Few of them are like sunlight based heat collectors, sun oriented photovoltaic, sun based heat power, sun based electricity powerhouse, sun-powered building design and counterfeit photosynthesis [1].

The most well-known use of the sun is electric energy generation through the photovoltaic (PV) frameworks. Sun based PV execution is reliant on the nearby climatic conditions and accessibility of sun-powered radiation. Sun based radiation appraisal and estimation is particularly useful for the appropriate structure of sun based energy change frameworks. The trap frequency and intensity of the Sun beams vacillate for the duration of the day and throughout the year. In this manner, sun-powered modules fixed at a specific point may not be completely advanced. Sun following PV frameworks can be structured, however they essentially raise the all-out expense of energy generation as they are exorbitant and require support. Another, encouraging approach to coordinate sun oriented PV into the developed condition is through multidimensional Photovoltaic frameworks [6-8]. One of the essential installation components for the establishment of all such sun based PV modules is land Rooftop sun oriented PV is an alluring choice, however, it has constrained space for modules to be masterminded in a cluster or multistoried building particularly inland limited urban territories. The Sun following PV frameworks can be a possible solution but they fundamentally raise the all out expense of energy as they are exorbitant and require support. Another challenge in this encouraging solar powered PV execution deals with issues like installation space requirement, low PV efficiency, approach to coordinate sun based PV frameworks etc.

Significant cost reductions have been observed in the past few decades along with high efficiency in photovoltaic based solar cells with strong increase in people acceptability. The PV technology now positioned to provide low-cost energy options and have enormous capacity for future energy generation. Improvement in energy-conversion efficiency reduces such costs, with mass manufacturing likely to bring down the additional costs associated for achieving higher efficiencies. This suggests that the industries will evolve beyond the conventional single-junction solar cells that presently dominate commercial production, where energy-conversion efficiencies are constrained by practical values of Shockley–Queisser limits to below 30%. Contrary to that nature's solar capturing technology, plants exploit the Sun more effectively [3]. The low effectiveness is due to associated losses like spectrum loss, reflection loss and heat generation and so forth.

In this article the overall prospects for a range of approaches that can potentially exceed these limits are based on ultimate efficiency prospects, material requirements and installations outlook. The challenges involved in this technique are considered and suggestions are proposed to overcome them. Finally the concept of solar trees is discussed to overcome above losses through law of conservation and minimization process.

II. SIMILARITIES BETWEEN SOLAR CELL AND PHOTOSYNTHESIS

Like plants, solar cells gather their energy from sunlight. Plant leaves collect sunlight and transform this energy into



chemical energy. Similarly, photovoltaic solar cells convert the sun's energy into electricity. They perform the same function but in different senses. The main resemblance between the two is that both photovoltaic and photosynthetic plants are supposed to absorb photons.

This task of absorption of sunlight is carried out by a molecule called chlorophyll in nature (plants). In the case of photovoltaic, scientists have designed materials where the electrons can absorb the energy of the sun's photons. In both cases, photons are absorbed by electrons, which take up the extra energy and get excited to a higher energy state known as excited electron. The process of conversion of the energy absorbed by excited electrons should be fast before they lose their energy and go back to their initial state. In the case of photosynthesis, this issue is resolved by moving the electron from one molecule to another until it finally settles in a molecule that can store this energy for a long time. In photovoltaic, the excited electrons are carried off into a circuit, where they are either used to operate directly or are routed into a battery for storage [4].

There is a type of solar cell that is designed to be as analogous to photosynthesis. This is a non-conventional photovoltaic cell that tries to reproduce the way photosynthesis works. Instead of moving an electron as quickly as possible through a crystal of identical atoms, this dye-sensitized solar cell absorbs energy in a dye molecule, then transfers the excited electron into another material located adjacent to the dye molecule. This technique makes sure that the electron doesn’t lose its energy ineffectively. When connected to a circuit, this electron makes its way through the second material without risking losing its energy.

III. PLANT ARCHITECTURE INSPIRE SOLAR CELL DESIGN

Plants exhibit a variety of systematic structures and structures, which are difficult to explain. Plants are dependent on the light energy they receive from photosynthesis, and strong sunlight is the most important factor in photosynthetic influence, but in addition the leaf temperature, leaf water status, and many other physiological procedures of the plant also influence photosynthesis. There are no fixed configuration rules in all conditions.

Geometrically it is determined by only two parameters: the propagation limit and the branch length. They are undoubtedly more cohesive and consistently attracted to authors of plant conditions and development models. In trees, the leaves outside the hiding place are different from those inside the tree cover. The outer leaves are called "sun leaves," and the inner leaves are "secret leaves." These leaves contain differences such as fiddle, internal structures of life, and science that enable the translation of certain energy to use different types of solar radiation more effectively. The leaves of the sun are generally normal, prolonged and larger than the secreted leaves, which contain many layers of chlorine-containing tissue and are very broad within the arteries. It is believed that the sun's leaves are better adapted to capture and use direct sunlight (if not too different to even think of causing heat and pressure related damage). Its long position is connected in parallel with the most important levels of sunlight reaching the inner cover where the foliage is located. Shaded leaves appear to make good use of sunlight, reaching into the inner shelter after exposure to different substances, such as outer leaves, in the form of direct light. Leaves of similar color can be found in an outdoor shelter such as the color towards the subtle sun. The leaf level of photosynthesis depends on the biomass of the leaf biomass as the emission of light into the leaf area. The scattering and course of leaf functioning within a tree can completely alter the effectiveness of a good harvest. The leaf area all tightened to a given plant is a very important aid that contributes to the partitioning of edible radiation. When the point from the light of the sun decreases from the equator to the highest positions, the trees of different lengths and widths have effectively changed the light capture technique. All kinds in a variety of colors to add an effort to capture a quick glow through the whole tree.

The leaf must "absorb" sunlight, water from the inside and carbon dioxide from the air that combines photosynthesis. Leaf photosynthesis, meanwhile, is influenced by these carefully regulated materials: enough sunlight and carbon dioxide for photosynthesis, yet not all that closely related heat storage or water misery. The leaves found on the Nature Trail trees are expensive and adequate or molded with a needle that is ideal for their water-conserving weather conditions and high maintenance during the day is an important way to persevere.

Plants show a striking assorted variety of structures and structures, which are hard to decipher. Plants rely upon the light energy that they catch by photosynthesis, and sun-powered radiation is the significant main impetus influencing photosynthetic movement, yet in addition leaf temperature, leaf water status, and numerous others. because nobody configuration rules in all conditions. The geometrically dictated by only two parameters: spreading edge and branch length. are undoubtedly multi-fractals and have consistently pulled in mathematicians for plant shape and development models. In trees, the leaves on the outside of the tree shelter contrast from those inside the tree covering. The outside leaves are alluded to as "sun leaves," while the inside leaves are "conceal leaves." These leaves have contrasts fit as a fiddle, interior life structures, and science that make an interpretation of specific capacities to utilize various types of sun based radiation successfully. Sun leaves are normally littler, progressively prolong, and thicker than concealed leaves, with more layers of chlorophyll-containing tissues and increasingly broad inside vascular frameworks. It is believed that sun leaves are better adjusted to catch and utilize direct

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 2, ISSN No. 2455-2143, Pages 217-221
Published Online June 2020 in IJEAST (http://www.ijeast.com)



sunlight based radiation (when it isn't too exceptional to even think about causing heat and different pressure related harm). Their lengthened shape is likewise connected with more significant levels of sunlight based radiation arriving at the inward covering where the shade leaves are found. Shade leaves appear to successfully utilize diffuse sun oriented radiation, which arrives at the internal shelter in the wake of having been dissipated by different items, similar to the external sun leaves, in the way of direct light. Shade leaves can likewise be found on the outside shelter as an afterthought that faces from the overarching sun. The leaf photosynthesis rate relies upon the biochemical limits of the leaf just as a light appropriation on the leaf surface. The dispersion and course of action of leaves inside a tree can firmly change the light reaping effectiveness. The all-out leaf zone bolstered by a given plant is the most essential auxiliary property that influences the division of consumed radiation. As the sunlight based tendency point diminishes from the equator to higher scopes, trees with contrasting stature to width proportion have inalienably changing efficiencies of light capture. Every single such variety to augment the immediate light capture attempt by the whole tree.

A leaf must "catch" sunlight, water from inside and carbon dioxide from the encompassing air by means of pores for photosynthesis. The leaf photosynthesis, at that point, is influenced by these exercises in careful control: enough daylight and carbon dioxide to run photosynthesis, yet not all that much-related warmth retention or water misfortune. The leaves found on the trees of the Nature Trail are either expensive and level or needle-moulded which appropriate to its climatic conditions for cautious preservation of water and most extreme daylight retention is a significant endurance system. A significant result of variety in a leaf design, its shape and size is the general light gathering and the proficiency factor for photosynthesis.

IV. SOLAR POWER TREE

Nature motivation centres around profoundly requested little bent surface geometry which follows the most extreme snare of daylight and water protection through photosynthesis. Likewise between a sun oriented cell and a tree is that both attempt to catch the most extreme sunlight based energy with least water misfortune or warmth age. At present utilization of PV, frameworks increment quickly because of ceaseless decrease in costs of sun based cells. Still, there are a few obstacles for the appropriation of this innovation like low productivity and enormous space necessities. Different methodologies [6] have been done to improve sun oriented energy factors, for example, low ingestion coefficient, heat loss, reflection coefficient and so forth, The charge bearer age and conduction impact the light conversion proficiency of a sunlight based cell. This is a roundabout way that relies upon the quantity of sun-powered cells introduced and their energy change limit.

Recently inspired by natural heat-trapping technology - photosynthesis, a new concept solar tree is introduced by many researchers [3,4]. Nature centres around profoundly requested little bent surface geometry like tree geometry. So a gathering of cells which are littler might be prolonged, a little bent and have a multilayer structure. The new idea- Solar Tree is a natural venture, an environmental model, a counterfeit sun based structure that seems as though sculptural trees exist from little scope (size of a bonsai tree) to enormous scope (about the size of a breeze turbine) power plant. In sun based trees, PV boards or cells are orchestrated in a Fibonacci arrangement design rather than leaves. It is an independent unit that produces green energy and provides a place of comfort and energy for a wide variety of services.

The structure is a ground-mounted solar system with a pole that supports many individual panels up in the air. The performance of sunlight based trees vary and they have been intended to give the capacity to face various situations. Sunlight based tree has remarkable properties as far as structure and multi-edge direction parameters. By utilizing the tallness parameter, the sun-powered tree requires less space utilization which can lessen the installation cost. Such a mix of the sun oriented cell should be possible on common trees to lessen the expense of mounting. Contrasted with the customary sun based PV homestead which is situated a solitary way, the multi-edge direction parameter from the sun based tree boards yields the potential ability to retain higher daylight force prompting higher yield vitality. Simultaneously, this parameter disposes of the need to introduce a sun oriented tracker which can decrease the operational expense. To build the assortment Generation of 1MW force from PV module framework requires the space of around 5-6 Acres for lodging the boards as it whereas a solar tree would take just 1% of the land territory in contrast with general PV lodging. So it is appropriate for places where there is a shortage of land. It can likewise be bolted at any situation to withstand the wind pressure because of substantial tempest influencing over the fundamental shaft/trunk. The boards will be normally looking with least block from different towards the sun at an edge as required so they can gather the greatest sun-powered energy in a daytime.

Sun-powered trees are actually a functional answer for household use. In this setting utilization of PV innovation for the local necessity as the sun oriented tree is the acceptable option when contrasted with traditional level or housetop mounting PV panels. Solar trees can be introduced close, before the house or on a porch, where there is no concealing consistently. To build assortment efficiency, the nanowire cell can be used rather than present PV cells. The Nanowire cells can gather multiple times progressively sun based vitality

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 2, ISSN No. 2455-2143, Pages 217-221
Published Online June 2020 in IJEAST (http://www.ijeast.com)

when contrasted with regular cells. The idea of nano leaf is a plan to expand the different types of energies, for example, sunlight based, warm and piezoelectric force by stalks, branches, and twigs of a tree.

Solar trees are really a practical solution for domestic use . In this context use of Photovoltaic technology for the domestic applications in the form of the solar tree is the good alternative when compared with conventional rooftop mountings. Solar trees can be installed in front of the houses, at the terrace, in streets or where there is no shading throughout the day.

V. Minimization of Solar Energy Loss

Sunlight increases temperature which affects photovoltaic conversion. This does not significantly affect the light generated current of the cell but reverse saturation current increases exponentially with temperature and also reduces open circuit voltage with temperature. Furthermore, it has been experimentally observed that temperature drops the efficiency of solar cell. It is imperative to reduce the heat build-up in the cell in order to improve conversion efficiencies. Recently, it is seen that solar tree energy conversion efficiency increases through minimizing the chlorophyll antennae [6]. Similarly to keep the heating impact of a solar cell low , various approaches are possible. The approaches are possible such as use of multispectral cells for low heat production, reduction in energy loss through current loss and the use of thermal generators for energy harvesting. The reduction in size expected a reduction in current with less utilization time. Recent quantum level technology developed tuneable band gaps across a wide range of energy levels by changing their size, whereas the bandgap is fixed by the choice of bulk materials. This reduces semiconducting particles below the size of the exciton Bohr radius. The quantum nature of existing particles set electron energies finite alike energies in an atom. The developed Quantum dots methodology is referred to as artificial atoms and energy levels are tuneable by changing their size, which in turn defines the bandgap [1]. The dots can be grown over a range of sizes, allowing them to express a variety of band gaps without changing the construction techniques. This makes it attractive for multi-junction solar cells and use of a variety of materials to improve efficiency by harvesting multiple parts of the solar spectrum. To increase the collecting efficiency of cell, the nanowire cells can be used instead of present Photovoltaic cells. The Nanowire cell can collect 15 times more solar energy as compared to conventional cells [1].

VI. Flexible Solar Tree

From the previous section, it is found that the natural light conversion technology photosynthesis is similar to solar cell technologies. Rather the natural conversion technology is found higher than solar cell light conversion technology [3,6]. These natural processes suggest various different possible solutions to these energy losses in a solar cell. The laws of science put some limits on such power conversion, but also leave some scope for improvement through multiple loss minimizations. Section II shows that sun spectrum absorption varies with nature and configuration of leaves, so give an idea of multi-spectrum material for solar cell construction. Also leaves shape, size effect light inclination and efficiency for light energy conversion. This inspiration focuses on small curved surface geometry to capture more sunlight. The multilevel solar cell - tandem cell can be fabricated to drastically reduce the transmission losses. The efficiency of natural photosynthesis also depends on water evaporations; similarly, heat dissipation play is a key factor in solar cell light conversion. .From the previous section, it is also clear that scaling down also affects efficiency drastically.

All these ideas include a modified version of the solar tree which is discussed in section II. The advancement of technology has reduced the size of solar panels to a size of 0.6 x 2.55 inches . This is comparable to the size of plant leaves and their individual identity power varies from 0.5 to 4 W. The cumulative power output will be quite high for various

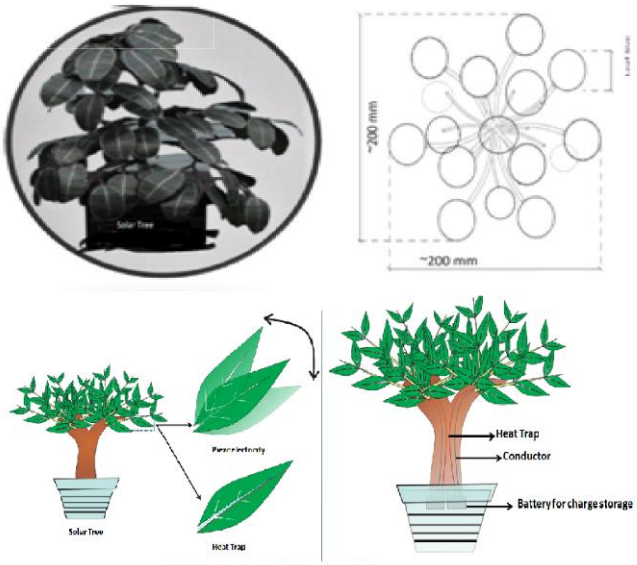


Fig 1 : Solar Tree Model

domestic applications. The concept of nano solar material solar tree leaf that would maximize the light conversion [1,5]. In order to trap maximum solar energy, the use of quantum wire PV boards is orchestrated in a Fibonacci arrangement design. One such plant pot size is proposed is shown in figure1. In this novel solar tree design, multispectral cell junctions can be fabricated to drastically reduce the transmission losses. On top of the leaves, the anti-reflective coating can be used to reduce reflection losses. Leaves are the independent unit that produces green energy and provides a place of comfort and energy for a wide variety of services. This kind of texturing of the rear surface will not be required



to achieve total internal reflection as the surface of flexible materials is already rough.

The leaf photosynthesis is also influenced by these exercises in careful control of carbon dioxide or too much-related water retention. In photosynthesis, this problem of heat generation is solved by molecules energy transfer to another until it settles in a molecule that can store energy for a long time. The conductive branches of photovoltaic solar trees can trap excited electrons into a battery for storage. This leaf-like structure acts as a thermionic generator thereby leading to the enhanced sensitivity. This design ensures the enhanced performance of a solar cell as a reduction in heat-up leads to higher efficiency. In addition to this the flexible nature of leaves also helps in harvesting wind energy. The various forms of energies such as solar, thermal and piezoelectric power by stalks, branches, and twigs of a tree. The amount of energy generation varies with the size/shape of leaves, the number of leaves etc. The flexible nature of coated piezoelectric solar cells helps in harnessing the wind energy throughout day and night. A significant amount of wind power is also generated due to wind-generated movement of small light weighted leaves. The heat energy can be converted simultaneously into electric energy using thermal converters. the energy of radiations. The various designs of flexible trees can be possible depending on geographical conditions and demands. The synchronization of solar energy with heat energy and wind energy supply continuous supply of energy even at night or when the sun is covered with clouds.

VII. CONCLUSIONS

The The sunlight based tree idea is effective to satisfy the expanding energy demands of the individuals, sparing area, and ought to be executed to give power without the issue of intensity cut-off. The concept of nano leaf is an idea to maximize the various forms of energies such as solar, thermal and piezoelectric power by stalks, branches, and twigs of a tree.

All this energy waste harvesting converts it into grid power. The adaptable nature and nano-size leaf fit it into a beautiful small size pot plant and a valuable device. This superb plant that requires no watering and sustains. The beginning venture cost of the sun based tree is likewise equivalent to the same limit PV frameworks as other framework parts are similar. The restitution time of proposed framework is by all accounts high yet because of persistent increment in the expense of network force and decrease in the expense of PV cell because of mechanical progression long recompense period can be compensated. The same photosynthesis move configuration can be reached out to different areas and applications including higher energy utilizations. At the last, Solar Tree front line eco-perfect innovation and should work in agreement with common trees.

VIII. REFERENCE

[1] Baskoutas, Sotirios; Terzis, Andreas F. (2006). "Size-dependent band gap of colloidal quantum dots". Journal of Applied Physics. 99 (1): 013708–013708–4.Bibcode:2006JAP....99a3708B. doi:10.1063/1.2158502

[2] Berny S. and others (2015), Solar Trees: First Large-Scale Demonstration of Fully Solution Coated, Semitransparent, Flexible Organic Photovoltaic Modules, , Advance Science, December 2015, https://doi.org/10.1002/advs.201500342

[3] C. Bhuvaneswari, R Rajeswari, C.Kalaiarasan and K.M.S. Muthukumararajaguru (2013), Idea to Design a Solar Tree Using Nanowire Solar Cell International Journal of Scientific and Research Publications, Volume 3, Issue 12, December 2013 1 ISSN 2250-3153.

[4] Economictimes (2019), Mini Solar Panel: Ideal power source for small devices/portable appliances, Read more at: Economictimes.indiatimes.com

[5] Kumar M. and other (2016), Traffic driven wind energy generator, Technology Letters, Vol.3 No.1, p10-13 (ISSN: 2348-8131).

[6] Melis A. (2009), Plant and Microbial Biology, University of California, Berkeley, CA 94720-3102, USA, Plant Science 177, p 272–280. Gupta S., and others (2015), "The benefits and applications of Solar Tree With Natural Beauty Of Trees", SSRG International Journal of Electrical and Electronics Engineering, Issue April 2015 Vol-4 Issue-2 2018 IJARIIE-ISSN(O)-2395-4396 7653.

[7] Srinivas (2016), A P R, "Design and Development of a SOLAR TREE", International Journal of Scientific & Engineering Research, Volume 7, Issue 10, October-2016 ISSN 2229-5518

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/342881094

Forecasting the Probability of Cost Overrun Risk of Indian Construction Projects using Fuzzy Model

Article · January 2020

CITATIONS
0

READS
30

3 authors, including:



Pradeep K. Goyal
Delhi Technological University
8 PUBLICATIONS **13** CITATIONS
SEE PROFILE



Prof.(Dr.) R.C. Chhipa
Suresh Gyan Vihar Univeristy, Jaipur
64 PUBLICATIONS **123** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

COST OVERRUN ANALYSIS MODEL IN FUZZY ENVIRONMENT FOR CONSTRUCTION PROJECTS View project

Physicochemical and spectroscopic analysis of effluents of textile and paper industries View project

All content following this page was uploaded by Prof.(Dr.) R.C. Chhipa on 12 July 2020.

The user has requested enhancement of the downloaded file.

International Journal on Emerging Technologies **11**(1): 10-22(2020)

ISSN No. (Print): 0975-8364
ISSN No. (Online): 2249-3255

Forecasting the Probability of Cost Overrun Risk of Indian Construction Projects using Fuzzy Model

Savita Sharma[1], Pradeep K. Goyal[2] and R.C. Chhipa[3]
[1]Research Scholar, Department of Civil Engineering, Suresh Gyan Vihar University, Jaipur (Rajasthan), India.
[2]Associate Professor, Department of Civil Engineering, Delhi Technological University, Delhi, India.
[3]Professor, Suresh Gyan Vihar University, Jaipur (Rajasthan), India.

(Corresponding author: Savita Sharma)
(Received 04 October 2019, Revised 11 November 2019, Accepted 10 December 2019)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: A fuzzy model for forecasting the probability of cost overrun risk for Indian construction project is developed by identifying, classifying and ranking the cost overrun factors of the construction industries. Fifty five cost overrun factors are identified, classified into eleven clusters through a detailed literature review process and discussion with experts from the Indian construction industry. A questionnaire survey in Indian construction industry was conducted for data collection to calculate Relative Importance Index (RII) of the factors and clusters. On the basis of the relative importance index of the factors and clusters, a model is developed using the fuzzy logic tool box of MATLAB program software. The model is developed in two stages. First the probability of cost overrun due to the clusters of cost overrun factors is determined and the final cost overrun risk probability of the project is determined taking into consideration the probability of all the clusters. Applicability of the fuzzy model has been tested on a real case study in India. With the help of the proposed model, it is possible to guide project managers and decision makers to make more informative decisions such as contingency estimation, mark-up estimation, bid price, selection of optimum procurement route, evaluation of different projects and preparation of project portfolios.

Keywords: Construction industry, Cost overrun, Probability, MATLAB, Fuzzy logic, India.

I. INTRODUCTION

To complete a construction project within stipulated budget is an important factor among all the norms required for the success of the project [1-3], yet construction projects are found to be suffering from the issue of cost overrun. Cost overrun is defined as the increased costs over estimated cost [4]. This issue of cost overruns is critical both for developed as well as developing countries [5-13]. Indian construction projects are also suffering from the problem of cost overrun. A joint study was conducted by PMI and KPMG with the support of MOSPI in India on infrastructure projects in which a total cost overrun of Rs 2.19 lakh crore for 1304 projects was reported in 2018 [14]. Generally the cost overrun risk analysis is ignored in the projects and the cost of a project is determined just by adding some contingency charges, consequently many projects fail to complete the project in stipulated budget. Therefore it is essential to estimate the probability of cost overrun risk for Indian construction projects to determine a reliable cost of the project.

Several theories and methods have been proposed for risk analysis for construction projects. Zhi, (1995) [15] used qualitative method by using P-I matrix of for risk analysis of overseas construction project, Baccarini and Archer (2001) [16] Computed a risk score for cost, time or quality of project by employing P-I model. Hull (1990) [17] proposed various models by using Monte Carlo Simulation (MSC) and PERT theory for the risk analysis of cost, schedule and technical performance. Özta & Ökmen (2004) [18] presented Monte Carlo simulation technique for determining the risks and uncertainties associated with project cost and duration. Thomas *et al.*, (2006) [19] used a Fault tree method to model risky scenarios by utilizing linguistic variables to evaluate risk likelihood and impact. Dey (2001) [20] presented risk analysis model based on AHP and decision trees. Dikmen and Birgonul (2006) [21] used AHP for risk and opportunity estimation of international construction projects. Nasir *et al.*, (2003) [22] used belief network and MCS for estimating activity duration. Kangari and Riggs (1989) [23] proposed a model for risk analysis by using FST. Tah and Carr (2000) [24] presented a model for hierarchical risk breakdown structure for qualitative risk assessment through the concepts of fuzzy association, fuzzy composition and application of fuzzy logic.

According to a study by Sharma and Goyal [25], it was concluded that each theory has its own advantages, disadvantages and no theory seems to be complete for handling the uncertainty. However, based on the work done earlier it may be suggested that fuzzy theory is more suitable for handling the complex problems in construction industry as the process is based upon experience, assumptions and human judgment. The fuzzy theory has been applied in the different areas of construction projects as decision support tool, enhancing the performance, evaluation and assessment, forecasting and modeling construction risks [26-35, 40].

The previous studies of cost overrun risk analysis in India were very few in which fuzzy techniques were applied. Therefore in this study a fuzzy model for forecasting the probability of cost overrun risk in Indian construction project is developed by identifying, classifying and ranking the cost overrun factors of the construction industries. The construction industry caters to building and other infrastructure development project. Data for the cost overrun factors are collected from medium to large construction industry with turnover varying from 50 crores to 600 crores. Collected data through questionnaire are validated from sample verification technique and expert opinion. Only reliable data are used in the proposed model. With the help of the proposed model, it is possible to guide project managers and decision makers to make more informative decisions such as contingency estimation, mark-up estimation, bid price, selection of optimum procurement route, evaluation of different projects and preparation of project portfolios.

II. METHODOLOGY

The adopted methodology for the development of the model and for obtaining the results of interest is explained diagrammatically in Fig. 1. This model is developed in two stages as shown in Fig.1. First the probability of cost overrun due to each cluster is assessed by the probability level of the factors of respective cluster. The probability of cost overrun of project is assessed by the probability output of each cluster as determined in the previous step. Major components of the methodology are explained in the following sections.

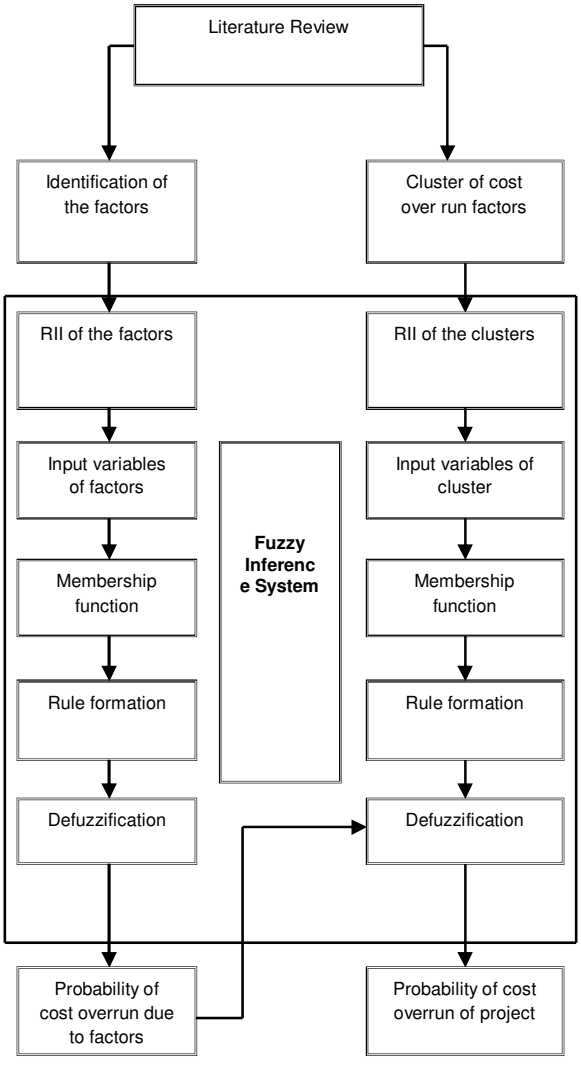


Fig. 1. Methodology.

A. Identification and classification of cost overrun factors

A number of studies conducted on the factors responsible for cost overrun in the construction projects are reviewed [6-14]. The cost escalation causes and their classification are given by Sharma and Goyal (2014) [25] for the construction industry. Fifty five important factors, responsible for cost overrun are identified through intensive literature review. These factors are grouped in eleven clusters named as 'Cluster Owner', 'Cluster Contractor', 'Cluster Consultant', 'Cluster Design', 'Cluster Project', 'Cluster

Material', 'Cluster Labor', 'Cluster Equipment', 'Cluster Contract', 'Cluster Coordination/communication' and 'Cluster External'.

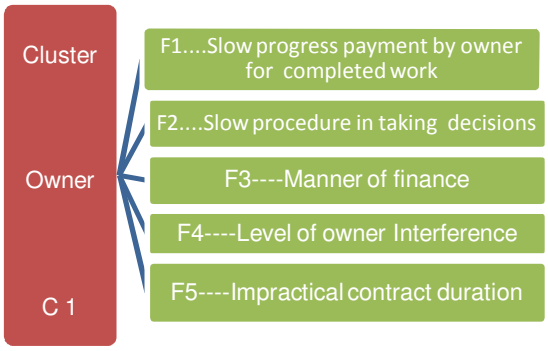


Fig. 2. Cost overrun factors of 'Cluster Owner'.

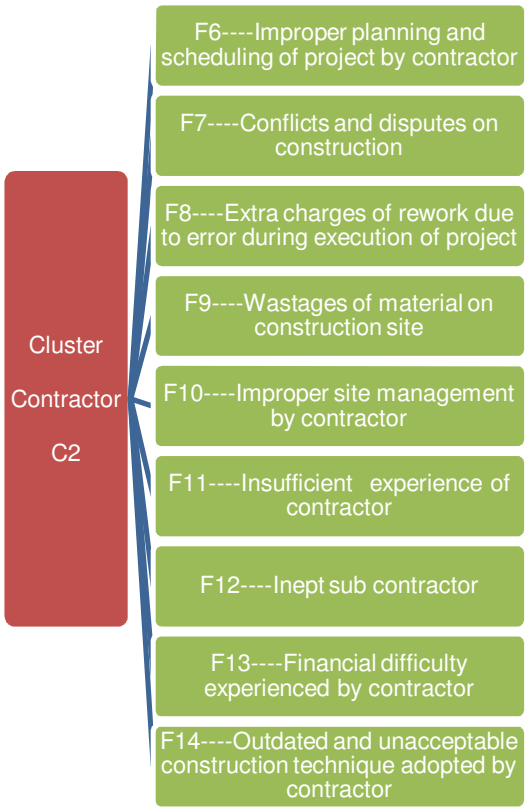


Fig. 3. Cost overrun factors of 'Cluster Contractor'.

All the factors are represented as F1, F2, F3 etc. and clusters are represented as C1, C2, and C3 etc. slow progress payment by owner for completed work, slow procedure in taking decisions, manner of finance, level of owner interference and impracticable contract duration are the factors of 'Cluster Owner' as shown in Fig. 2. improper planning and scheduling of project by contractor, conflicts and disputes on construction sites, extra charges of rework due to error during execution of project, wastages of material on construction site, improper site management by contractor, insufficient experience of contractor, inept sub contractor, financial difficulty experienced by contractor and outdated and unacceptable construction technique adopted by contractor are the factors of 'Cluster Contractor' as shown in Fig. 3.

In correct estimation of time and cost, contract management, delay in inspecting of work, quality assurance & quality control and experience level of consultant are the factors of 'Cluster Consultant' as shown in Fig. 4. In a similar manner all the factors of 'Cluster Design', 'Cluster Project', 'Cluster Material', 'Cluster Labor', 'Cluster Equipment', 'Cluster Contract', 'Cluster Coordination/communication' and 'Cluster External' are shown in Fig. 4-12.

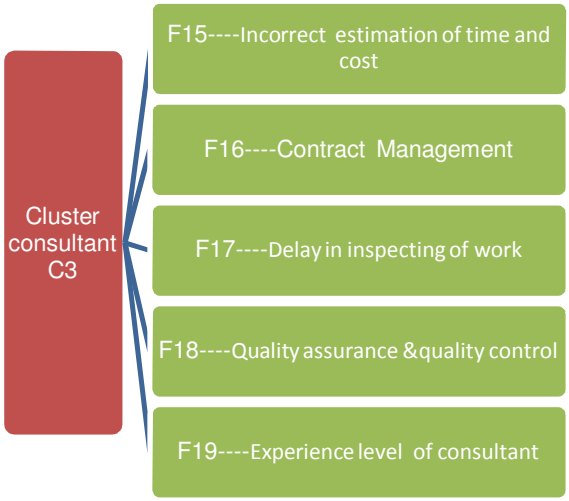


Fig. 4. Cost overrun factors of 'Cluster Consultant'.

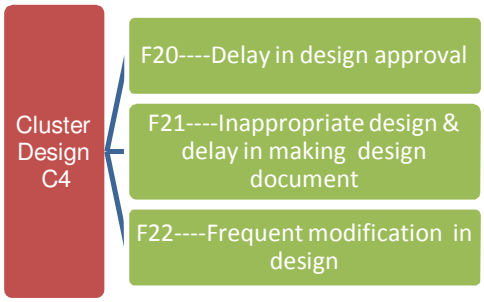


Fig. 5. Cost overrun factors of 'Cluster Design'.

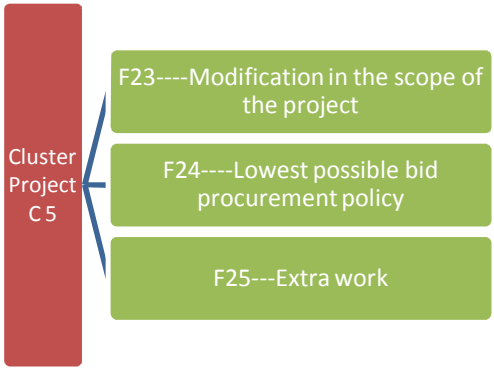


Fig. 6. Cost overrun factors of 'Cluster Project'.

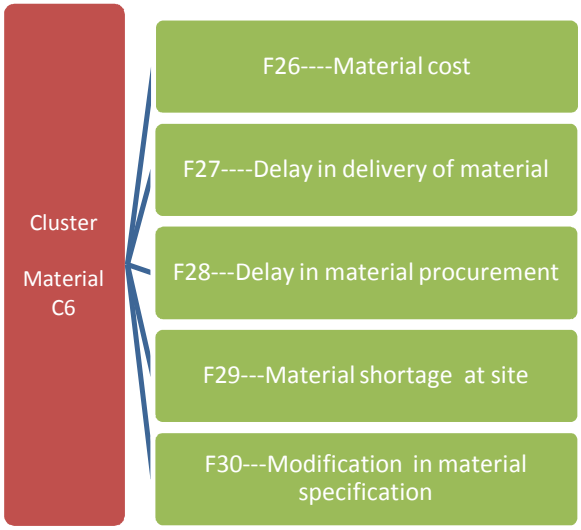


Fig. 7. Cost overrun factors of 'Cluster Material'.

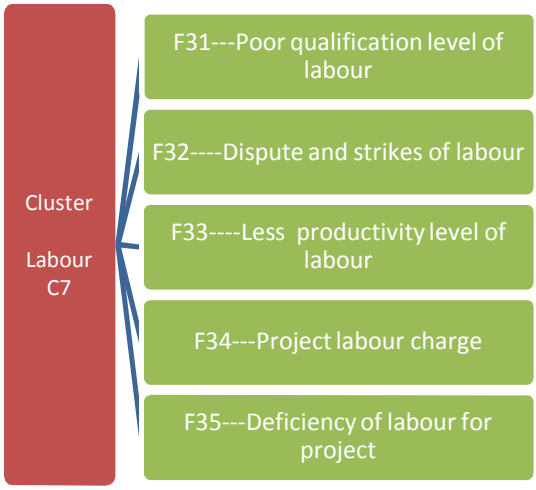


Fig. 8. Cost overrun factors of 'Cluster Labor'.

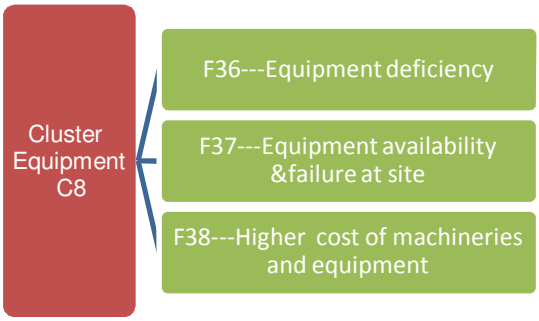


Fig. 9. Cost overrun factors of 'Cluster Equipment'.

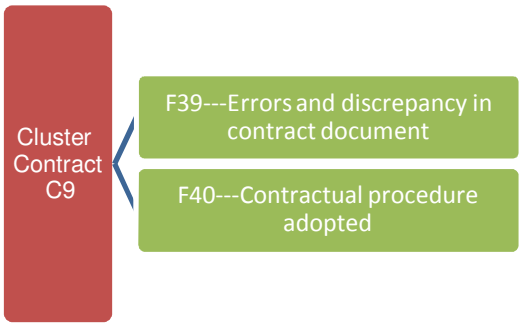


Fig. 10. Cost overrun factors of 'Cluster Contract'.

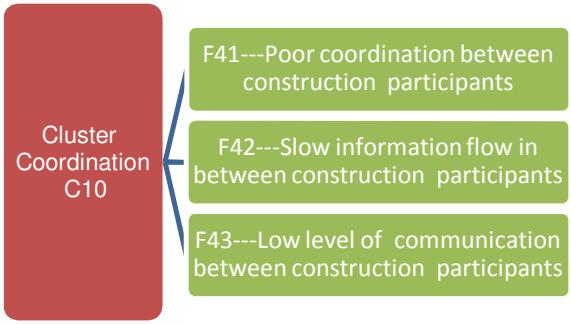


Fig. 11. Cost overrun factors of 'Cluster Coordination'.

B. Relative importance index of factors and clusters

For ranking the cost overrun factors and cluster relative importance index (RII) is calculated by collecting data through a questionnaire survey in Indian construction. The population for survey was drawn from the practitioners of the Indian construction industry, which included owners comprising government sectors (key decision makers), contractors and consultants. A non-probability sampling technique known as convenience and snow ball is used to collect the sample. The questionnaire was distributed through friend and referral network. Personal interviews were also conducted for obtaining the response quickly. Total 250 questionnaires were distributed to the potential respondents within the Indian construction company comprising owner, contractor and consultant. 100 questionnaire set to owner, 100 questionnaire set to contractor and 50 questionnaire set to consultant were distributed. Out of 250 questionnaire set, only 135 responses were returned. Out of total respondents, 62 (45%) of the respondents are working in clients' organization while 47 (34.81%) and 26 (19.2%) respondents are professionals that work in contracting and consulting organizations respectively.

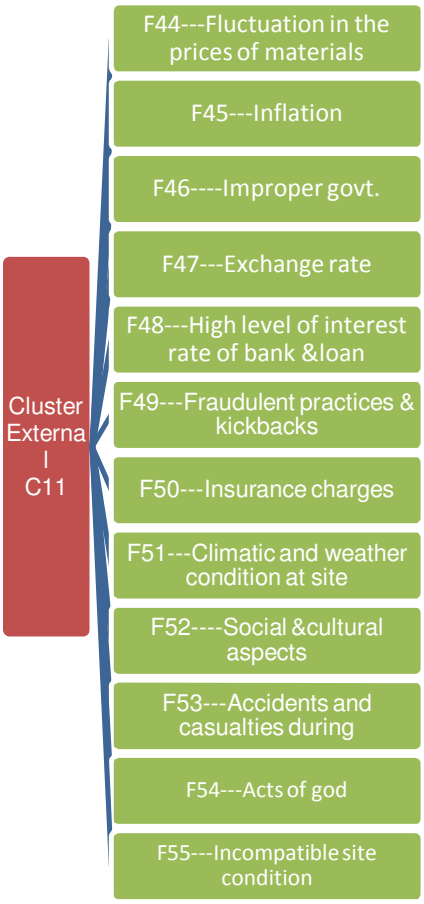


Fig. 12. Cost overrun factors of 'Cluster External'.

The respondents were asked for judging the importance level of identified cost overrun factors. A likert scale of 1 to 5 was used to evaluate the importance level of each factor. These numbers were allocated to the respondents' score as '1 = Very low; 2=Low; 3=Medium; 4=High; 5=Very high'. For ranking these cost overrun factors and clusters the following relation was used to determine the relative importance index (RII)

$$RII = \sum W / (A*N) \quad (1)$$

where W is the weighting for each factor provided by the respondent (it ranges from 1 to 5), A is the maximum weight and N is the number of overall respondents.

C. Model development using fuzzy theory

The fuzzy model to assess the probability of cost overrun is developed in fuzzy inference system (FIS) of MATLAB [36]. Fuzzy inference is a procedure where application of fuzzy logic is employed for mapping from a set of given input to obtain an output. The important features of fuzzy inference system have been explained here in brief.

Fuzzy set. The notion of fuzzy logic was originally pioneered by Zadeh in a seminal article on fuzzy sets theory in 1965 [37]. The concept of the theory is founded on multiple-valued logic; it processes and introduces interpretations which are expressed in vague linguistic terms rather than exact and precise crisp value. The theory is very useful for handling the partial truth issues.

A fuzzy set A of a universe of discourse X can be defined through Eqn. (2):

$$A = \{(x, \mu_A(x)) \mid x \in A, \mu_A(x), \in [0, 1]\} \quad (2)$$

where μ_A (x) is the membership function which provides X a membership values in the interval of 0 to 1.

Fuzzy inference system. As per litreture review Mamdani and Sugeno type of fuzzy inference methods are generally employed. For this research study, Mamdani type fuzzy inference is used because of its wide application [38]. The system of fuzzy inference is basically a rule-based. The crisp input members are fuzzified first to convert them into fuzzy sets by defining the membership functions. After that inference engine of the fuzzy logic system repossesses the knowledge contained in a rule-base where all inputs are combined as per the constructed fuzzy rules. The output fuzzy sets obtained through this process then aggregated and defuzzified where they are converted from a fuzzy set to obtain a crisp number.

The model is developed in two stages as shown in Fig. 2. First the probability of cost overrun due to each cluster is assessed by the probability level of the factors of respective cluster. The probability of cost overrun of project is assessed by the probability output of each cluster as determined in the previous step. The process of fuzzy inference system is applied for both the stages.

The various steps of fuzzy inference process employed for model development are outlined below:

1. The first step of fuzzy inference system is to define all input and output variables for model development. For the first stage of model development the various 'factors' are taken as the input variables and output is the 'probability of cost overrun due to cluster'. For the second stage various clusters are taken as the input variables and 'Probability of cost overrun of project' which is abbreviated as 'PCOP' is defined as the output variable.
2. The input variables are then fuzzified with the help of membership function in this step. The membership function is the fuzziness degree and gives a mathematical value to all input variables. Different shapes of membership functions such as trapezoidal, triangular, bell-shaped, piecewise-linear, Gaussian, g bell-shaped etc. can be used. in this study, g bell-shaped are used as Gaussian and bell membership functions have the advantage of being smooth and nonzero at all points [39]. It can be graphically represented as shown in Fig. 13.

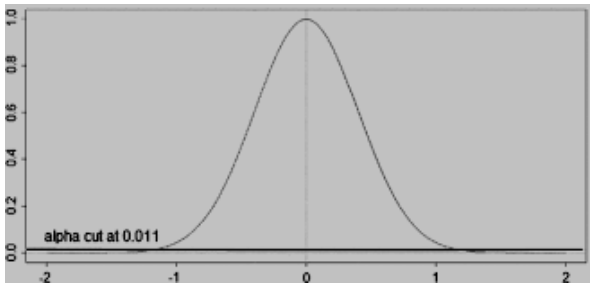


Fig. 13. Gbell-shaped fuzzy membership functions.

3. Fuzzy rules are defined in this step to perform the inference. Fuzzy rules are logical inference and connect

the input variables to output variables. IF-THEN forms of fuzzy rules are used here in this research study. Fuzzy rule for the input linguistic variables x_1, x_2 with their corresponding fuzzy values A_1, A_2 and output linguistic variable y with B as its fuzzy value, can be explained as follows.

If x_1 is A_1 AND x_2 is A_2 then y is B

If part of this rule "x_1 is A_1 AND x_2 is A_2" is called antecedent and then part of the rule "y is B" is known as consequent.

Fuzzy operators "AND", "OR" and "NOT" are also incorporated to construct fuzzy rules.

The estimated values of relative importance for the cost overrun factors are allocated as fuzzy rules weight. An interview was developed with the experts of construction industry in India for the construction of fuzzy rules for forecasting the probability of cost overrun.

4. Defuzzification process takes place finally, where the fuzzy results of the inference engine are converted into a crisp one. Different types of defuzzification methods have been recommended in literature. In this study 'centroid of area' method is applied to forecast 'the probability of cost overrun of the project'. This method of 'centroid of area' calculates the weighted average of a fuzzy set. The outcome of defuzzification to a fuzzy conclusion by using this method can be expressed by the formula

$$Z_{COA} = (\int_Z \mu_A (Z)\, Z\, dz) / (\int_Z \mu_A(Z)\, dz) \quad (3)$$

where Z_{COA} is the crisp output, μ_A (Z) is the aggregated membership function and z is the output variable.

D. Model validation

To test the reliability of the proposed fuzzy cost overrun assessment model, it is tested in a real case study. For this purpose an interview was conducted with a team of the experts of a leading Indian construction company. The panel of the experts included top 10 executive engineers, project managers and site engineers of the company. The experts were requested to examine the cost overrun factors considered for this research and filled in the required information regarding cost overrun factors to assess the proposed model. They were also requested to estimate the cost risk level of the project.

III. NUMERICAL STUDY AND RESULTS

The summary of the calculated, relative importance index (RII) of the cost overrun factors in Indian construction industry and the proposed fuzzy model for these factors has been discussed in the following sections:

A. RII of cost overrun factors and important cost overrun factors in Indian construction industry

The data collected in the questionnaire are analyzed from owner, contractor and consultant's perspective**.** The relative importance index (RII) is computed for each factor causing cost overrun to determine the important factors causing cost overrun in Indian construction industry.

RII of Cluster Owner. RII of the factors slow progress payment by owner for completed work, slow procedure in taking decisions, manner of finance, level of owner interference and impracticable contract duration are 0.498, 0.522, 0.425, 0.357 and 0.676 respectively as shown in Fig. 14.

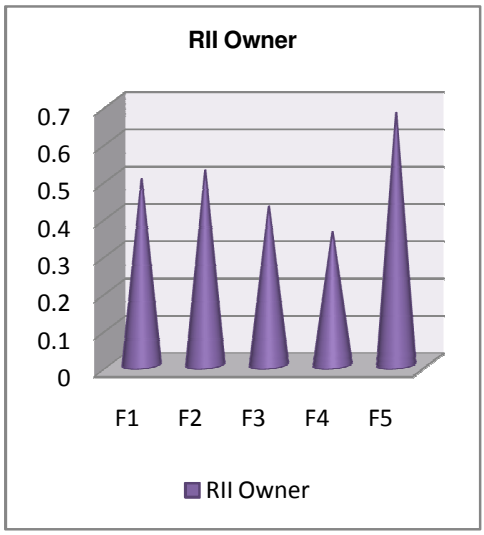


Fig. 14. RII of cluster owner.

RII of Cluster Contractor. RII of the factors improper planning and scheduling of project by contractor, conflicts and disputes on construction sites, extra charges of rework due to error during execution of project, wastages of material on construction site, improper site management by contractor, insufficient experience of contractor, inept sub contractor, financial difficulty experienced by contractor and outdated and unacceptable construction technique adopted by contractor are 0.670, 0.548, 0.495, 0.445, 0.385, 0.352, 0.413, 0.675 and 0.320 respectively as shown in Fig. 15.

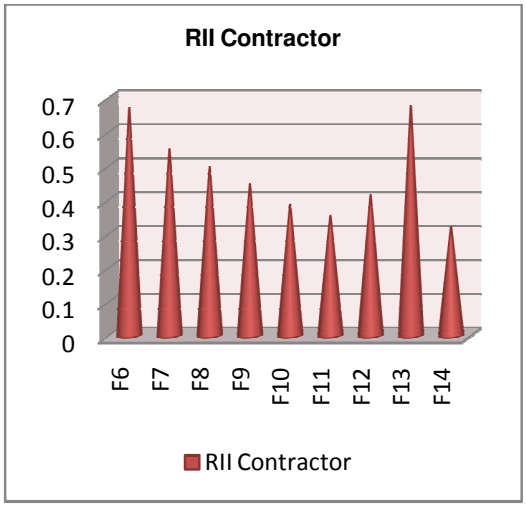


Fig. 15. RII of cluster Contractor.

RII of Cluster Consultant. RII of the factors incorrect estimation of time and cost, contract management, delay in inspecting of work, quality assurance & quality control and experience level of consultant are 0.742, 0.556, 0.391, 0.430 and 0.394 respectively as shown in Fig. 16.

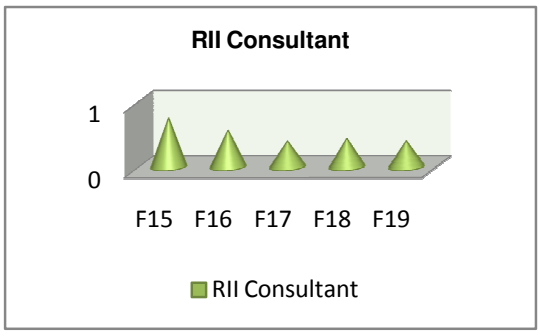


Fig. 16. RII of cluster Consultant

RII of Cluster Design. RII of the factors delay in design approval, inappropriate design & delay in making design document and frequent modification in design are 0.442, 0.478 and 0.704 respectively as shown in Fig. 17.

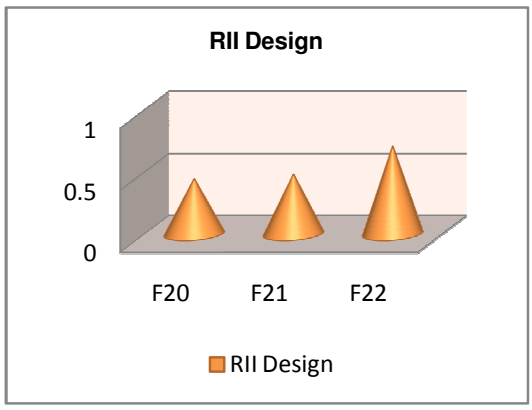


Fig. 17. RII of cluster Design.

RII of cluster Project. RII of the factors modification in the scope of the project, lowest possible bid procurement policy and quantity of extra work done are 0.556, 0.800 and 0.724 respectively as shown in Fig. 18.

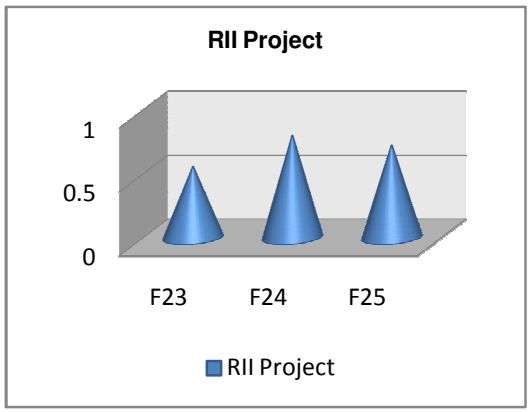


Fig. 18. RII of cluster Project.

RII of cluster Material. RII of the factors material cost, delay in delivery of material, delay in material procurement, material shortage at site and modification in material specification are 0.565, 0.343, 0.317, 0.371 and 0.411 respectively as shown in Fig. 19.

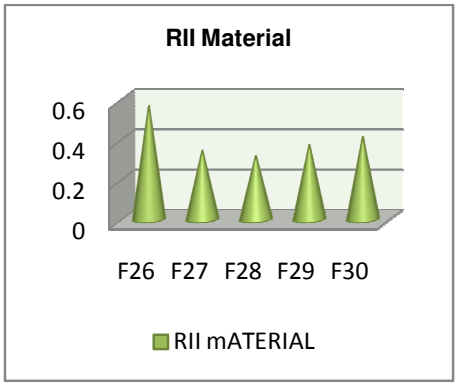


Fig. 19. RII of cluster Material.

RII of cluster Labor. RII of the factors poor qualification level of labor, dispute and strikes of labor, Less productivity level of labor, project labor charge and deficiency of labor for project are 0.341, 0.328, 0.352, 0.628 and 0.370 respectively as shown in Fig. 20.

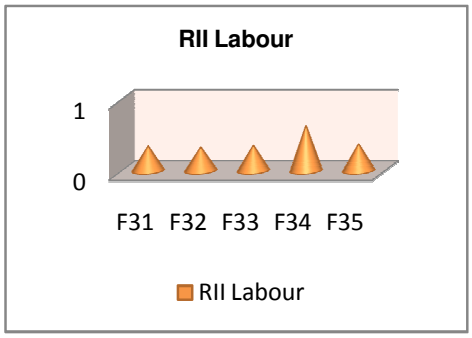


Fig. 20. RII of cluster Labour.

RII of cluster Equipment. RII of the factors equipment deficiency, equipment availability & failure at site and higher cost of machineries and equipment are 0.317, 0.302 and 0.582 respectively as shown in Fig. 21.

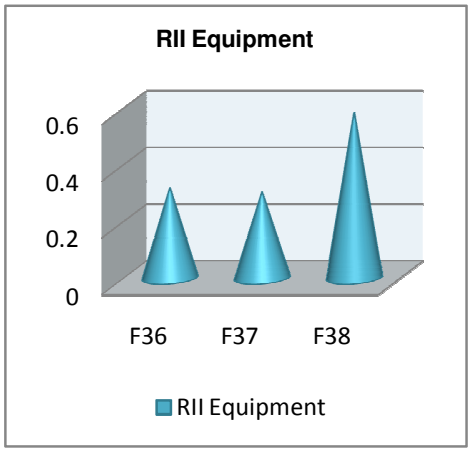


Fig. 21. RII of cluster Equipment.

RII of cluster Contract. RII of the factors errors and discrepancy in contract document and contractual procedure adopted are 0.728 and 0.410 respectively as shown in Fig. 22.

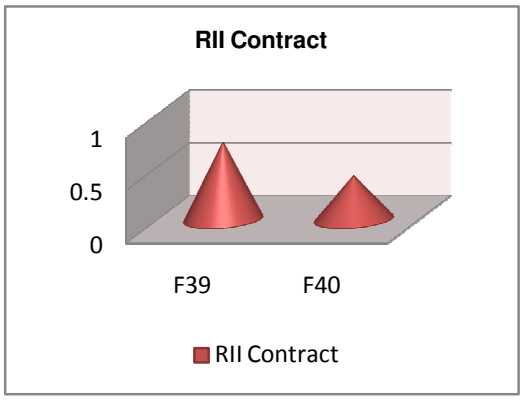


Fig. 22. RII of cluster Contract.

RII of cluster Coordination. RII of the factors are Poor coordination between construction participants, slow information flow in between construction participants and Low level of communication between construction participants are 0.659, 0.440 and 0.430 respectively as shown in Fig. 23.

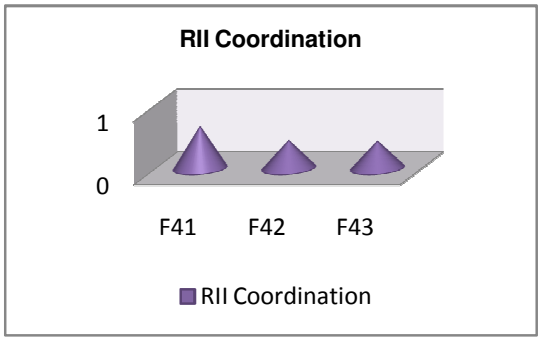


Fig. 23. RII of cluster Coordination.

RII of cluster External. RII of the factors fluctuation in the prices of materials, inflation, improper govt. policy, exchange rate, high level of interest rate of bank &loan, fraudulent practices & kickbacks, insurance charges of project, climatic and weather condition at site, social & cultural aspects, accidents and casualties during construction period, acts of god and incompatible site condition are 0.829, 0.788, 0.780, 0.621, 0.560, 0.520, 0.490, 0.485, 0.423, 0.370, 0.321 and 0.665.

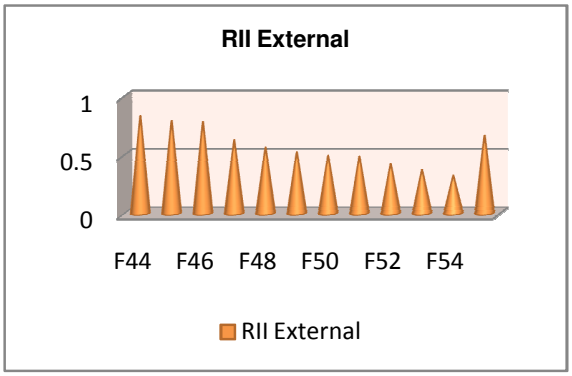


Fig. 24. RII of cluster External.

RII of the clusters. The RII of the clusters is shown in Fig. 25. RII of the clusters of ' Cluster Owner', 'Cluster Contractor', 'Cluster Consultant', 'Cluster Design', 'Cluster Project', 'Cluster Material', 'Cluster Labor', 'Cluster Equipment', 'Cluster Contract', 'Cluster Coordination/communication' and 'Cluster External' are 0.495, 0.477, 0.502, 0.541, 0.693, 0.401, 0.391, 0.4, 0.569, 0.509 and 0.570 respectively.

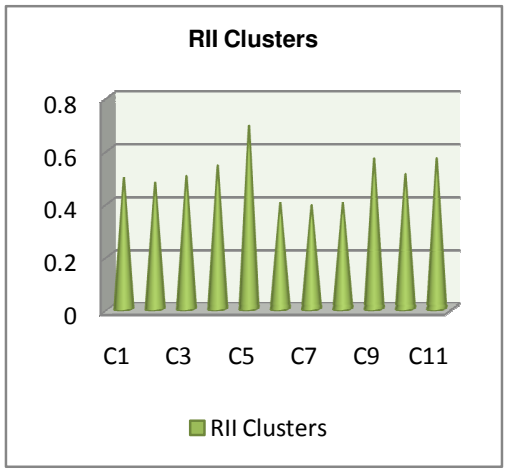


Fig. 25. RII of clusters.

B. Model development

The model is developed in two stages. First the probability of cost overrun due to each cluster is assessed by the probability level of the factors of respective cluster such as the probability of cost overrun due to the 'Cluster Owner' is estimated by probability of cost overrun due to the factors of 'Cluster Owner'.

The probability of cost overrun of project is then assessed by the probability output of each cluster as determined in the previous step.

Define input and output variables. For the first stage of the model development the cost overrun factors are taken as the input variables and output probability of cost overrun due to the cluster is determined from the model. The input and output variables for 'Cluster Owner' are described in Table 1. The various factors of cost overrun of 'Cluster Owner' are Slow progress payment by owner for completed work, Slow procedure in taking decisions, Manner of finance, Level of owner interference and Impracticable contract duration. These factors are considered as input variable for determining the output 'the probability of cost overrun due to the 'Cluster Owner'. During model development these factors are considered as F1, F2, F3, F4, and F5. And the output variable is abbreviated as 'POC1'. Similarly, probability of cost overrun due to the 'Cluster Contractor', 'Cluster Consultant', 'Cluster Design', 'Cluster Project', 'Cluster Material', 'Cluster Labor', 'Cluster Equipment', 'Cluster Contract', 'Cluster Coordination/communication' and 'Cluster External' is determined using factors related to that cluster.

As mentioned earlier the cost overrun probability of the project is determined taking into consideration the probability of all the clusters. Therefore for the second stage of model development clusters such as 'Clusters Owner', 'Cluster Contractor', 'Cluster Consultant', 'Cluster Design', 'Cluster Project', 'Cluster Material',

'Cluster Labor', 'Cluster Equipment', 'Cluster Contract', 'Cluster Coordination/communication' and 'Cluster External' are considered as input variables.

Table 1: Input and output variables for determining 'probability of cost overrun due to cluster 'Owner'.

Input variables		Output variable
F1	Slow progress payment by owner for completed work	probability of cost overrun due to cluster' C1 (Owner) POC1
F2	Slow procedure in taking decisions	
F3	Manner of finance	
F4	Level of owner interference	
F5	Impracticable contract duration	

And the output variable is 'probability of cost overrun of project' as shown in Table 2. During model development these clusters are considered as C1, C2, C3, C4, C5, C6, C7, C8, C9, C10 and C11.and the output variable is 'Probability of cost overrun of project' which is abbreviated as **'PCOP'.**

Table 2: Input and output variables for determining the probability of cost overrun of project.

Input variables		Output variable
C1	Owner	Probability of cost overrun of project (PCOP)
C2	Contractor	
C3	Consultant	
C4	Design	
C5	Project	
C6	Material	
C7	Labour	
C8	Equipment	
C9	Contract	
C10	Coordination	
C11	External	

The input variables in the form of factors and output probability of owner related 'POC1' cluster are shown in the FIS Editor window of Fuzzy Logic tool box of MATLAB Program Software in Fig. 26.

Fuzzy Membership Functions. The membership function of input variables in the form of factors and clusters of factors and output variable Probability of cost overrun of project 'PCOP 'were generated using the membership function editor in the Matlab fuzzy logic toolbox as shown in Fig. 27.

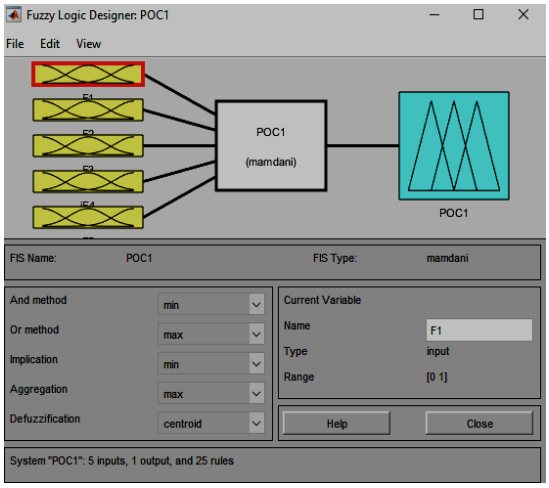


Fig. 26. Defining Input and output variables.

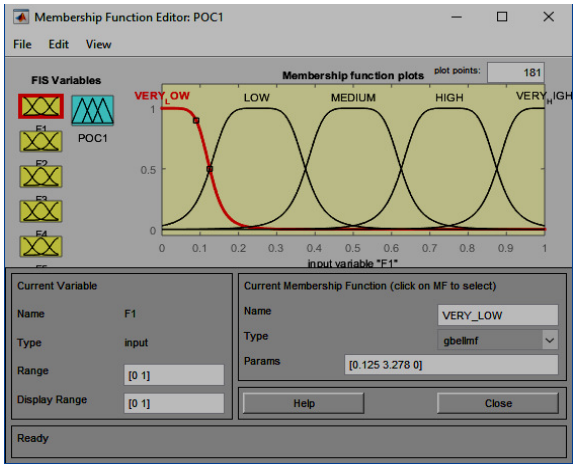


Fig. 27. Defining membership of Input and output variables.

Fuzzy rules and weighting of rules. Considering 55 cost overrun factors, total 275 rules are formed. Samples of the fuzzy rules of cluster 'Owner' are presented in Table 3.

Table 3: Sample fuzzy rules of factors of 'Cluster Owner'.

Rule	Antecedent	consequence	weighting
1	If the probability of cost overrun due to the factor 'Slow progress payment by owner for completed work ' is very low	Then the probability of cost overrun due to Cluster 'owner' is very low	0.498
6	If the probability of cost overrun due to the factor 'Slow procedure in taking decisions' is very low	Then the probability of cost overrun due to Cluster 'owner' is very low	0.522
11	If the probability of cost overrun due to the factor 'Manner of finance ' is very low	Then the probability of cost overrun due to Cluster 'owner' is very low	0.425
16	If the probability of cost overrun due to the factor 'Level of owner interference' is very low	Then the probability of cost overrun due to Cluster 'owner' is very low	0.357
21	If the probability of cost overrun due to the factor 'Impracticable contract duration' is very low	Then the probability of cost overrun due to Cluster 'owner' is very low	0.676

Table 4: Sample Fuzzy rules for estimating the probability of cost overrun of the project.

Rule	Antecedent	Consequence	Weighting
1	If the probability of the cost overrun due to 'Cluster Owner' is very low	Then the 'probability of cost overrun of the project' is very low	0.495
6	If the probability of the cost overrun due to the 'Cluster Contractor' is very low	Then the 'probability of cost overrun of the project' is very low	0.477
11	If the probability of the cost overrun due to the 'Cluster Consultan' is very low	Then the 'probability of cost overrun of the project' is very low	0.502
16	If the probability of the cost overrun due to the 'Cluster Design' is very low	Then the 'probability of cost overrun of the project' is very low	0.541
21	If the probability of the cost overrun due to the 'Cluster Project' is very low	Then the 'probability of cost overrun of the project' is very low	0.693
26	If the probability of the cost overrun due to the 'Cluster Material' is very low	Then the 'probability of cost overrun of the project' is very low	0.401
31	If the probability of the cost overrun due to the 'Cluster Labour' is very low	Then the 'probability of cost overrun of the project' is very low	0.391
36	If the probability of the cost overrun due to the 'Cluster Equipment' is very low	Then the 'probability of cost overrun of the project' is very low	0.4
41	If the probability of the cost overrun due to the 'Cluster Contract' is very low	Then the 'probability of cost overrun of the project' is very low	0.569

Since the probability of cost overrun due to the cluster 'Owner' is directly related to the probability of cost overrun due to slow progress payment by owner for completed work, slow procedure in taking decisions, Manner of finance, Level of owner interference and Impracticable contract duration, therefore the probability of cost overrun of these factors will influence the probability of cost overrun due to the cluster 'Owner'. The calculated RII of the factors are the weighting of the rules.

Considering 11 clusters of cost overrun factors, total 55 rules are formed. Since the probability of cost overrun of the project is directly related to the probability of cost overrun due to the 'Clusters Owner', 'Cluster Contractor', 'Cluster Consultant', 'Cluster Design', 'Cluster Project', 'Cluster Material', 'Cluster Labor', 'Cluster Equipment', 'Cluster Contract', 'Cluster Coordination/communication' and 'Cluster External'. Therefore the probability of cost overrun due to these clusters will influence the probability of cost overrun of the project. Fuzzy rules for estimating the probability of cost overrun of the project are presented in Table 4.

For fuzzy inference the rules are made as follows:

If F1 is 'very low' then POC1 is 'very low': Which means that if the probability of cost overrun due to the factor 'slow payment of complete work by owner' is 'very low' then the probability of cost overrun due to Cluster 'Owner' is 'very low'. The rule editor window of Fuzzy Logic tool box of MATLAB Program Software is shown in Fig. 28.

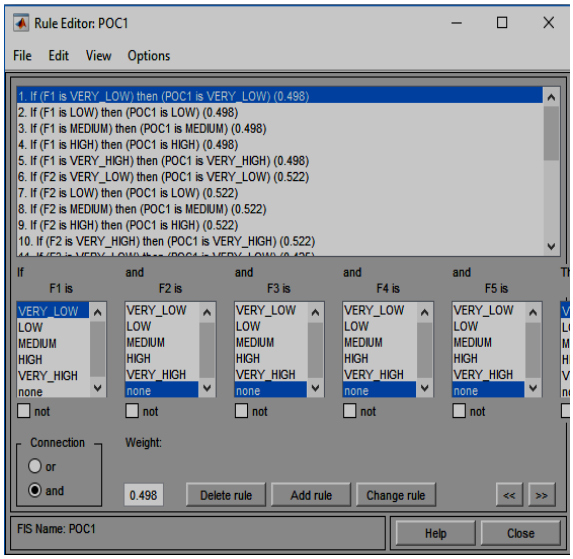


Fig. 28. Rule formation.

Defuzzification. Finally the 'probability of cost overrun due to Cluster 'Owner' PCO1 and 'probability of cost overrun' of project' is estimated by defuzzification by 'centroid of area' method as shown in Fig. 29.

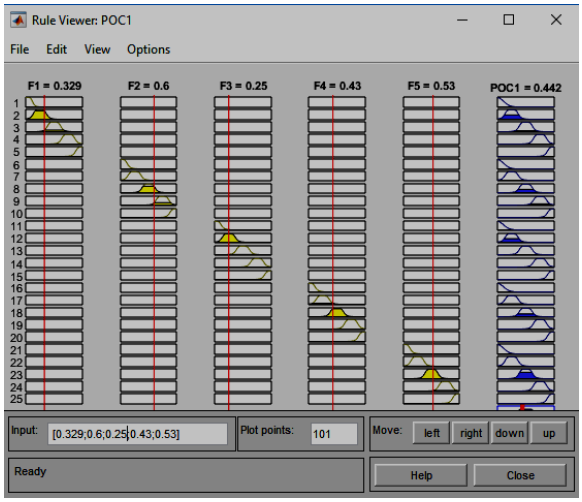


Fig. 29. Defuzzification process.

C. Testing of the model

To test the reliability of the proposed fuzzy cost overrun assessment model, an interview has been conducted with a team of the experts of a leading Indian construction company. The panel of the experts consists of the top 10 executive engineers, project managers and site engineers of the company. The experts were requested to examine the cost overrun factors considered for this research and filled in the required information regarding cost overrun factors to assess the proposed model. They were also requested to estimate the probability of cost overrun of the project. The information provided by the group of the experts are presented as shown in the following table in the Table 5.

Table 5: Information provided by the group of the expert.

S.No.	Factors	Probability of occurrence of factor
1	F1	0.329
2	F2	0.6
3	F3	0.25
4	F4	0.43
5	F5	0.53
6	F6	0.4
7	F7	0.3
8	F8	0.25
9	F9	0.3
10	F10	0.5
11	F11	0.35
12	F12	0.25
13	F13	0.55
14	F14	0.25
15	F15	0.42
16	F16	0.36
17	F17	0.27
18	F18	0.38
19	F19	0.23
20	F20	0.35
21	F21	0.62
22	F22	0.58
23	F23	0.22
24	F24	0.8
25	F25	0.5
26	F26	0.6
27	F27	0.33
28	F28	0.28
29	F29	0.15
30	F30	0.24
31	F31	0.27
32	F32	0.16
33	F33	0.23
34	F34	0.32
35	F35	0.21
36	F36	0.18
37	F37	0.21
38	F38	0.23
39	F39	0.25
40	F40	0.27
41	F41	0.56
42	F42	0.42
43	F43	0.38
44	F44	0.87
45	F45	0.82
46	F46	0.93
47	F47	0.62
48	F48	0.72
49	F49	0.4
50	F50	0.43
51	F51	0.85
52	F52	0.35
53	F53	0.27
54	F54	0.23
55	F55	0.74

The determined probabilities of cost overrun due to clusters by model are given in Table 6.

Table 6: Probability level of clusters.

S.No.	Cluster	Name of cluster	Probability of cost overrun due to cluster	Probability of cost overrun of the project
1	C1	Owner	0.442	0.435
2	C2	Contractor	0.405	
3	C3	Consultant	0.406	
4	C4	Design	0.476	
5	C5	Project	0.524	
6	C6	Material	0.436	
7	C7	Labour	0.289	
8	C8	Equipment	0.26	
9	C9	Contract	0.263	
10	C10	Coordination	0.465	
11	C11	External	0.609	

The probability of cost overrun of the project was calculated by the experts was in the range of 0.35-0.50. Provided information was tested using the proposed model and The probability of cost overrun of the project obtained was 0.435 as shown in Fig. 30, which is much closer to the estimated cost risk level of the project estimated by the panel of the experts of the company.

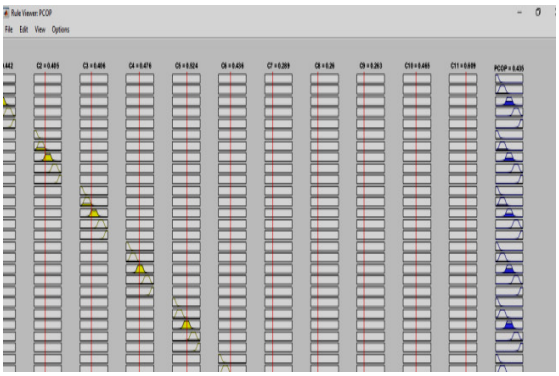


Fig. 30. Probability of cost overrun of the project.

V. CONCLUSION

It is observed that very few projects get completed within stipulated costs in India, despite the importance of economy in construction projects. The cost of a project is calculated just by adding some contingency charges and proper analysis of cost overrun risk is ignored, consequently many projects fail to complete the project in stipulated budget. Therefore in this study a fuzzy model for forecasting the probability of cost overrun risk in Indian construction project is developed by identifying, classifying and ranking of the factors responsible for the cost overrun in the construction industries. The fuzzy theory is chosen for developing the model as the theory is found suitable for quantifying the probability of cost overrun risk of the construction project when the variables regarding the cost overrun factors are expressed in linguistic terms. The fuzzy set theory (FST) is equipped to take care of imprecise, incomplete and uncertain data expressed in linguistic form. Through an extensive literature survey and discussion with the experts of the Indian construction industry, fifty five cost overrun factors were identified and then classified into eleven clusters. A questionnaire survey was conducted in Indian construction industry for data collection to calculate relative importance index (RII) of the recognized factors causing cost overrun and clusters. Total 250 questionnaires were distributed to the potential respondents within the Indian construction company comprising owner, contractor and consultant. Out of 250 questionnaire set, only 135 responses were returned. The calculated relative importance index of the factors and clusters was used to develop model using the fuzzy logic tool box of MATLAB program software. First the probability of cost overrun due to the clusters of cost overrun factors is determined. The final cost overrun risk probability of the project is then determined taking into consideration the probability of all the clusters. For all input and output linguistic variables five membership functions are defined using g-bell shaped fuzzy numbers. Fuzzy rules for estimating the probability of cost overrun of the project are constructed with the discussion of experts of the Indian construction industry in the simple forms of the Mamdani-style. The calculated RII of the factors and clusters are taken as the weighting of the rules. Centre of area (COA) method is chosen for defuzzification. The model has been tested by taking into consideration a real case study in India. To test the model an interview was conducted with a team of the experts of a leading Indian construction company. The panel of the experts consisted of the top 10 executive engineers, project managers and site engineers of the company. The experts were requested to examine the cost overrun factors considered for this research and filled in the required information regarding cost overrun factors to assess the proposed model. They were also requested to estimate the cost risk level of the project. The probability of cost overrun of the project was calculated by the experts was in the range of 0.35-0.50. Provided information was tested using the proposed model and The probability of cost overrun of the project obtained was 0.435, which is much closer to the estimated cost risk level of the project estimated by the panel of the experts of the company. With the help of the proposed model, it is possible to guide project managers and decision makers to make more informative decisions such as contingency estimation, mark-up estimation, bid price, selection of optimum procurement route, evaluation of different projects and preparation of project portfolios.

VI. LIMITATION OF APPROACH AND FUTURE SCOPE

The following are the limitations of the proposed model: The sample size of population for the identification cost overrun factors is very small as compared to the size of construction industry. Therefore the calculated the relative importance index (RII) does not represent a true picture of actual cost overrun scenario. The model can differ as per the shapes and number of membership function adopted in fuzzy logic tool. The constructed rules for fuzzy model are based on expert opinion. It is based on subjectivity. For different attitudes of the experts rules can also be changed and in this condition the model will produce different result. Results can vary according to the defuzzification method adopted for obtaining output. Fuzzy rules are based on literature survey findings expert judgment and, it is clear that a completely different model may be proposed by other researchers based on different expert opinions. The model could also be developed for estimating the performance and delay probability in construction and other industrial projects using fuzzy logic theory.

Conflict of Interest. No.

REFERENCES

[1]. Hajarat, A.D., & Smith J. N. (1993). Exposure envelopes: an assessment of the exposure to time and cost overruns during construction projects. *Int. J. Proj. Manag., 11*(4), 227–231.

[2]. Azhar, N., Farooqui R.U., & Ahmed, S.M. (2008). Cost overrun factors in construction industry of Pakistan. In: Paper presented at the first international conference on construction in developing countries (ICCIDC-I). Advancing and integrating construction education, research & practice.

[3]. Memon, A.H., Rahman, I.A., Asmi, A., & Azis, A. (2010). Factors Affecting Construction Cost in Mara Large Construction Project: Perspective of Project Management Consultant. *International Journal of Sustainable Construction Engineering & Technology, 1*(2), 41-54.

[4]. Akinci, B., & Fischer, M. (1998). Factors affecting contractors' risk of cost overburden, *Journal of Management in Engineering, 14*(1), 67-76.
[5]. Reina, P., & Angelo, W. J. (2002). Megaprojects need more study up front to avoid cost overruns. *ENR*, *249*(3).
[6]. Frame, J. D. (1997). Establishing project risk assessment teams. In: K. Kähkönen & K. Artto (eds.), Managing risks in projects, London: E & FN Spon, 17-19.
[7]. Jackson, S. (2002). Project cost overrun and risk management in Greenwood. In: Proceedings of Association of Researchers in Construction Management, 18th Annual ARCOM Conference, Newcastle, Northumber University, UK, 2-4.
[8]. Chang, A.S.T. (2002). Reasons for Cost and Schedule Increase for Engineering Design Projects. *Journal of Management in Engineering, 18*(1), 29–36
[9]. Olawale, Y.A., & Sun, M. (2010). Cost and time control of construction projects: Inhibiting factors and mitigating measures in practice. *Construction Management and Economics, 28*(5), 509–526.
[10]. Cantarelli, C.C., Flyvbjerg B., Molin, E.J.E., & Wee, B. V. (2010). Cost *Overruns* in Large-Scale Transportation Infrastructure Projects: Explanations and Their *Theoretical* Embeddednes. *European Journal of Transport and Infrastructure Research*, *10*(1), 5-18.
[11]. Koushki P.A., Al-Rashid, K., & Kartam, N. (2005). Delays and cost increases in the construction of private residential projects in Kuwait. *Construction Management and Economics*, *23*(3), 285–294.
[12]. Gupta, N. (2009). Avoiding Time and Cost Overruns in the Construction of Rohtang Tunnel. Institute for Defence Studies and Analyses.
[13]. Frimpong, Y., Oluwoye, J., & Crawford, L. (2003). Causes of delays and cost overruns in construction of groundwater projects in developing countries; Ghana as a case study. *International Journal of Project Management, 21*, 321-326.
[14]. KPMG (2019). Study on Assessment of infrastructure projects and corrective recommendations for performance. A joint study conducted by PMI and KPMG in India on infrastructure projects in India, supported by Ministry of Statistics and Programme Implementation (MoSPI).
[15]. Zhi, H. (1995). Risk management for overseas construction projects. *International Journal of Project Management, 13*(4), 231-237.
[16]. Baccarini D., & Archer, R. (2001). The risk ranking of projects: a methodology. *International Journal of Project Management 19*(3), 139-145.
[17]. Hull, J. (1990). Application of risk analysis techniques in proposal assessment. *International Journal of Project Management, 8*(3), 152-157.
[18]. Özta, A, & Ökmen, Ö (2004). Risk analysis in fixed-price design-build construction projects. *Building and environment* 39(2)**,** 229-237.
[19]. Thomas, A., Kalidindi, S., & Ganesh, L. (2006). Modelling and assessment of critical risks in BOT road projects. *Construction Management and Economics, 24*(4)**,** 407-424
[20]. Dey P K (2001). Decision support system for risk management: a case study. *Management Decision, 39*(8): 634-649.
[21]. Dikmen, I., Birgonul, & M. T. (2006). An analytic hierarchy process based model for risk and opportunity assessment of international construction projects. *Canadian Journal of Civil Engineering, 33*(1): 58-68.
[22]. Nasir, D., McCabe, B., & Hartono, L. (2003). Evaluating Risk in Construction–Schedule Model (ERIC–S): Construction Schedule Risk Model. *Journal of Construction Engineering and Management, 129*, 518.
[23]. Kangari, R., & Riggs, L. S. (1989). Construction risk assessment by linguistics. *IEEE transactions on engineering management*, *36*(2), 126-131.
[24]. Tah, J., & Carr, V. (2000). A proposal for construction project risk assessment using fuzzy logic. *Construction Management and Economics, 18*(4), 491-500.
[25]. Sharma, S., & Goyal, P. K. (2014). Cost overrun factors and project cost risk assessment in construction industry- a state of the art review. *International journal of civil engineering, 3*(3), 139-154.
[26]. Oliveros, A. V. O., & Fayek, A. R. (2005). Fuzzy logic approach for activity delay analysis and schedule updating. *J. Constr. Eng. Manage*, *131*(1), 42–51.
[27]. Dikmen, I., Birgonul, M.T., & Han, S. (2007b). Using fuzzy risk assessment to rate cost overrun risk in international construction projects. *International Journal of Project Management, 25*(5), 494-505.
[28]. KarimiAzari, A., Mousavi, N., Mousavi, S. F., & Hosseini, S. (2011). Risk assessment model selection in construction industry. *Expert Systems with Applications*, *38*(8), 9105-9111.
[29]. Nieto-Morote, A., & Ruz-Vila, F. (2011). A fuzzy approach to construction project risk assessment. *International Journal of Project Management, 29*, 220–231.
[30]. Yeung, J., Chan, A., & Chan, D. (2012). Fuzzy set theory approach for measuring the performance of relationship-based construction projects in Australia. *J. Manage. Eng.,* 10.1061/(ASCE)ME.1943-5479 .0000083, 181–192.
[31]. San Cristobal, J. (2013). Critical path definition using multi criteria decision making: The PROMETHEE method. *J. Manage. Eng.,* 10.1061/(ASCE)ME.1943-5479.0000135, 158–163.
[32]. Gunduz, M., Nielsen, Y., & Ozdemir, M. (2015). Fuzzy Assessment Model to Estimate the Probability of Delay in Turkish Construction Projects. *Journal of Management in Engineering, ASCE,* DOI: 10.1061/(ASCE)ME.1943-5479.0000261
[33]. Shaktawat, A., & Vadhera, S. (2016, December). Fuzzy logic based determination of cost overrun of hydro power plant. In *2016 International Conference on Electrical Power and Energy Systems (ICEPES)* (pp. 301-304). IEEE.
[34]. Khalek, H. A. E., Aziz, R. F. & Kamel, H. M. (2016). Risk and Uncertainty Assessment Model in Construction Projects Using Fuzzy Logic. *American journal of civil engineering, 4*(1), 24-39.
[35]. Manoliadis, O. (2018). The Assessment of Construction Project Risks with the Use of Fuzzy Delphi Methodology- Case Study Photovoltaic Project in Greece. *International Journal of Management and Fuzzy Systems, 4*(1), 7-14.
[36]. Fuzzy Logic Toolbox™ 2 User's Guide (2008), MATLAB, The MathWorks, Inc.
[37]. Zadeh, L.A. (1965). Fuzzy sets. *Information and control, 8*(3), 338-353.
[38]. Mamdani, E. H., & Assilian, S. (1975). An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, *7*, 1-13.
[39]. Mittal, H., & Bhatia, P. (2007). Optimization criteria for effort estimation using fuzzy technique. *CLEI Electronic Journal*, *10*(1), 1-11.
[40]. Sharma, S. and Goyal, P. K. (2019). Applying "Fuzzy Techniques" in construction project management. *International Journal on Emerging Technologies*, *Research Trend, 10*(2): 384-391.

How to cite this article: Sharma, Savita, Goyal, Pradeep K. and Chhipa, R. C. (2020). Forecasting the Probability of Cost Overrun Risk of Indian Construction Projects using Fuzzy Model. *International Journal on Emerging Technologies*, *11*(1): 10–22.

View publication stats

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Fuzzy Logic based Automatic Generation Control

MD Shadan Salik
Electrical Engineering
Delhi Technological University
Delhi, India
shadansaik@gmail.com

Shaurya Priyal
Electrical Engineering
Delhi Technological University
Delhi, India
shauryapriyal1312@gmail.com

Neha Khokher
Electrical and Electronics Engineering
Delhi Technological University
Delhi, India
nehakhokher9@gmail.com

DR. Narendra Kumar
Electrical Engineering
Delhi Technological University
Delhi, India
narendra.kumar@dce.edu

***Abstract*— Automatic Generation Control (AGC) is a controlling technique used in electrical power generation to maintain a constant nominal frequency and power interchange with other linked systems by creating a balance of the generation of the system with system load losses. The frequency of the system deviates from its desired value if there is any sudden load change. If the frequency deviates in a large amount the system may get collapsed. Hence, a fast and accurate controller action is required to maintain stability and nominal frequency. For AGC many controlling techniques have been studied. In this paper, a PI controller is compared with a fuzzy controller. A two area system is considered both being thermal. Different step load changes have been considered and the performance of both the controllers has been observed in different load conditions. Simulation results have been brought up to show the better working of the fuzzy controller than the PI controller.**

Keywords— AGC, area control error, fuzzy logic controller, pi controller, two area system

I. INTRODUCTION

The purpose of electrical power generating units is to generate a certain amount of power that fulfills the required amount of load in a specific region. These generating units are referred to as power plants. There are different types of plants like a thermal plant, hydro plant, nuclear plant, etc. The load conditions do not remain the same throughout the time. It changes very rapidly and not only in one direction but in both directions that is it may go higher or lower than the normal demand. It depends on the consumers how much they are utilizing the load. Due to these variations in load demand the system frequency gets affected. It deviates from the nominal frequency. The deviation of frequency from its nominal value is undesirable as it affects the speed of the machine. The speed of a synchronous machine which is used as in generating units directly depends on the frequency. Therefore, as the frequency fluctuates, the speed of machines also fluctuates due to which the output of the machine is inconsistent. Hence the frequency of the system must remain constant.

To maintain a constant frequency of the system a control scheme is known as Automatic Generation Control (AGC) is used that controls the amount of the steam entering the turbine [1]. The primary purpose of the controller is to set the frequency of the system uniform whatever the changes in load demand may be [7]. As the load demand changes the frequency suddenly increases or decreases, immediately the current frequency is compared with a reference frequency or nominal frequency and according to it an error signal is sent to the controller which the takes adjust the amount of steam valve to be open or closed. Certain limits of load fluctuation are controlled by the controller or AGC but if load fluctuations are very high then the controllers need to be reset either manually or automatically [1]. The controller has a very vital role in AGC, and its proper and fast functioning is a must.

There are a lot of controllers that are used like P, PI, PID, fuzzy, genetic algorithm, PSO, etc [8-10]. Different controllers have different advantages. Many works have been followed out for inspecting different controller's behavior by considering a certain load change in a certain area. In this paper different performance of conventional PI controller has been contrasted with the fuzzy logic controller. A two area system is simulated in MATLAB Simulink. Different load changes are given to area 1 of the two area system. First taking the PI controller and then fuzzy controller the system is simulated and frequency variations and power variations are plotted. These plots are presented to show a comparative analysis of both the controllers.

II. CONTROL AREA CONCEPT

A power system in which a common generation control technique is used is known as a control area [2],[9]. Many control areas are attached through tie-lines [3]. In a control area, the action of a controller is to regulate the amount of inlet of the steam valve so that the amount of steam input to the turbine is regulated. The controller uses Area Control Error to bring the frequency at a nominal value [4].

Area Control Error (ACE) is the variation between actual and scheduled power generation in a control area considering frequency bias setting [1],[2]. In this paper two area thermal-thermal system is considered.

Mathematically ACE is given as:

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:53:05 UTC from IEEE Xplore. Restrictions apply.

For area 1: $ACE_1 = \Delta P_{12} + B_1 \Delta f_1$
For area 2: $ACE_2 = \Delta P_{21} + B_2 \Delta f_2$

where:

ΔP_{12} and ΔP_{21} denotes the departures from the scheduled interchange
B_1 and B_2 denotes frequency bias coefficient of area 1 and area 2 individually.
Δf_1 and Δf_2 area frequency deviation of area 1 and area 2 individually.

III. TWO AREA SYSTEM

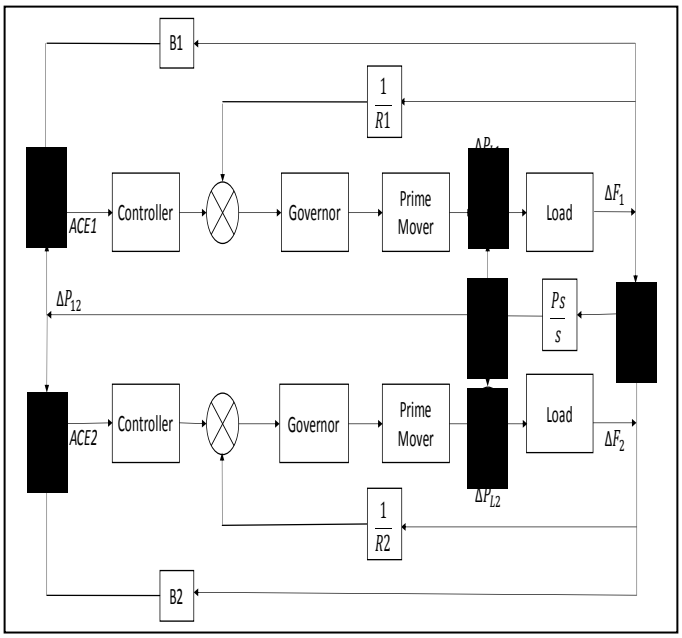


Fig 1: Two area thermal-thermal system block diagram

As shown in the above figure each area of the two-area system has its own controller and both are connected through a tie-line. Load changes (ΔP_L) may occur in any of the area which results in the frequency variations of area 1 i.e. ΔF_1 and in area 2 i.e. ΔF_2.
In this paper load changes of 0.1 per unit (pu), 0.2 pu and 0.3 pu have been considered in area 1. The controllers are first taken as PI controller and then fuzzy logic controller and then the system is simulated in MATLAB Simulink.

A. Parameters used

Time constant of turbine= 0.5 s
Time constant of governor = 0.2 s
Inertia constant of governor = 5 s
Speed regulation of governor= 0.05
Rated output of turbine = 250 MW

IV. FUZZY LOGIC CONTROLLERS

The fuzzy system is one of the key agents of computational intelligence. Prof. A Zadeh proposed the idea of a fuzzy system in 1965. A Fuzzy system differs from conventional computing as conventional computing involves Boolean logic whereas a fuzzy system involves multivalued logic. It deals with uncertainties due to ambiguity, imprecision, and vagueness. Fuzzy logic is much more general than conventional logic systems. It is mostly used in the systems with uncertainties due to ambiguity, imprecision, vagueness ambiguity, randomness, partial truth, and approximation.
The fuzzy logic controller has three steps [5],[6]:

1. Fuzzification
2. Inference rule
3. Defuzzification

It takes input variables in the form of membership functions. Membership functions can be of various types like triangular, Gaussian, trapezoidal, etc. Here, a triangular membership function is used. Every input variable is assigned with a specific number of membership functions that contain values of certain limits. These values are assigned with some linguistic terms [6]. This step is known as fuzzification.
After fuzzification there comes the rule base of the fuzzy logic controller. In the rule base, rules are assigned formulated on the knowledge of the system. Outputs are also given linguistic terms and input variable mapping is done to assign a particular output linguistic variable following the given matching of input linguistic variables. The full rule base considered in the system is mentioned in the following section.
In the final step that is defuzzification, the output linguistic variables are changed to corresponding numbers which decide the controlling action accordingly.

A. Membership functions

The following variables are created using MATLAB fuzzy command with five membership functions each.

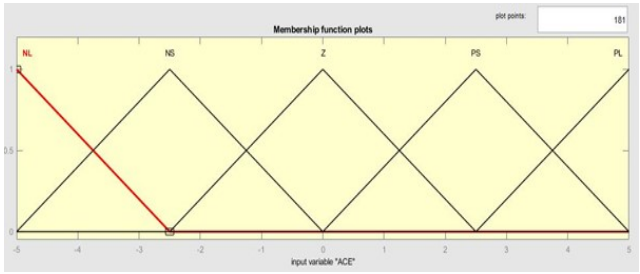


Fig 2: Input variable ACE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:53:05 UTC from IEEE Xplore. Restrictions apply.

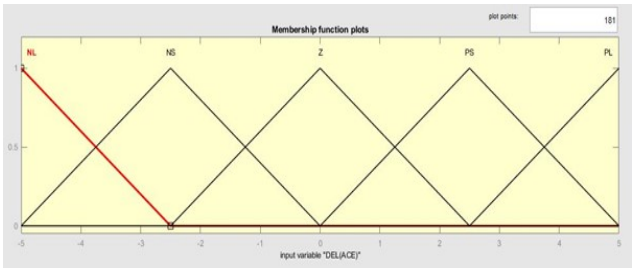


Fig 3: Input variable del(ACE)

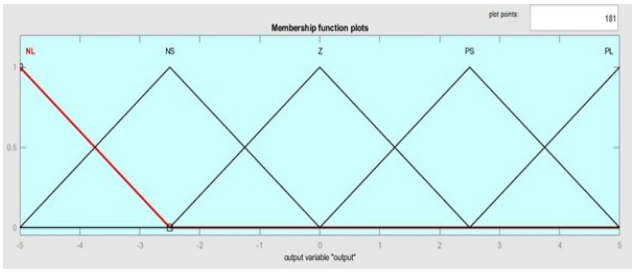


Fig 4: Output variable

B. Fuzzy Rule Base

The rules are assigned with the help of MATLAB fuzzy toolbox. Following table displays the rules that are set for input variables ACE and del(ACE) in order to give necessary output. Here the ACE and del(ACE) are used as input variables and assigned with values from -5 to 5 and given linguistic terms : More negative (NL), Less negative (NS), Zero (Z), Less positive(PS), More positive (PL).

TABLE I. RULE BASE [4]

		ACE				
		NL	**NS**	**Z**	**PS**	**PL**
DEL(ACE)	**NL**	NL	NL	NS	Z	Z
	NS	NL	NL	NS	PS	Z
	Z	NS	Z	Z	PS	PS
	PS	NS	Z	PS	PL	PL
	PL	Z	Z	PS	PL	PL

V. SIMULATION AND RESULTS

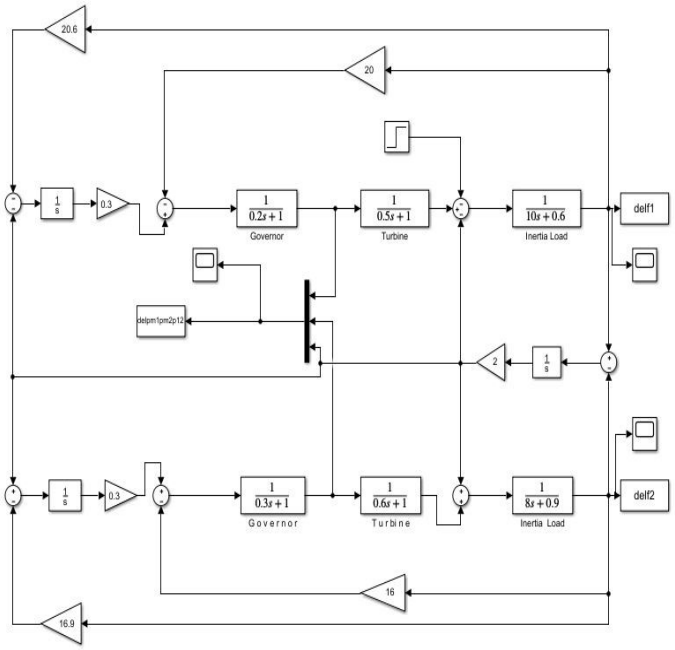


Fig 5: Two area system Simulink model by using PI control

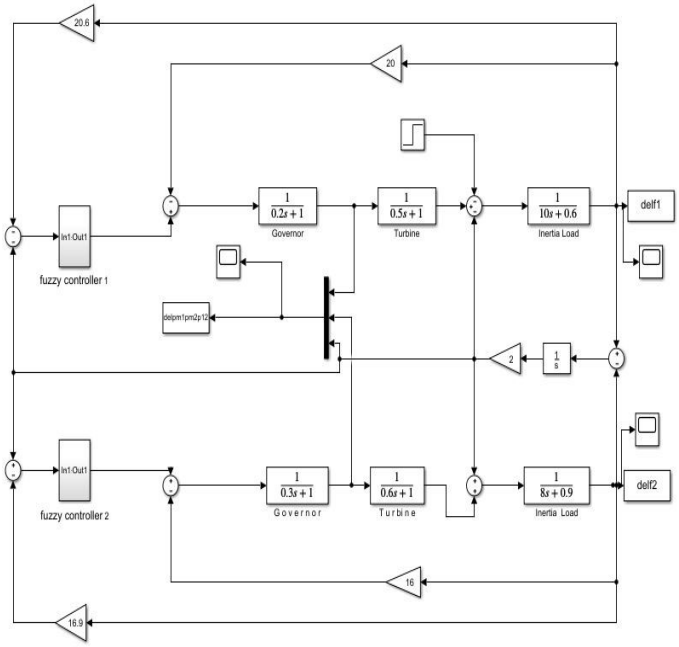


Fig 6: Two area system Simulink model by using fuzzy control

The following plots show the behavior of frequency as well as power fluctuations in both areas using a fuzzy controller and PI controller. For each load change, four plots are shown in the following sections.

A. Results for 0.1 pu load change

A change of 0.1 pu load is applied in area1 and the following results have been obtained.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:53:05 UTC from IEEE Xplore. Restrictions apply.

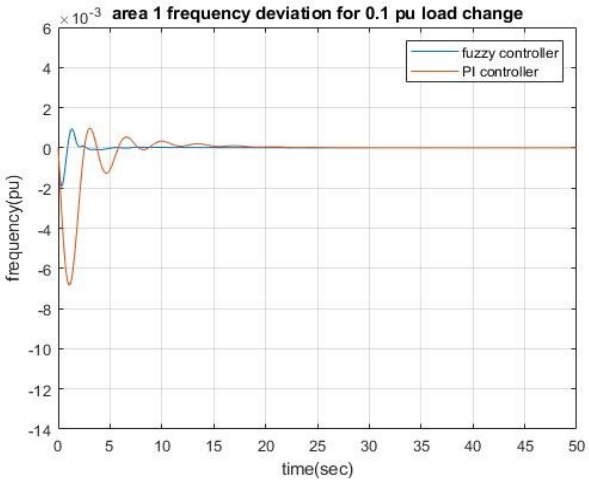


Fig 7: Area 1 frequency variations

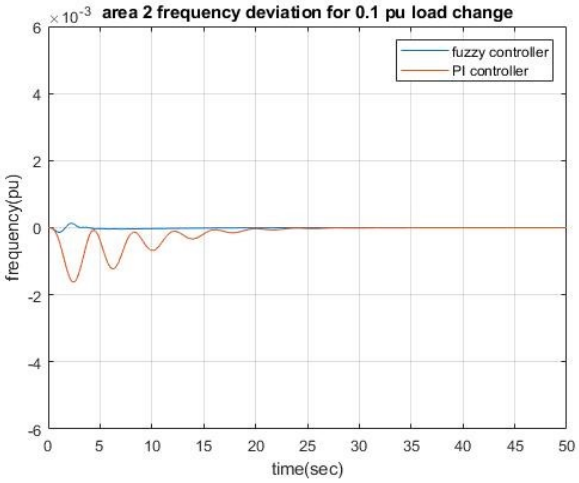


Fig 8: Area 2 frequency variations

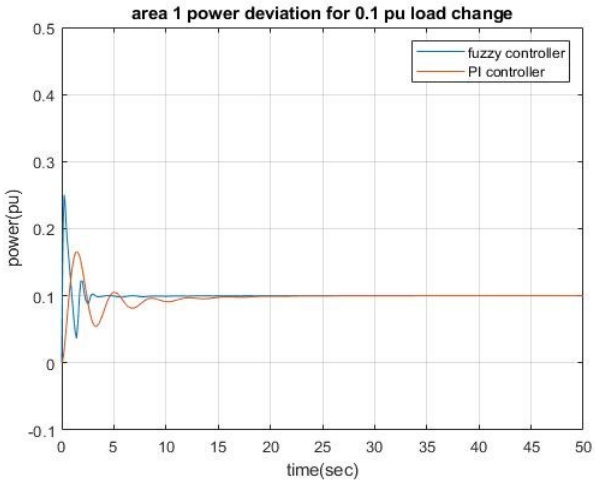


Fig 9: Area 1 power variations

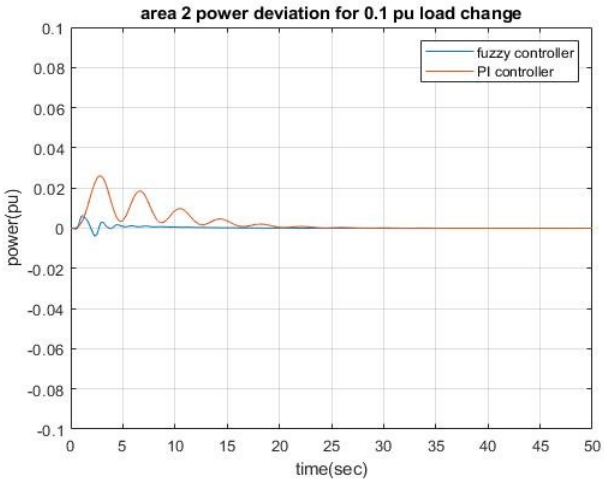


Fig 10: Area 2 power variations

B. Results for 0.2 pu load change

A change of 0.2 pu load is applied in area1 and the following results have been obtained.

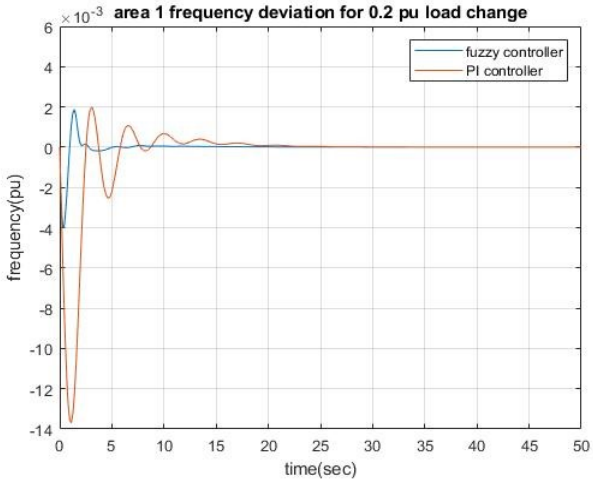


Fig 11: Area 1 frequency variations

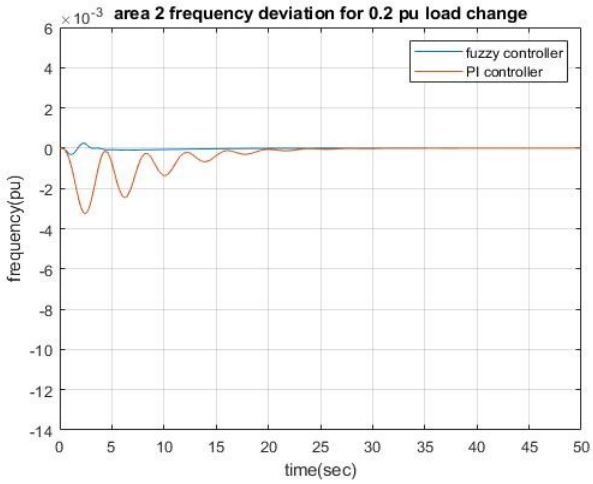


Fig 12: Area 2 frequency variations

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:53:05 UTC from IEEE Xplore. Restrictions apply.

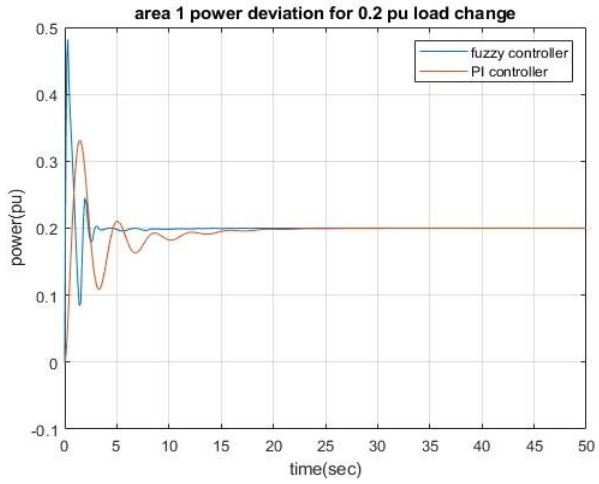


Fig 13: Area 1 power variations

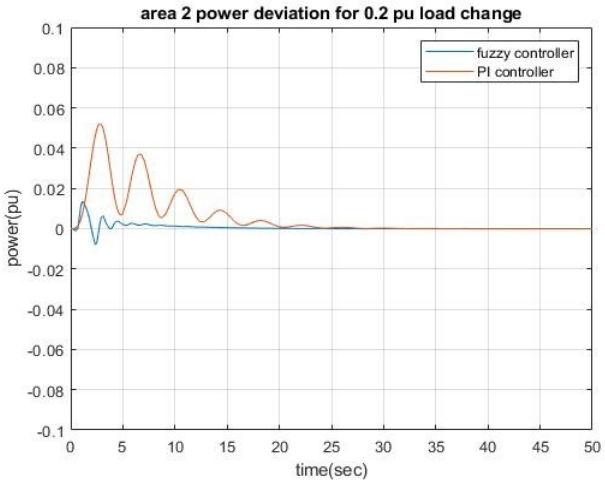


Fig 14: Area 2 power variations

C. *Results for 0.3 pu load change*

A change of 0.3 pu load is applied in area1 and the following results have been obtained.

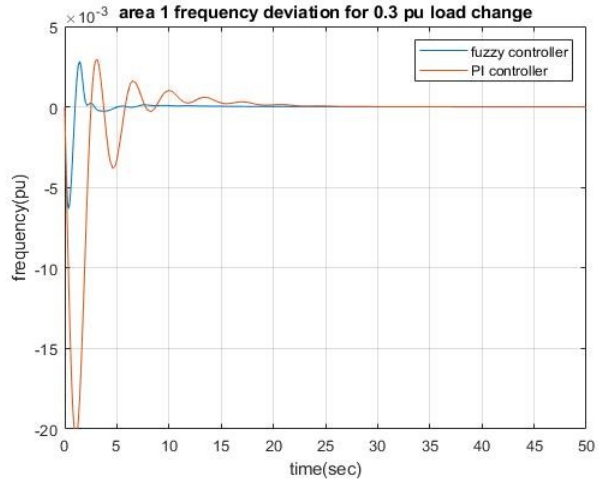


Fig 15: Area 1 frequency variations

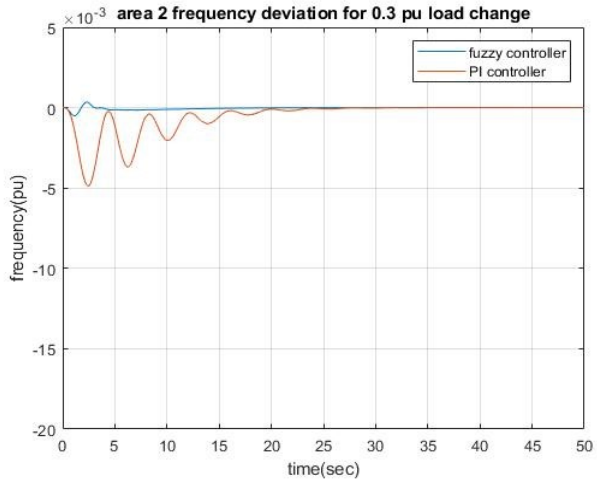


Fig 16: Area 2 frequency variations

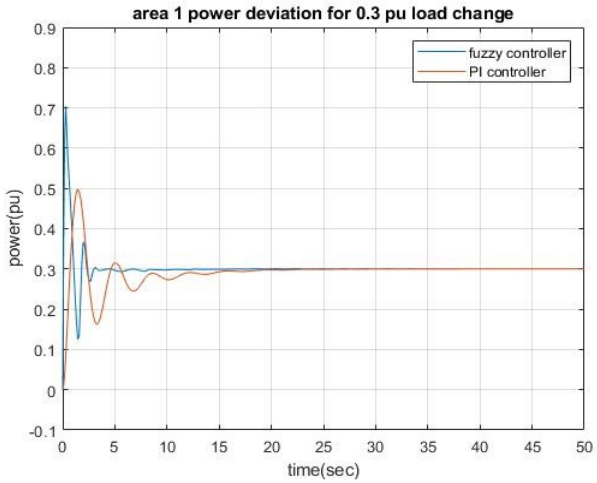


Fig 17: Area 1 power variations

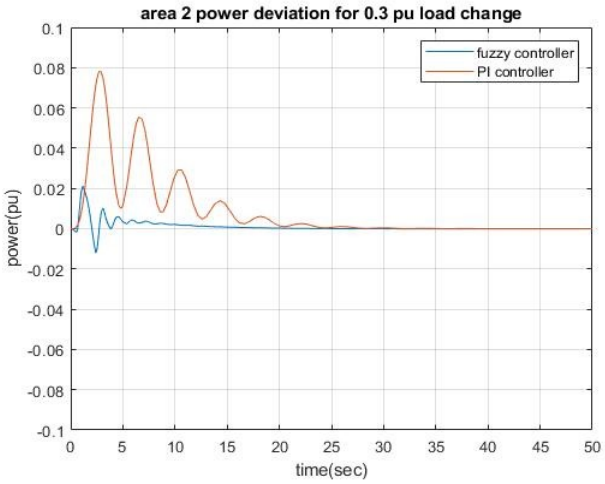


Fig 18: Area 2 power variations

From all the above plots the settling time and maximum deviation have been calculated with the help of MATLAB and following table is prepared which shows settling time and maximum fluctuation of frequency change in area 1 (ΔF1), frequency change in area 2 (ΔF2), power change in area 1 (ΔP1), power

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:53:05 UTC from IEEE Xplore. Restrictions apply.

change in area 2 (ΔP2) for all three load changes.

TABLE II. COMPARATIVE RESULT OF SETTLING TIME AND MAXIMUM DEVIATION OF FREQUENCY AND POWER

1% change	PI		Fuzzy	
	Settling time(sec)	max fluctuation (pu)	Settling time(sec)	max fluctuation (pu)
ΔF1	18.3083	0.0068	9.1211	0.0037
ΔF2	23.3675	0.0016	14.6863	0.000345
ΔP1	18.124	0.1647	11.7476	0.2463
ΔP2	23.8816	0.0261	13.8683	0.002
2% change	PI		Fuzzy	
	Settling time(sec)	max fluctuation (pu)	Settling time(pu)	max fluctuation (pu)
ΔF1	22.2716	0.0136	10.2409	0.0078
ΔF2	27.3657	0.0032	17.8777	0.0012
ΔP1	23.041	0.3282	12.7378	0.4925
ΔP2	26.8495	0.0501	16.4937	0.0039
3% change	PI		Fuzzy	
	Settling time(sec)	max fluctuation (pu)	Settling time(pu)	max fluctuation (pu)
ΔF1	23.3151	0.0201	13.4343	0.0075
ΔF2	31.3705	0.0049	18.7375	0.0017
ΔP1	22.012	0.4907	13.7400	0.7087
ΔP2	27.8761	0.0778	19.6394	0.0065

VI. CONCLUSION

By observing all the plots of frequency and power deviations of both the areas using two controlling techniques for a particular load change the frequency returns to its nominal value in lesser time in case of the fuzzy logic controller in contrast to the traditional PI controller and also power deviations are less in case of fuzzy logic controller. Moreover, due to the increased the amount of step load change, the deviation of transients of frequency and power in the conventional controller are increased much before coming to its nominal value whereas in the case of the fuzzy logic controller the deviations are not much increased by increasing the load changes. Hence fuzzy logic is a more reliable and effective controller as compared to the traditional PI controller.

LIMITATIONS AND FUTURE SCOPE

In the fuzzy controller, a triangular membership function is used. In place of triangular other membership functions can be considered like trapezoidal, Gaussian, etc. In place of fuzzy control more advanced artificial intelligence techniques can be used.

REFERENCES

[1] Saadat, H., 1999. *Power system analysis*. WCB/McGraw-Hill

[2] Wood, A.J., Wollenberg, B.F. and Sheblé, G.B., 2013. *Power generation, operation, and control*. John Wiley & Sons..

[3] Subbaraj, P. and Manickavasagam, K., 2008. Automatic generation control of multi- area power system using fuzzy logic controller. *European transactions on electrical power*, *18*(3), pp.266-280.

[4] Chandrashekar, M.J. and Jayapal, R., 2013, December. Design and comparison of I, PI, PID and Fuzzy logic controller on AGC deregulated power system with HVDC link. In *2013 International conference on Circuits, Controls and Communications (CCUBE)* (pp. 1-6). IEEE.

[5] Jain, S.K., Bhargava, A. and Pal, R.K., 2015, September. Three area power system load frequency control using fuzzy logic controller. In *2015 International Conference on Computer, Communication and Control (IC4)* (pp. 1-6). IEEE.

[6] Prakash, S., Sinha, S.K., Pandey, A.S. and Singh, B., 2009. Impact of slider gain on load frequency control using fuzzy logic controller. *ARPN J. Eng. Appl. Sci*, *4*(7), pp.20-7.

[7] Çam, E. and Kocaarslan, I., 2005. Load frequency control in two area power systems using fuzzy logic controller. *Energy conversion and Management*, *46*(2), pp.233-243.

[8] Mathur HD, Manjunath HV. Extended fuzzy logic based integral controller for three area power system with generation rate constraint. In2006 IEEE International Conference on Industrial Technology 2006 Dec 15 (pp. 917-921). IEEE.

[9] AnilKumar A, Srikanth NV. Load frequency control for diverse sources of interconnected two area power system: An adaptive fuzzy approach. In2016 International Conference on Control, Computing, Communication and Materials (ICCCCM) 2016 Oct 21 (pp. 1-4). IEEE.

[10] Moon YH, Ryu HS, Kim B, Kim S, Park SC. Fuzzy logic based extended integral control for load frequency control. In2001 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No. 01CH37194) 2001 Jan 28 (Vol. 3, pp. 1289-1293). IEEE.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:53:05 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Generating Visible Spectrum Images from Thermal Infrared using Conditional Generative Adversarial Networks

1st Neeraj Bhat
Dept. of Computer Science and Engineering
Delhi Technological University
New Delhi, India
neerajbhat98@gmail.com

2nd Navneet Saggu
Dept. of Computer Science and Engineering)
Delhi Technological University
New Delhi, India
navneetbt2k16@dtu.ac.in

3rd Pragati
Dept. of Computer Science and Engineering
Delhi Technological University
New Delhi, India
pragatisingh7890@gmail.com

4th Sanjay Kumar
Dept. of Computer Science and Engineering
Delhi Technological University
New Delhi, India
sanjay.kumar@dtu.ac.in

***Abstract*—Thermal Infrared cameras have a unique set of capabilities that enables them to see and capture images in situations where vision is severely obscured, including dark, shade, and foggy conditions. Conversion of TIR image to visible spectrum makes them more interpretable, but automating transformation is a challenging task as it requires estimation of both chrominance and luminance for each pixel, posing the need for a more complex model that is capable of capturing higher-level semantics of a TIR image to generate a perceptually realistic RGB image.The use of Conditional Generative Adversarial Networks (cGANs) is proposed with a combination of content and adversarial losses to tackle this problem. The proposed cGAN uses a U-Net auto-encoder network as generator, and a patch based discriminator for adversarial training. The cGAN learns the most effective loss function during training, producing state-of-the-art perceptually realistic outputs. Qualitative and quantitative analyses show that our approach outperforms existing methods.**

***Index Terms*—Conditional Generative Adversarial Networks, Deep Learning, Image to Image Transformation, Thermal Infrared Imaging**

I. Introduction

Thermal infrared imaging technology originally found applications in the fields of Military and Intelligence [8]. Thermal infrared cameras capture images and videos in the thermal spectrum, which is invisible to the human eye. Infrared radiation is reflected and emitted by all objects, even in the dark. Thus, thermal infrared cameras can capture images in the dark, and in conditions where visibility is obscured e.g. foggy conditions. In recent years, the technology has significantly improved. Resolution of thermal sensors has increased, and thermal infrared cameras have become less expensive. This has led to the application of thermal infrared cameras in entirely new fields like; automotive computer vision systems, surveillance systems, medical imaging and specialized equipment for various scientific and industrial applications including archaeology, geology, mining etc. Other significant advantages of thermal cameras are that they can capture images even in very low light, they are robust to extreme changes in light and shade conditions, and they preserve the anonymity of the subjects being captured in surveillance systems. As the use of thermal infrared cameras becomes more common, there is a growing need to develop efficient and accurate methods to transform thermal infrared (TIR) images to visible spectrum (RGB) images. Perceptually realistic visible spectrum (RGB) images are more intuitive to interpret and easier to understand for the untrained human eye, and will help improve the viewing performance of the human operator that operates a thermal infrared imaging system.

Transformation of black and white (grayscale) images into colored (RGB) images is a similar problem that has been well researched before [2]–[6]. Autoencoder networks have been used to obtain impressive results in grayscale colorization e.g. the "Colorful Image Colorization" [19] and "Let there be color!" [18]. When colorizing grayscale inputs, a given pixel of the input grayscale image can correspond to many different colors in the colorized output. As each pixel in a grayscale image only captures brightness (luminance), the color (chrominance) of the output remains ambiguous. An input grayscale pixel intensity can map to an infinite number of combinations of red, green and blue intensity values that produce the same luminance as the input but output very different colors. This is known as chrominance ambiguity in colorization methods. The problem of transforming thermal infrared images to visible spectrum images poses an entirely new challenge. A thermal infrared image captures intensities of infrared radiation only. There is no way to estimate visible spectrum intensities knowing only the thermal infrared intensi-

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:49:24 UTC from IEEE Xplore. Restrictions apply.

ties, as they both lie in different radiation spectra. Therefore, it has been required to resolve not only chrominance ambiguity, but also luminance ambiguity for generating a perceptually realistic visual spectrum image from thermal infrared input. Colorization methods have previously been researched that utilise autoencoder Convolutional Neural Networks (CNN), as they are able to extract a semantic representation of input and then generate an appropriate output. This approach was used by Berg et al. (TIR2Lab) [17] for thermal infrared colorization and Limmer et al. [20] for near infrared colorization. However, the images generated using this approach were distorted and blurred Therefore, a more complex system is needed that is able to generate the visible spectrum image from a more abstract representation of the input image.

II. Method

The use of a Conditional Generative Adversarial Network (cGAN) [16] is proposed for generating visible spectrum images from thermal infrared input. This is an image to image transformation problem, a class of problems which has previously been approached using autoencoder networks. Such networks must train on carefully hand engineered loss functions to work effectively. An ill-engineered loss will eventually result in blurry and unrealistic output, which is a common problem in autoencoder models. The proposed Conditional Generative Adversarial Network architecture not only learns the input-output mapping, but also learns the loss function that trains the complete network resulting in clearer and more realistic looking output image.

A. Conditional Generative Adversarial Networks

Conditional Generative Adversarial Networks provide an architecture for effective training of generator networks to learn a conditional mapping from given input image, to a target output image. The generator, a deep convolutional neural network, first extracts the relevant semantic features from the input infrared image, and generates colorized output from that abstract semantic representation. This generator network trains on a combination of adversarial and content losses, in parallel with a discriminator, using the cGAN architecture. Our cGAN was trained on a large dataset (50,184 images) of pairs of thermal infrared and visible spectrum images.

A cGAN consists of two primary components, a generator network (G), and a discriminator network (D). The generator model is an autoencoder network that maps the input image to its corresponding output. With each training iteration, it adjusts it's weights to produce a more realistic image, that is closer to the actual output provided in the dataset. The discriminator model is typically a convolutional neural network that is trained to correctly identify original output images (output images sampled from the dataset) from fake output images (output images generated using the generator model). Unlike in traditional GANs, in cGANs, the discriminator receives both the input image and the target image as input, and it predicts whether the output is a valid transformation of the given input. This aids in learning a conditional mapping between the input and the output. The training process of a GAN is a min-max game, with an objective function as given in eq. 1

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data(x)}}[\log D(x)] + \mathbb{E}_{x\sim p_z(x)}[\log(1 - D(G(x))] \quad (1)$$

Here, G and D are the generator and discriminator network respectively. $D(x)$ represents the discriminator prediction for the given input image (x), $D(G(x))$ represents the discriminator prediction for the generated image ($G(x)$) for the given input image. With the objective of averting detection by generating more realistic outputs, the generator aims to increase the discriminator's classification error, while the discriminator trains to better classify real and fake samples, decreasing the classification error. This min-max game drives the training of a GAN, and adversarial competition trains the generator to produce more genuine looking images.

B. Network Architecture

The generator of the cGAN architecture proposed is an autoencoder network with skip connections that has previously been used for image segmentation and image synthesis (U-Net) [15]. Our discriminator is a convolutional neural network that classifies 70x70 sized patches of the image as real or fake (PatchGAN), producing a tensor as output.

1) Generator Architecture: U-Net consists of a contracting encoder network (left) and an expanding decoder network (right)(see Fig. 1). The encoder network uses convolution operation to downsample the image and obtain an abstract representation of the image at the bottleneck. The decoder network then generates the output image from this feature map by means of upsampling.

2) Encoder: The encoder resembles most Convolutional Neural Networks, with added skip connections to aid in direct transfer of low level information directly from initial layers towards the output layers. The entire encoder section consists of 8 stages. With each encoding stage, two convolutions using a 3x3 filter is applied. The activation function used is Rectified Linear Unit (ReLU). The convolution layer is followed by a 2x2 max pooling layer that has a stride of 2 for downsampling. This downsampling is what lends the encoder network it's contracting path shape. With each stage, the number of channels is doubled at the downsampling step.

3) Decoder: The decoder network consists of 8 decoding stages. Each stage of the decoder applies a simple 2D up sampling operation to the feature map, expanding the map. After that, a 2x2 up convolution is applied, that halves the number of channels in the feature map. A skip connection then concatenates the resulting feature tensor with the output from a corresponding encoding stage. This is followed by application of two 3x3 convolutions, using Rectified Linear

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:49:24 UTC from IEEE Xplore. Restrictions apply.

Unit (ReLU) activation.

4) Discriminator Architecture: The PatchGAN discriminator [9] is another unique element of the proposed architecture (Fig. 2). The function of the discriminator is to correctly classify the given images as being original or fake. Our discriminator classifies patches of the image of NxN size, instead of making predictions for the entire image at once. The primary benefit of using a patch based discriminator is that there are fewer parameters to train in the discriminator network, resulting in faster training. Using a patch based discriminator also encourages sharper detail in generated images [10], [11]. When using a Patch discriminator, the image is essentially modeled as a Markov field, where pixel values are separated by more than a span of N can be considered to be conditionally independent. Through extensive parameter tuning, it has been discovered that a patch size of 70x70 pixels is optimal, making fast discriminator predictions without much loss in image quality. With a patch size of 70x70, there are 900 patches of an output image to predict, resulting in an output tensor of 30x30.

C. Objective Functions

The use of a combination of per-pixel content loss, and the cGAN adversarial loss, are used to train the network. Through the adversarial loss, the high level semantic structure of the input image can be captured, while content loss aids in the capture of low frequency details in the input.

1) Content Loss: This is a per-pixel loss. It evaluates the average difference between each pixel of the target image and the generated image, thus effectively minimizing the per-pixel error in generated images. The L1 content loss is used as it has been observed that use of L2 content loss tends to generate blurry output images.

$$L_{content} = \frac{1}{WH}\sum_{i=1}^{W}\sum_{j=1}^{H}||y_{i,j} - G(x)_{i,j}||_1 \quad (2)$$

where W and H denote the height and width of the image respectively, and $||.||_1$ denotes the L1 norm. $y_{i,j}$ and $G(x)_{i,j}$ represent the pixel values for the sample output image(y) and the generated image ($G(x)$) respectively.

2) Adversarial Loss: Adversarial loss in a cGAN can be expressed as:

$$L_{adversarial}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(G(x), z))] \quad (3)$$

Equation 3 represents a quantitative measure of the classification error of the discriminator. G, the generator network, and D, the discriminator network are involved in a two player min-max game, with the objectives of respectively maximising and minimising this classification error.

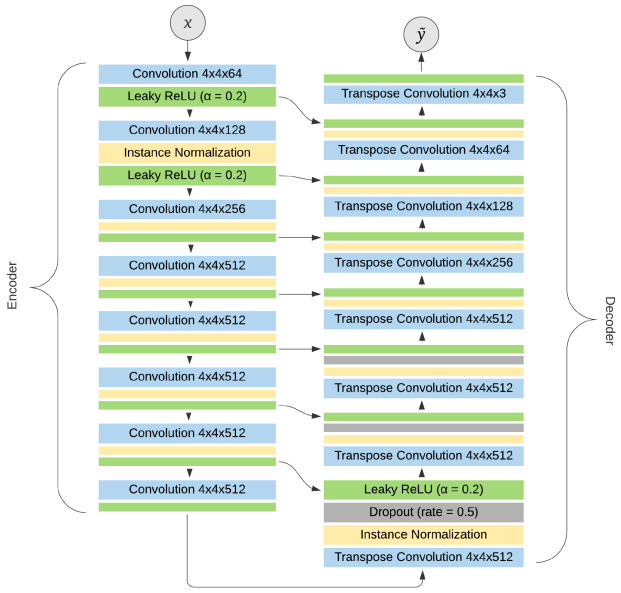


Fig. 1. Generator Architecture

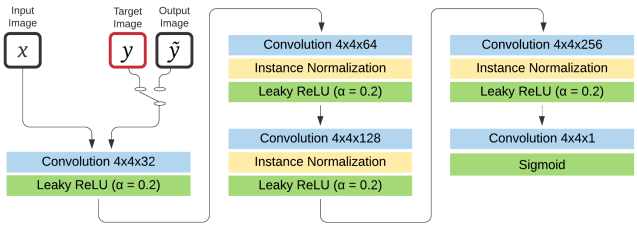


Fig. 2. Discriminator Architecture

3) Final Objective: The final objective function is a weighted sum of the content loss and the adversarial loss.

$$L_{total} = \lambda_{content} L_{content} + \lambda_{adversarial} L_{adversarial} \quad (4)$$

After experimenting and fine tuning parameters on the training data-sets, the optimal values of the weights are obtained as: $\lambda_{content} = 1, \lambda_{adversarial} = 100$

III. Experiments

The model is trained on the publicly available KAIST-MS dataset [7], which is benchmark dataset used for training and testing multi-spectral image transformation models. The dataset consists of 50,184 sample mappings of visible spectrum RGB images and their corresponding TIR images. All images were resized to 256x256 three channel RGB with a train-test split ratio of 80:20. The generator network was initialized with random weights, with a normal distribution $N(0, 0.02)$. The dimensions of the patch discriminator were set to 70x70 after fine-tuning. Training was done with a

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:49:24 UTC from IEEE Xplore. Restrictions apply.

learning rate of 0.0002, $\beta_1 = 0.5$ and $\beta_2 = 0.9$, using the Adam Optimizer. Publicly available Nvidia Tesla P100 (15.9 GB) GPUs were used to train the model on Kaggle.

IV. Results

There is no single widely accepted technique to evaluate the quality of computer generated images, and a variety of multiple methods and measures have been used by researchers working in this field. The quality of our results are evaluated using a number of such quantitative measures in IV-A, and also compared them with baseline results for a subjective qualitative analysis in IV-B.

A. Quantitative Analysis

The model is also compared to its contemporaries on the basis of relevant quantitative metrics. The metrics used will also include the structure similarity index (SSIM) [14], multiscale structure similarity index (MSSIM) [13] and peak signal-to-noise ratio (PSNR). Higher values indicate better quality of generated images. Comparing our method with the Naive and TIR2Lab methods, both trained on the same KAIST dataset, it has been observed that our model outperforms these baselines. Average scores for the compared methods are presented in table I.

1) Structure Similarity Index (SSIM): Structure Similarity Index is a measure used to estimate the perceived quality of cinematic movies and television picture, as well as various types of digital images and videos. The index is calculated on a number of windows of the image. The SSIM for two image windows p and q, having size $N \times N$ is:

$$SSIM(x,y) = \frac{(2\mu_p\mu_q + c_1)(2\sigma_{pq+c_2})}{(\mu_p^2 + \mu_q^2 + c_1)(\sigma_p^2 + \sigma_q^2 + c_2)} \quad (5)$$

Where μ_p is mean, σ_q is standard devation, and σ_{pq} is co-variance.

2) Peak Signal-to-Noise Ratio (PSNR): PSNR is a mathematical ratio between the maximum power of a given signal and the power of noise error. It is commonly used to measure the image reconstruction quality of lossy compression methods. PSNR is used to compare the quality of the generated RGB image with the actual visual spectrum image sample. A simple PSNR can be defined by mean squared error.

$$\frac{1}{WH}\sum_{x=1}^{W-1}\sum_{y=1}^{H-1}(P(x,y) - Q(x,y))^2 \quad (6)$$

Where the images P and Q are $W \text{x} H$ sized.

3) Multi-scale SSIM (MSSIM): Multi-scale SSIM calculates the structural similarity index by calculating the SSIM index for several differently scaled versions of an image, and then combining them. The MS-SSIM index is more robust than SSIM index.

TABLE I
AVERAGE RESULTS OF ALL THE TEST IMAGES OF THE KAIST DATA-SET.

Method	PSNR	SSIM	MSSIM
Naive	9.15	0.32	0.13
TIR2Lab	14.6	0.52	0.51
Proposed	**16.12**	**0.55**	**0.53**

B. Qualitative Analysis

Fig. 3 shows the results generated using our proposed model. Comparing them with the TIR2Lab baseline in Fig. 4, It can also be seen that the obtained results are highly detailed, and perceptually realistic. Our model is robust and correctly generates outputs for both day time, and night time inputs. This was ensured by training the cGAN on day and night images separately.

V. Conclusion

In this paper, a deep learning approach is proposed to solve the colorization problem of thermal infrared images. In particular, a multi-term objective function associated with conditional generative adversarial networks is used to learn the mapping for this image to image translation task. Our method has been evaluated on a publicly available dataset. Our qualitative results produce perceptually realistic images, while retaining the high level information of the given input. When tested on quantitative measures, our model gave a comparable SSIM score, as well as a better PSNR value. Both qualitative and quantitative analysis methods demonstrate that our proposed work outperforms all previous traditional approaches such as the use of a Convolutional Neural Network as an auto-encoder network.

References

[1] Ali, E. A., Qadir, H., Kozaitis, S. P. (2014, June). Color night vision system for ground vehicle navigation. In Infrared Technology and Applications XL (Vol. 9070, p. 90700I). International Society for Optics and Photonics.

[2] Cao, Y., Zhou, Z., Zhang, W., Yu, Y. (2017, September). Unsupervised diverse colorization via generative adversarial networks. In Joint European conference on machine learning and knowledge discovery in databases (pp. 151-166). Springer, Cham.

[3] Cheng, Z., Yang, Q., Sheng, B. (2015). Deep colorization. In Proceedings of the IEEE International Conference on Computer Vision (pp. 415-423).

[4] Deshpande, A., Lu, J., Yeh, M. C., Jin Chong, M., Forsyth, D. (2017). Learning diverse image colorization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 6837-6845).

[5] Gupta, R. K., Chia, A. Y. S., Rajan, D., Ng, E. S., Zhiyong, H. (2012, October). Image colorization using similar images. In Proceedings of the 20th ACM international conference on Multimedia (pp. 369-378).

[6] Huang, Y. C., Tung, Y. S., Chen, J. C., Wang, S. W., Wu, J. L. (2005, November). An adaptive edge detection based colorization algorithm and its applications. In Proceedings of the 13th annual ACM international conference on Multimedia (pp. 351-354).

[7] Hwang, S., Park, J., Kim, N., Choi, Y., So Kweon, I. (2015). Multispectral pedestrian detection: Benchmark dataset and baseline. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1037-1045).

[8] Gade, R., Moeslund, T. B. (2014). Thermal cameras and applications: a survey. Machine vision and applications, 25(1), 245-262.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:49:24 UTC from IEEE Xplore. Restrictions apply.

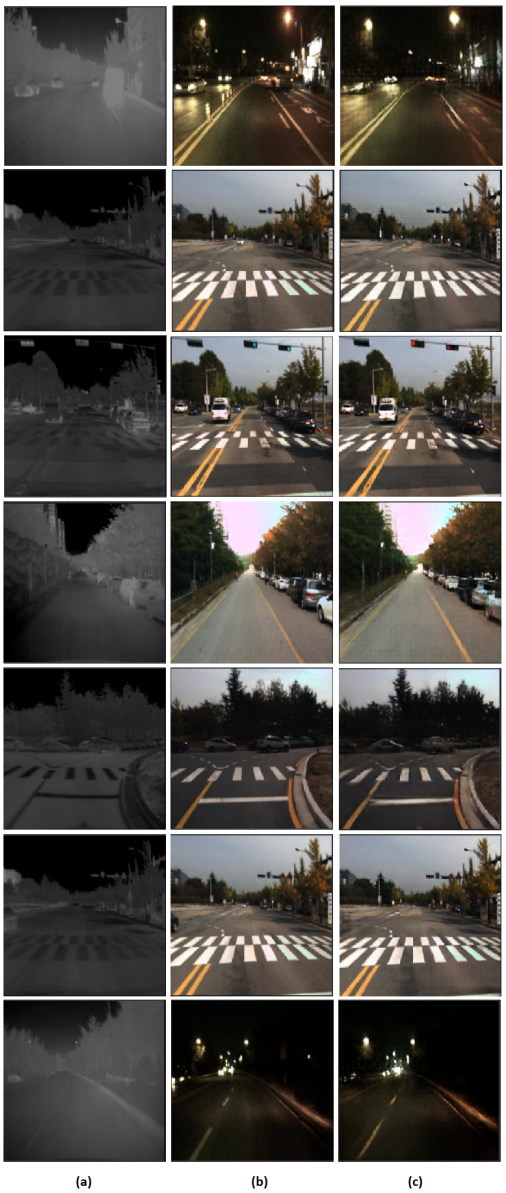


Fig. 3. Our generated results (a)Thermal Infrared Images, (b) Ground Truth(True RGB images), (c)Proposed



Fig. 4. Our results compared with TIR2Lab (a)Thermal Infrared Images, (b) Ground Truth(True RGB images), (c)TIR2Lab, (d)Proposed

[9] Li, C., Wand, M. (2016, October). Precomputed real-time texture synthesis with markovian generative adversarial networks. In European conference on computer vision (pp. 702-716). Springer, Cham.
[10] Gatys, L. A., Ecker, A. S., Bethge, M. (2016). Image style transfer using convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2414-2423).
[11] Li, C., Wand, M. (2016). Combining markov random fields and convolutional neural networks for image synthesis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2479-2486).
[12] Damera-Venkata, N., Kite, T. D., Geisler, W. S., Evans, B. L., Bovik, A. C. (2000). Image quality assessment based on a degradation model. IEEE transactions on image processing, 9(4), 636-650.
[13] Wang, Z., Simoncelli, E. P., Bovik, A. C. (2003, November). Multiscale structural similarity for image quality assessment. In The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003 (Vol. 2, pp. 1398-1402). Ieee.
[14] Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, 13(4), 600-612.
[15] Ronneberger, O., Fischer, P., Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.
[16] Mirza, M., Osindero, S. (2014). Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.
[17] Berg, A., Ahlberg, J., Felsberg, M. (2018). Generating visible spectrum images from thermal infrared. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 1143-1152).
[18] Iizuka, S., Simo-Serra, E., Ishikawa, H. (2016). Let there be color! Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. ACM Transactions on Graphics (ToG), 35(4), 1-11.
[19] Zhang, R., Isola, P., Efros, A. A. (2016, October). Colorful image colorization. In European conference on computer vision (pp. 649-666). Springer, Cham.
[20] Limmer, M., Lensch, H. P. (2016, December). Infrared colorization using deep convolutional neural networks. In 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 61-68). IEEE.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:49:24 UTC from IEEE Xplore. Restrictions apply.

GloVeInit at SemEval-2020 Task 1: Using GloVe Vector Initialization for Unsupervised Lexical Semantic Change Detection

Vaibhav Jain
Delhi Technological University
Delhi, India
vaibhav29498@gmail.com

arXiv:2007.05618v1 [cs.CL] 10 Jul 2020

Abstract

This paper presents a vector initialization approach for the SemEval2020 Task 1: Unsupervised Lexical Semantic Change Detection. Given two corpora belonging to different time periods and a set of target words, this task requires us to classify whether a word gained or lost a sense over time (subtask 1) and to rank them on the basis of the changes in their word senses (subtask 2). The proposed approach is based on using Vector Initialization method to align GloVe embeddings. The idea is to consecutively train GloVe embeddings for both corpora, while using the first model to initialize the second one. This paper is based on the hypothesis that GloVe embeddings are more suited for the Vector Initialization method than SGNS embeddings. It presents an intuitive reasoning behind this hypothesis, and also talks about the impact of various factors and hyperparameters on the performance of the proposed approach. Our model ranks 13th and 10th among 33 teams in the two subtasks. The implementation has been shared publicly.[1]

1 Introduction and Background

Lexical Semantic Change (LSC) Detection is an active research topic in the field of natural language processing, and has been applied for diachronic (across time) and synchronic (across domains) tasks (Schlechtweg et al., 2019). Previously limited to manual "close-reading" approaches, the availability of large-scale corpora have allowed the use of computational methods for this task (Tahmasebi et al., 2018). This topic has found various applications in various disciplines such as improving information retrieval from historical documents (Morsy and Karypis, 2016), preventing cross-domain ambiguity in requirements elicitation interviews (Jain et al., 2020), and studying the impact of societal and cultural changes on word meanings and usage (Tahmasebi and Risse, 2017).

LSC detection involves the use of two corpora C_1 and C_2 which, in the diachronic case, belong to different time periods t_1 and t_2 respectively. The various approaches found in the literature usually involve the construction of a word embedding space specific to each corpus. The embeddings can be constructed through count-based methods such as Positive Pointwise Mutual Information and Random Indexing, or predictive methods such as Skip-Gram with Negative Sampling (SGNS). Most of these models are stochastic in nature which means that separately trained embedding models live in their own space. In order to project them onto a *unified space*, alignment techniques such as vector initialization and orthogonal Procrustes are used. The LSC of a word is then quantitatively determined by measuring the contextual dissimilarity between the word's representations (Tahmasebi et al., 2018).

Recently, there have been efforts to evaluate these various methods by comparing their results with manually-annotated data (Schlechtweg et al., 2019; Ahmad et al., 2020). The SemEval-2020 Task 1 is one such effort which is based on LSC detection in corpora of English, German, Latin, and Swedish languages (Schlechtweg et al., 2020). Its aim is to provide an evaluation framework for unsupervised LSC detection systems by comparing their results against a ground truth, as annotated by native speakers or scholars. It consists of two subtasks: *Given two corpora C_1 and C_2 (for time periods t_1 and t_2), for a set of target words,*

[1] github.com/vaibhav29498/GloVeInit

This work is licensed under a Creative Commons Attribution 4.0 International License. License details: http://creativecommons.org/licenses/by/4.0/.

1. *decide which words lost or gained senses between t_1 and t_2, and which ones did not; as annotated by human judges.*

2. *rank them according to their degree of LSC between t_1 and t_2 as annotated by human judges. A higher rank means stronger change.*

2 System Overview

2.1 Vector Initialization

The Vector Initialization (VI) alignment method, which was first used by Kim et al. (2014), involves the embedding space for t_1 to be trained independently on C_1. It is then used to initialize the embedding space for t_2 which is then subsequently trained from C_2. The underlying idea is that the embedding for a word w will get considerably updated if it is used within different contexts in C_1 and C_2, otherwise it will receive only a slight update. A study by Schlechtweg et al. (2019), which applied the VI method on SGNS embeddings, found it to perform significantly weaker on LSC detection tasks than other methods such as orthogonal Procrustes (OP). They attributed this to the sensitivity of the VI method to the frequency of a word in C_2. The high frequency of a word in C_2 will result into the word's embedding getting frequent updates away from its initial state. This leads to a large divergence between the word's representations even if it does not undergo any LSC from t_1 to t_2.

In a recent shared task on LSC detection in the German language, a modified VI approach achieved the third-best result and outperformed certain OP approaches (Ahmad et al., 2020). Instead of only initializing on the word embeddings obtained from C_1, the modified approach initializes on the complete SGNS model which includes the hidden layer. However, it also theoretically suffers from sensitivity to high frequency.

2.2 Comparison of GloVe and SGNS

GloVe (Global Vectors) and SGNS are unsupervised algorithms for obtaining word embedding spaces (Pennington et al., 2014; Mikolov et al., 2013). GloVe makes use of co-occurrence matrix X; its (i, j) entry, X_{ij} is the number of times the word w_j appears in the context of the word w_i (as defined by the window-size L). It trains the word embeddings by minimizing the cost function

$$J = \sum_{i=1}^{V} \sum_{j=1; j \neq i}^{V} f(X_{ij})(u_j^T v_i - \log X_{ij})^2, \tag{1}$$

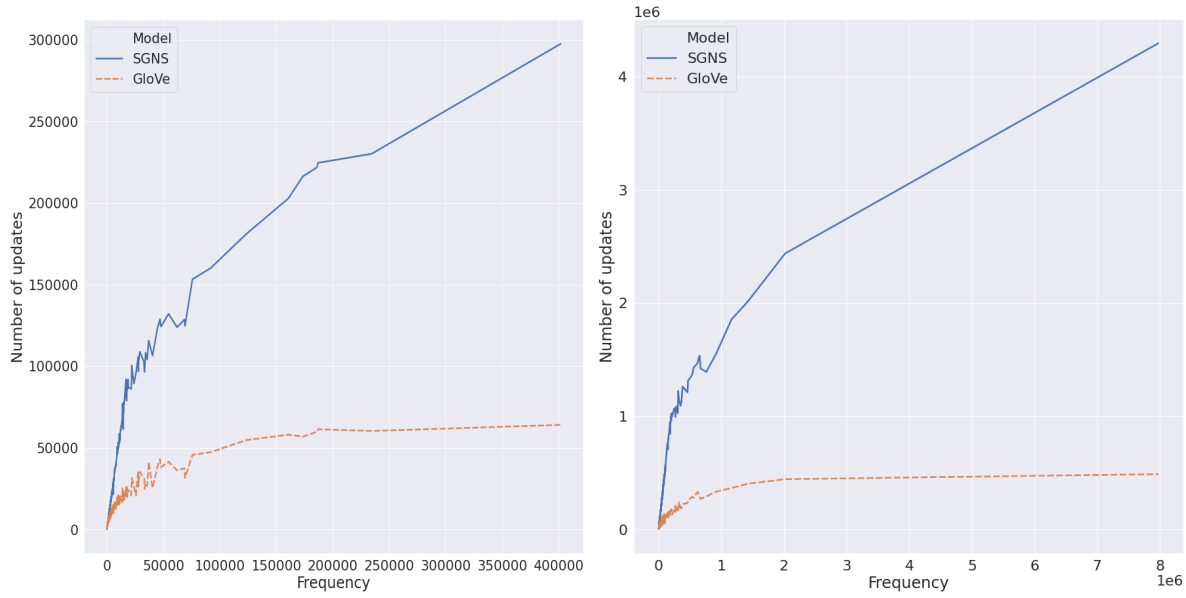
where V is the vocabulary size and $u, v \in \mathbb{R}^D$ are the word and context word vectors respectively. The final word embeddings can be obtained by summing or averaging the two.

On the other hand, SGNS is a predictive model which relies on a shallow two-layer neural network which, given a word, predicts the set of its context words. To avoid the expensive softmax function in the training objective, negative sampling is used by drawing a few negative samples from a noise distribution.

An important distinction between GloVe and SGNS from this paper's point of view is the number of updates that are made to a word embedding during training. In a single epoch, the number of updates to the SGNS embedding of a word is roughly equal to the number of words that appear in its context throughout the corpus. Hence, the number of updates is proportional to the frequency of the word which has an upper-bound of the total number of words in the corpus. However, this relationship is not exactly linear because of downsampling of frequent words and negative sampling. In the GloVe model, the number of updates received in a single epoch is equal to the number of *distinct* context words. This number is limited by an upper-bound of vocabulary size V, which is usually much less than the total number of words in the corpus.

As discussed in Section 2.1, the VI method can falsely give a high LSC score to words with high frequency. Our hypothesis is that the GloVe model is more suited for the VI method due to its lesser sensitivity to high frequency. To give an indication of the extent of this difference, the relationship between a word's frequency and the number of updates to its embedding is depicted in Figure 1 for both GloVe and

SGNS models trained on the C_2 corpora of all the four languages. These results are based on the official GloVe implementation[2] and the `gensim`[3] implementation of SGNS. The models were trained for a single epoch and only words with a frequency of greater than or equal to 5 were considered. It is clear that the proportion of updates for frequent and rare words is comparatively more balanced in the case of GloVe. On the other hand, there is a high rate of growth in the number of updates with respect to the frequency in the case of SGNS, which indicates that it is biased towards estimating a high LSC for frequent words.



(a) CCOHA 1960-2010 (English)

(b) BZ and ND 1946-1990 (German)

(c) LatinISE 0-2000 (Latin)

(d) KubHist 1895-1903 (Swedish)

Figure 1: Plot of frequency of a word (x-axis) vs the number of updates to its embedding (y-axis). The subcaptions include the name of the corpora C_2, the corresponding time-period t_2, and its language.

[2] `github.com/stanfordnlp/GloVe`
[3] `radimrehurek.com/gensim`

3 Results and Experimental Setup

The dataset provided by the competition organizers consists of the corpora C_1 and C_2, and the list of target words for four languages: English, German, Latin, and Swedish (Schlechtweg et al., 2020). All of the corpora were already prepossessed: they were in tokenized and lemmatized form with punctuation marks and one-word sentences removed.

Subtask 1 is concerned with identifying the loss or gain of one or more word sense(s), where subtask 2 tests a model's ability to detect fine-grained changes in the two sense frequency distributions. For example, consider the word *cell* which historically referred to either a chamber or the smallest unit of an organism. However, the relative usages of both these senses have decreased with time, with mobile phone being the predominant meaning now. For subtask 2, the ground-truth ranking of the target words is determined by calculating the Jensen-Shannon divergence between their normalized sense frequency distributions from t_1 and t_2 (Donoso and Sánchez, 2017). A submission is scored by its Spearman's rank-order correlation coefficient against the ground-truth ranking.

We used cosine distance as the metric to calculate the distance between a word's vectors in diachronic spaces, and every word with a distance of more than 0.55 (for German) or 0.45 (for other languages) were classified to have gained of lost word sense(s).

In all of our experiments, only words having a frequency of at least 5 were considered. Our best-performing solution during the evaluation phase achieved an accuracy of 60% in subtask 1 and a score of 0.352. It was obtained by training word embeddings with a dimensionality of 50 on both C_1 and C_2 for 60 epochs each with a window-size of 10. The subtask-2 score was close to but less than the modified SGNS-based VI approach discussed in Section 2.1, which was proposed by the team *IMS* and received a score of 0.372. We ranked 13th and 10th out of 33 participants in the two subtasks respectively.

Further experiments on subtask-2 were conducted after the competition for analyzing the impact of hyperparameters like window-size L and embedding dimensionality d and are reported in Table 1. Number of training epochs was limited to 20 for models with window-size 10 because of their high computational requirements. The results suggest that training the models for higher number of epochs can produce better results if the embedding dimensionality is high, but can backfire in the opposite case. Using a larger window-size improves the average result in most cases.

L	**Epochs**	***d***	**Scores**				
			English	**German**	**Latin**	**Swedish**	**Average**
5	20	5	0.012	0.366	0.391	0.15	0.23
		10	-0.144	0.362	0.394	0.114	0.182
		20	0.193	0.402	0.368	0.244	0.302
		50	**0.278**	0.446	0.315	0.229	0.317
		100	0.226	0.312	0.254	0.186	0.244
	60	5	-0.018	0.354	0.432	0.121	0.222
		10	-0.095	0.283	0.379	0.108	0.169
		20	0.212	0.41	0.377	0.324	0.331
		50	0.228	0.464	0.329	**0.366**	**0.347**
		100	0.269	0.312	0.276	0.309	0.291
10	20	5	0.011	0.334	**0.485**	0.218	0.262
		10	-0.036	0.351	0.409	0.146	0.218
		20	0.144	**0.479**	0.397	0.187	0.302
		50	0.198	0.477	0.275	0.267	0.304
		100	0.199	0.349	0.257	0.227	0.258

Table 1: Results of the post-evaluation experiments

3.1 Error Analysis

We defined the ranking error for each target word as the difference between its predicted rank and its true rank divided by the total number of target words. It lies in the range $(-1, 1)$ and has an ideal value of zero. A positive value indicates that the model overestimated the LSC for a word, and a negative value indicates otherwise. A regression plot between the relative frequency of the target words and their ranking error as per our best performing model is depicted in Figure 2. A word's relative frequency is defined as its frequency divided by the total frequency of all words in the corpus.

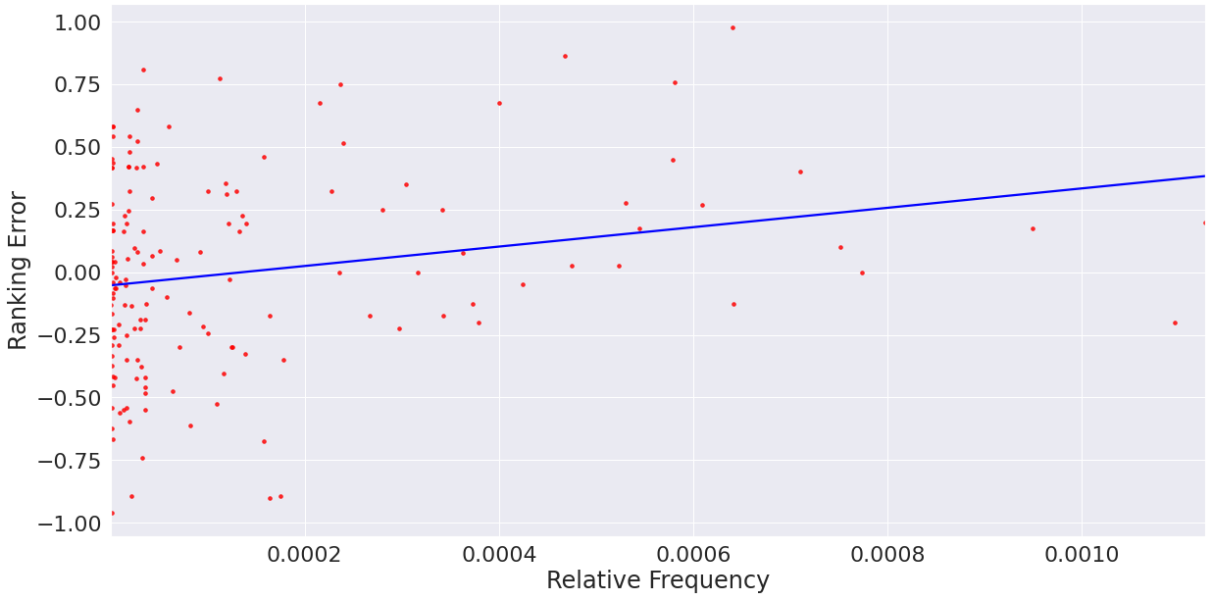


Figure 2: Regression plot between the frequency of the target words and their ranking error

There is a statistically significant correlation between the ranking error and the relative frequency, with the result of the Spearman's rank-order correlation test being $\rho = 0.186$ and $p = 0.020$. This indicates that the GloVe model's relative insensitivity to high frequency can lead to it assigning a rank lower than the true rank to such words. We believe that the frequency of the target words is not large enough for SGNS' sensitivity to high frequency to have a major impact on the results. This explains our model's inferior performance as compared to SGNS-based model in this task.

4 Conclusion and Future Work

In this paper, we reported our work in the SemEval2020 Task 1. We proposed a GloVe-based VI approach which achieved the 11th rank out of 32 participating teams. We gave a theoretical reasoning behind why GloVe is less sensitive towards high frequency than SGNS and thus more suited for VI method, and empirically showed the magnitude of the difference between the number of updates to word embeddings in the two models. We believe that despite a lower-than-expected performance in this competition, our work presents a good case for the suitability of GloVe model when corpora of larger size are involved. However, proving this quantitatively is a challenging task because of the limitations associated with manual annotation-based evaluation.

Planned future work includes the study of how techniques such as dimension-wise mean-centering and length normalization, which have proved beneficial in OP-based approaches, can be applied for VI method.

Acknowledgments

We would like to thank the task organizers for preparing the dataset and organizing the competition, Jens Kaiser (member of the team *IMS*) from University of Stuttgart for sharing the results from his work on the SGNS-based VI alignment method, and the anonymous reviewers for their detailed and helpful feedback.

References

Adnan Ahmad, Kiflom Desta, Fabian Lang, and Dominik Schlechtweg. 2020. Shared task: Lexical semantic change detection in german. *CoRR*, abs/2001.07786.

Gonzalo Donoso and David Sánchez. 2017. Dialectometric analysis of language variation in twitter. In *Proceedings of the Fourth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial)*, pages 16–25, Valencia, Spain, April. Association for Computational Linguistics.

Vaibhav Jain, Ruchika Malhotra, Sanskar Jain, and Nishant Tanwar. 2020. Cross-domain ambiguity detection using linear transformation of word embedding spaces. In *Joint Proceedings of REFSQ-2020 Workshops, Doctoral Symposium, Live Studies Track, and Poster Track co-located with the 26th International Conference on Requirements Engineering: Foundation for Software Quality (REFSQ 2020)*, volume 2584 of *CEUR Workshop Proceedings*, Pisa, Italy. CEUR-WS.org.

Yoon Kim, Yi-I Chiu, Kentaro Hanaki, Darshan Hegde, and Slav Petrov. 2014. Temporal analysis of language through neural language models. In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, pages 61–65, Baltimore, MD, USA, June. Association for Computational Linguistics.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS13, page 31113119, Red Hook, NY, USA. Curran Associates Inc.

Sara Morsy and George Karypis. 2016. Accounting for language changes over time in document similarity search. *ACM Trans. Inf. Syst.*, 35(1), September.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar, October. Association for Computational Linguistics.

Dominik Schlechtweg, Anna Hätty, Marco Del Tredici, and Sabine Schulte im Walde. 2019. A wind of change: Detecting and evaluating lexical semantic change across times and domains. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 732–746, Florence, Italy, July. Association for Computational Linguistics.

Dominik Schlechtweg, Barbara McGillivray, Simon Hengchen, Haim Dubossarsky, and Nina Tahmasebi. 2020. SemEval-2020 Task 1: Unsupervised Lexical Semantic Change Detection. In *To appear in Proceedings of the 14th International Workshop on Semantic Evaluation*, Barcelona, Spain. Association for Computational Linguistics.

Nina Tahmasebi and Thomas Risse. 2017. On the uses of word sense change for research in the digital humanities. In *Research and Advanced Technology for Digital Libraries*, pages 246–257, Cham. Springer International Publishing.

Nina Tahmasebi, Lars Borin, and Adam Jatowt. 2018. Survey of computational approaches to lexical semantic change. *CoRR*, abs/1811.06278.

The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/1934-8835.htm

Gratitude, social intelligence, and leadership among university teachers: mediation and moderation analysis

Naval Garg
University School of Management and Entrepreneurship, Delhi Technological University, Delhi, India

Anuradha Jain
Vivekanand School of Economics, Vivekananda Institute of Professional Studies, Pitampura, India, and

B.K. Punia
Haryana School of Business, Guru Jambheshwar University of Science and Technology, Hisar, India

Received 18 January 2020
Revised 8 April 2020
20 May 2020
Accepted 24 May 2020

Abstract

Purpose – The purpose of this study is to investigate the linkage between gratitude and leadership among university teachers. It also tended to examine mediating and moderating effect of social intelligence amidst the relationship between gratitude and leadership. As social intelligence is a multi-dimensional concept, this paper explored mediating and moderating effect of three dimensions of social intelligence, i.e. social awareness, social skills and social information processing.

Design/methodology/approach – Mediation effect was examined with help of Baron and Kenny's (1986) technique, Sobel test and bootstrapping estimates. Significance of interaction between independent variable (gratitude) and moderator (social intelligence) was investigated for moderation analysis.

Findings – The findings reported partial mediating effect of all three dimensions of social intelligence. No significant moderation effect was observed.

Originality/value – This is one of the pioneer studies that explored interrelationship between gratitude, social intelligence and leadership among Indian University teachers.

Keywords Gratitude, Social intelligence, Leadership, Mediation, Moderation

Paper type Research paper

Introduction

A growing number of researchers in positive psychology have proved that gratitude improves well-being (Emmons and McCullough, 2003; Watkins *et al.*, 2014; Wood *et al.*, 2010), facilitates life satisfaction (McCullough *et al.*, 2002; Garg, 2017a), elicits prosocial behavior (Bartlett and DeSteno, 2006) and reduces aggression, envy and narcissism (Froh *et al.*, 2011). However, researchers still believe that the domain of gratitude is yet to be explored adequately. Emmons and Crumpler (2000) highlighted the scarcity of studies on gratefulness in the sphere of social science and organizational psychology. Only handfuls of researchers have investigated the role of gratefulness in modern organizations (Waters, 2012; Grant and Wrzesniewski, 2010).

International Journal of Organizational Analysis
© Emerald Publishing Limited
1934-8835
DOI 10.1108/IJOA-01-2020-2016

According to Emmons and McCullough (2003), gratitude directly influences organizations through improving organizational climate. It also enhances employee's well-being and reduces negative emotions like resentment and envy at the workplace. Also, Fehr *et al.* (2017) proposed that gratitude may play an essential role in various prominent organizational phenomena such as leadership, organizational citizenship and customer service. The feeling of thankfulness may also promote employee efficiency, success, productivity and loyalty (Grant and Wrzesniewski, 2010; Garg, 2018). Previous studies support the positive association of gratitude with organizational citizenship behaviors (Dik *et al.*, 2014; Spence *et al.*, 2013; Garg *et al.*, 2019), prosocial work behaviors (Grant and Gino, 2010), altruism (Spence *et al.*, 2013) and social support at the workplace (Kaplan *et al.*, 2014). In a critical study, Edmondson (2002) reported that gratitude facilitates psychological safety at the workplace, which promotes trust, faith and courage to take interpersonal risks. It influences the attitude of employees toward their colleagues, superiors and management, which enhances their job satisfaction by facilitating better interpersonal relationships (Hu and Kaplan, 2014; Giles and Yates, 2014). Gratuitous organizational culture ensures the recognition of employees' contributions to the workplace. Authors also reported that grateful employees approach their colleagues in a more positive note, which leads to more magnificent team spirit and altruistic behavior.

Although these studies have highlighted the potential benefits of gratitude for modern organizations, they are often criticized for three main reasons. Firstly, most of the studies are theoretical; thereby, the domain calls for empirical investigation of these hypothesized relationships. Secondly, a theoretical explanation of gratitude (see broaden-and-build theory of positive emotions, social-cognitive theory of gratitude and find, remind and bind theory of gratitude) is primarily based on western context and researchers advocate the application of these theories in eastern countries for better acceptability and research rigor. Then finally, research studies are yet to explore the direct impact of gratitude on individual and organizational outcomes such as motivation, morale, intelligence, customer satisfaction and profitability (Fredrickson, 2004; Fehr *et al.*, 2017). Fredrickson (2004) highlighted the need to investigate the role of gratitude in organizations. "In organizations, does gratitude lead to lower employee turnover, more customer loyalty, higher net sales or more profitable outcomes?" (Fredrickson, 2004, p. 160). Unfortunately, the research gap highlighted by Fredrickson is yet to be fully answered. Other researchers, including Grant and Wrzesniewski (2010), Watkins *et al.* (2003), Garg (2020) and Garg and Gera (2019), have also called for more empirical researches on gratitude, especially in the modern organization.

While explaining the broaden-and-build theory of positive emotions in the context of gratefulness, Fredrickson (2004) suggested an investigation of the mechanism through which gratitude strengthens organizations. Probably, the author was aspiring for future researchers to examine moderating and mediating variables amidst the relationship between gratitude and proposed organizational outcome variables. Algoe and Stanton (2012) find, remind and bind the theory of gratitude and Watkins's *et al.* (2014) amplification theory of gratitude also reiterated the same research gap. Based on these theories, this study proposes a model that tends to examine mediating and moderating social of social intelligence in the relationship between gratitude and leadership. The mediating variable explains the mechanism through which the effect of independent variables transmits to the dependent variable. In other words, the independent variable influences the mediating variable, which ultimately influences the dependent variable. Then, the moderating variable influences the relationship between predictor and criterion variable. The moderators act as a catalyst that either increases or decreases the intensity of association between two variables. Following a section of this paper reviews previous literature on gratitude, social intelligence

and leadership. Next, several hypotheses regarding the relationship between gratitude, three dimensions of social intelligence (social awareness (SA), social skills (SS) and social information processing), and leadership are proposed in Section 2. Propositions of mediating and moderating effects of three dimensions of social intelligence are also introduced here. Then, the methodology, sample, questionnaires and findings of the study are discussed in Section 3. This paper ends with a discussion on the theoretical and practical implications of the study in Section 4. The limitations of the study, along with scope for future researchers, were also discussed in Section 5 of the paper.

Theory and conceptual framework

Gratitude

Gratitude has been conceptualized in several forms, including emotion, trait, virtue, habit and coping mechanism (Emmons and McCullough, 2003). "In various contexts, gratitude can refer to an emotional state, an emotional expression, a character trait or even a virtue" (Watkins *et al.*, 2014, p. 13). Researchers suggested that Rosenberg's (1998) "level of analysis" is helpful for an adequate understanding of various facets of gratitude (Watkins *et al.*, 2014). Rosenberg (1998) argued that an affective phenomenon could be divided into three levels of analysis, namely, emotion, mood and trait. Similarly, gratitude could also be approached using this typology, which assists researchers to conceptualize gratefulness appropriately.

As emotions, gratitude may be experienced when people feel "that something good has happened to them, and they recognize that someone else is largely responsible for this benefit" (Watkins *et al.*, 2014, p. 17). Moral affect theory of gratitude (McCullough *et al.*, 2001) reported that gratitude is an honest emotion in three ways. Firstly, it acts as a moral barometer. A barometer is an instrument that indicates a change in atmospheric pressure. In the same way, gratitude represents the emotional response of beneficiary toward benefactor's intentional act of benevolence. Secondly, gratitude acts as a moral motivator. A grateful individual is more likely to act morally and pro-socially. Previous researches (Peterson and Stewart, 1996; Graham, 1998) specially elaborated that a beneficiary possesses a higher probability of contributing to the welfare of the benefactor or even to the third party. Then finally, it can reinforce moral behaviors. A benefactor who receives sincere gratitude is more likely to help their beneficiary again, and thereby, gratuitous behaviors are reinforced (Watkins *et al.*, 2014).

As an affective trait, it is defined as a "general tendency to recognize and respond with grateful emotions to the roles of other people's benevolence in the positive experiences and outcomes that one obtains" (McCullough *et al.*, 2002, p. 112). This approach considers gratitude as a personality trait of an individual. "Trait gratitude may be viewed more as the being of gratitude, whereas gratitude emotion is related to the feelings of gratitude" (Watkins *et al.*, 2014, p. 22). Watkins (2009) defined three essential components of trait gratitude and termed them as "three pillars of gratitude." Firstly, grateful people are poised to have a strong *sense of abundance* or lack of *sense of deprivation*. A sense of abundance or lack of sense of deprivation allows an individual to acknowledge and appreciate all physical and material things that a person possess. Such persons start appreciating previously unnoticed blessings such as a human body, sound mind, three square meal, family, friend, job and home. Although these items seem quite reasonable and insignificant, a mere realization that a large number of people still dreams of having a life similar to them does the miracle. This realization helps a grateful person to acknowledge the fact that he is far more blessed than most of the others. Appreciation of simple pleasure is expressing thankfulness for mundane aspects of life such as a pleasant meal, excellent weather, sudden and

unplanned meeting with an old friend. Secondly, grateful people have a strong sense of *appreciation for the simple pleasures* of life. Even mundane, routine, and day-to-day benefits are cheered by people possessing trait gratitude. Appreciation of simple pleasure is expressing thankfulness for all aspects of life such as a pleasant meal, excellent weather, sudden and unplanned meeting with an old friend, etc. Then finally, grateful people are also characterized by *social appreciation* or *appreciation of others*: they appreciate the contribution of others in their life. This component of trait gratitude also values the importance of expression of appreciation to others.

In addition to these pillars of gratitude, McCullough *et al.* (2002) elaborated four facets of trait gratitude, namely, *intensity*, *frequency*, *span* and *density*. Frequency is the number of times an individual experiences gratitude in a given set of time. Span refers to several events for which one feels gratefulness in a given period. Density involves the number of persons for which an individual experiences gratefulness regarding one particular positive outcome. Then finally, intensity refers to a degree of experience of gratefulness. To summarize, people high in trait gratitude should show facets of intensity, span, frequency and density of gratitude. Little work has investigated gratitude as mood probably because emotion and mood states are not appreciatory distinguished in previous researches (Watkins *et al.*, 2014). However, Rosenberg's approach clarified that moods are longer in duration and are more rooted in awareness than emotions. Grateful moods are not about something particular. Then they are assumed to facilitate grateful emotions. In a nutshell, researchers hailed gratitude as a multidimensional concept that can take several forms (Alkozei *et al.*, 2017). Although researchers have discussed several perspectives of gratitude, they broadly agree on four facets of gratefulness. These facets are, firstly, benefactor favors beneficiary intentionally without being bound of duty; secondly, the gift has certain utility for the receiver; thirdly, the recipient recognizes the goodness of favor/gift and finally, the receiver aspires to reciprocate the gift received from the giver.

Social intelligence

Social intelligence is defined as "the capability to understand and manage humanity to act wisely in human relations (Thorndike, 1920, p. 228). According to Marlowe (1986), social intelligence is one's capability to understand the feelings, thoughts and behaviors of self and others in social circumstances and consequently act according to that understanding. It is accessed with the help of the Tromso Social intelligence scale (TSIS) that investigates three dimensions of social intelligence, namely, SA, SS and Social Information Processing (SIP). SA refers to one's ability to sense social cues about another person's thoughts, aspirations, and emotions. It helps an individual to understand the social situation appropriately (Goleman, 2006). Goleman (2006) empathized that the empathic acknowledgment of other's feelings and roles is the basis of SA. Further, SS are competencies that promote interaction and communication with others. They are cumulative inherent and learned behaviors that influence one's relationship with others (Agran *et al.*, 2016; Davies *et al.*, 2015). Examples of such behaviors are praising and appreciating others, cooperation, helping and seeking help from others, collaborating with others. Then finally, SIP describes the process of cognitive and behavioral response generation is affected by the SIP mechanism. It refers to the ability to understand social cues and other's feelings and aspirations. It helps an individual to predict one's beliefs, expectations, wishes and behavior through observing their facial expression, gestures, eye movement and body language. In a nutshell, it sensitizes employees for their and other feelings.

Leadership

Northouse (2007) defines leadership as "a process whereby an individual influences a group of individuals to achieve a common goal" (p.75). Also, Greer ad Plunkett (2000) stated that "leadership is an art that can be acquired and developed by anyone with the motivation to do so." (p. 157). Researchers have propagated various theories to explain how people become leaders. One of the most eminent scholars, Bass (1990) contributed greatly in the development of three leadership theories. Firstly, trait theory acknowledged the contribution of certain desirable personality traits in evolution of a leader. Secondly, situational theory gave credit to a situation or event that provided an opportunity to rise above the occasion. Such situations are so stimulating and challenging that they elicit extraordinary leadership skills from an ordinary-looking man. Then finally and most importantly, transformational leadership theory emphasized on four essential dimensions of leadership i.e. individualized consideration, intellectual stimulation, inspirational motivation and idealized influence. The theory stated that "transformational leaders are people with a vision who demonstrate the ability to accomplish greatness, are capable of obtaining results through the injection of energy and enthusiasm, and are generally trusted, respected and admired by their followers" (pp. 23–25). The underlying premise of the present study is based on the transformational theory of leadership. Here, gratitude is proposed to have influence over four inherent dimensions of transformational leadership. Garg (2020) also supported the association between gratitude and different constructs of leadership. The theoretical premise of the interrelationship between gratitude, social intelligence and leadership are explained in the following paragraphs.

Gratitude, social intelligence and leadership

Proposed associations between gratitude, social intelligence and leadership are theoretically based on Fredrickson's (1998) *broaden-and-build theory of positive emotions*, and Algoe and Stanton (2012) *find, remind and bind argument of gratitude*. According to Fredrickson, positive emotions like gratitude "broaden people's momentary thought-action repertoires and build their enduring social resources" (Fredrickson, 2004, p. 147). She elaborated that positive emotion broadens thought-action tendencies in three broad ways, namely, through broadening the scope of attention, the range of cognition and scope of action. It is reported that positive emotions elicit a well-thought response from beneficiary instead of a narrow tit-for-tat reaction. Grateful people creatively explore several activities as probable reflections of their gratitude. They tend to think consciously about how to meet the needs of their benefactor, which broadens their attention toward the giver's needs (Watkins *et al.*, 2014). Several studies, including Peterson and Stewart (1996), Graham (1998), McCullough *et al.* (2001), also favored the broad perspective of the theory.

Fredrickson also argued that gratitude builds resources in at least four ways. Firstly, gratitude is assumed to develop and strengthen social bonds. Secondly, in addition to establishing a connection between benefactor and beneficiary, gratitude also strengthens one's relationship with the divine. Several researchers, including Uhder *et al.* (2010), Watkins *et al.* (2014), also supports the fact that gratitude builds and strengthens spirituality. Thirdly, gratitude promoted civil communities and organizations. Then finally, gratitude is poised to expand one's "more general skills for loving" (Fredrickson, 2004, p. 152). In a nutshell, gratitude is poised to transform both at individual and organizational levels. Following the same analogy, this study proposed that gratitude broadens one's social intelligence, and it builds leadership skills.

Now, this study proposes that gratitude improves SA, SS, and social information processing. Broaden perspective of positive emotion elaborated expansion of thought-action

repertoires, the scope of attention, cognition and action and awareness of other's needs. Then, these expanded repertoires seem to have a direct bearing on the social intelligence of the employees. Few researchers, including Covey (1996), Boyatzis and Saatcioglu (2008) reported a positive relationship between gratitude and some elements of social intelligence. Further, banking on Fredrickson's explanation, it is proposed that gratitude broadens social intelligence, which builds leadership resources (skills) for a manager.

The proposed association between gratitude and leadership could also be hypothesized based on Algoe and Stanton (2012) *find, remind and bind theory of gratefulnes*s. Algoe's work is an extension of Fredrickson's philosophy, and it tends to explain the mechanism through which gratitude builds resources for an individual. According to this theory, gratitude helps one find new relationships in social and organizational settings. A grateful person creates and avails more opportunities to initiate a social exchange with others, which facilitates the establishment of a new connection. It also maintains an ongoing relationship by reminding the blessing of association with others. A grateful person cherishes the social context, and thus, makes conscious and consistent efforts to manage them. This effort tends to bind these social affiliations and makes them long-standing. Again, based on this explanation of Find, Remind and Bind theory of gratitude, it could be proposed that gratitude works on one's social intelligence, which bestows them with permanent and favorable social affiliation. Then, good social association with colleagues and followers is the hallmark of success for a leader.

Garg and Gera (2019) explored the mediating effect of social intelligence amidst the relationship between gratitude and leadership. They explained that social intelligence partially mediates the intended association. The current study aspired to further strengthen the findings of the authors by exploring the mediating effect of three constructs of social intelligence. It has also been argued that an individual's personality traits like social intelligence may influence the relationship between gratitude and leadership. Accordingly, the present tends to investigate both mediation and moderation effect with the help of the following theoretical models (Figures 1 and 2) and hypotheses.

H1. Gratitude is positively related to leadership.

H2. Gratitude is positively related to SA.

H3. Gratitude is positively related to SS.

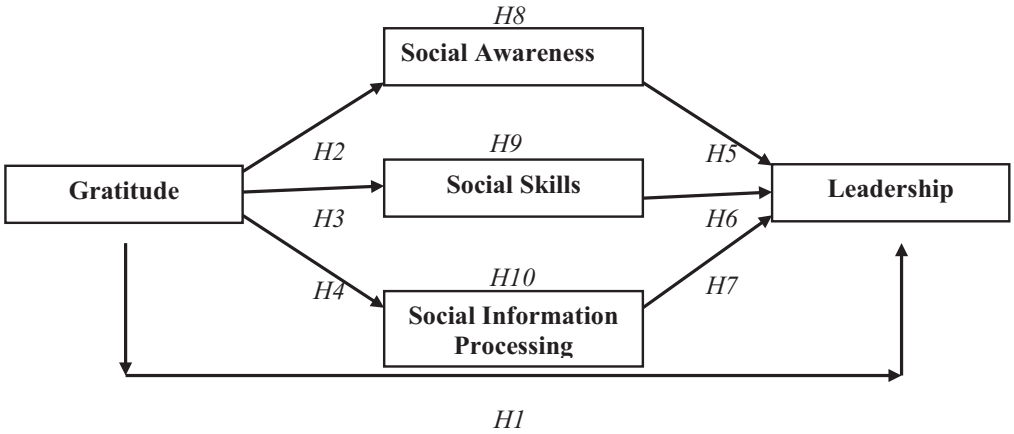


Figure 1.
Mediation model

H4. Gratitude is positively related to social information processing.

H5. SA is positively related to leadership.

H6. SS is positively related to leadership.

H7. SIP is positively related to leadership.

H8. SA mediates the relationship between gratitude and leadership.

H9. SS mediates the relationship between gratitude and leadership.

H10. SIP mediates the relationship between gratitude and leadership.

H11. SA moderates the relationship between gratitude and leadership.

H12. SS moderates the relationship between gratitude and leadership.

H13. SIP moderates the relationship between gratitude and leadership.

Methodological framework

Objectives, sample and procedures

The principal aim of this study was to investigate the linkage between gratitude and leadership among university teachers. It also tended to examine the mediating and moderating effect of social intelligence amidst the relationship between gratitude and leadership. As social intelligence is a multi-dimensional concept, this paper explored the mediating and moderating effect of three dimensions of social intelligence, i.e. SA, SS and social information processing. The sample comprised of 252 university professors teaching across various private, government, and autonomous institutes across Delhi- National Capital Region (NCR), Haryana and Punjab. Respondents were approached either personally or through emails. A total of 300 University teachers were approached and 255 teachers reverted. Three questionnaires were discarded for being incomplete. Now, colleagues of these 252 teachers were approached. Respondent was asked to respond to gratitude questionnaire (GQ-6) and TSIS and one of the colleagues of the respondent was approached to access leadership skills of respondent through multifactor leadership questionnaire (MLQ-5X). The convenience sampling method was used to reach out to these 300 respondents. Respondents were assured regarding the anonymity of responses and for strictly academic usage of collected data. Out of 252 respondents, 124 (49%) were male and 128 (51%) were female. Also, 19% of teachers were below 25 years, 26% were between 26-35 years, 21% were between 36-45 years, 16% were between 46-55 years and the remaining 18% were above the age of 55 years. Finally, the majority of respondents (51%) were from Delhi-NCR.

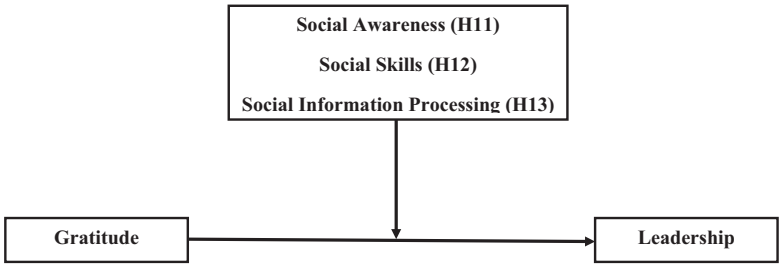


Figure 2.
Moderation model

Measures

Gratitude. It was measured with the help of a GQ-6, a self-report instrument developed by McCullough *et al.* (2002). This scale has six items rated on a seven-point scale ranging from 1 (strongly disagree) to 7 (strongly agree). Out of six items, two items are reverse scored on this scale. GQ-6 is one of the most practiced gratitude scales in academic researches (Froh *et al.*, 2008; Yüksel and Oguz Duran, 2012; Sun and Kong, 2013; Kong *et al.*, 2015; Caputo, 2016; Langer *et al.*, 2016). The six items include "If I had to list everything, I felt grateful, it would be a very long list," "I have so much in life to be thankful for" and "Long amounts of time can go by before I feel grateful to something or someone" (reverse-scored).

Social intelligence. Social Intelligence was measured with the 21 items TSIS developed by Silvera *et al.* (2001). TSIS accesses three dimensions of social intelligence, namely, SA (SA), SS (SS) and SIP. Seven statements are reverse scored, and each statement is measured on the seven-point rating scale. SA is measured with items like "Other people become angry with me without me being able to explain why" and "I have often hurt others without realizing it." Examples of items included to measure SS are "I fit in easily in social situations" and "I have a hard time getting along with other people." SIP is accessed with items like "I know how my actions will make others feel" and "I understand others' wishes."

Leadership. Leadership is measured with the help of the 21 items MLQ-5X developed by Rowold (2004). It accesses seven dimensions of leadership (inspirational motivation, idealized influence (attributed), idealized influence (behavior), individualized consideration, intellectual stimulation, contingent reward and management by exception) rated on a five-point rating scale ranging from 1 (strongly disagree) to 5 (strongly agree).

Reliability and validity

Reliability refers to the consistency of a measure. It is the degree to which similar results are obtained under consistent conditions (Malhotra and Dash, 2018). In this paper, internal consistency reliability, which is the consistency of the respondent's response across different items of the scale, is accessed Cronbach's alpha. It measures split-half correlation through splitting items into two sets, and Cronbach's alpha represents mean of all possible split-half correlations for a set of items. According to George and Mallery (2003), Cronbach's alpha should be greater than 0.70 to confirm reliability. All measures which confirm internal consistency reliability are reported in Table 1 (Cronbach's alpha value for gratitude = 0.877, SA = 0.748, SS = 0.709, SIP = 0.736 and leadership = 0.778). It is suggested to calculate composite reliability along with Cronbach's alpha (Cronbach and Shavelson, 2004). Composite reliability estimates should also be greater than 0.70 to reaffirm reliability (Hair *et al.*, 2010). Also, Convergent and Discriminant validity is determined with average variance extracted (AVE), which was calculated by summation of square of each factor

Variable	Mean	S.D.	α	CR	AVE	VIF
Gratitude	4.68	1.231	0.888	1.270	0.877	1.396
Social awareness	3.83	0.565	0.800	0.584	0.748	2.062
Social skills	3.51	0.643	0.782	0.633	0.709	2.727
Social information processing	3.34	0.761	0.794	0.764	0.736	2.475
Leadership	3.65	0.732	0.773	0.713	0.778	1.231

Notes: S.D. – Standard deviation; α – Cronbach's alpha; CR – Composite reliability; AVE – Average variance extracted; VIF – Variance inflation index
Source: Primary data

Table 1.
Descriptive statistics

loading divided by the number of indicators. AVE must be greater than 0.50 for convergent validity (Fornell and Larcker, 1981). AVE values are reported in Table 1. Discriminant validity indicates distinctiveness. It reflects the degree to which one construct is different from another construct (Bryman and Bell, 2014). Fornell and Larcker (1981) suggested a simple method to ensure Discriminant validity i.e. AVE's value should be higher than the squared inter-correlations of constructs. Table 3 depicts the results of Fornell and Larcker's (1981) results. AVEs were compared (highlighted in bold in the diagonal) with the squared intercorrelations (non-bold entries and below the diagonal). It was found that AVE was greater than the squared correlations, and thereby, it supports the existence of discriminant validity.

Multicollinearity and common method bias

According to Kline (2009), the problem of multi-collinearity surfaces when the correlations among variables are high. In this study, multi-collinearity was explored with the variance inflation factor (VIF), which should be less than five (Kline, 2009). The present study did not have an issue with multi-collinearity, as all VIF values are less than five (Table 1). Detrimental effects of method variance include biased estimates of reliability and validity, inflated or deflated estimates of the relationship between two constructs, affect hypothesis testing, lead to Type I or Type II error and deflated interaction effect (Podsakoff *et al.*, 2012). To minimize common method bias, this study adopted both procedural and statistical remedies. Firstly, pre-validated measurement scales with a different rating scale for predictor (seven-point) and criterion (five-point) were used. Secondly, the anonymity of the respondents was assured and strictly maintained. Also, respondents (university teachers) were explained how the study and its findings might be useful for them. It motivates respondents to respond more accurately and positively (Podsakoff *et al.*, 2012). Also, the independent variable and dependent variables were measured from two different sources. Respondent was asked to respond to GQ-6 and TSIS, and one of the colleagues of the respondent was approached to access the leadership skills of the respondent through MLQ-5X. Apart from these procedural methods, the Harman one-factor was also used to check common method bias. More than one factor was reported when items of different measures were subjected to factor analysis.

Statistical tools

The collected data is analyzed with the help of different statistical tools such as the measure of central tendency, standard deviation, correlation, hierarchical regression, Sobel test and bootstrapping.

Data analysis and interpretation

Table 1 shows means, standard deviations, and reliability, validity and multicollinearity estimates of variables of the study (Tables 1 and 2). Table 2 elucidates the correlation among variables of the study. Results suggest that gratitude is positively associated with the leadership of the teacher and with three dimensions of social intelligence. Hence, *H1–H4* are supported. Also, SA, SS, and SIP are positively correlated with leadership. Thus, *H5–H7* are confirmed (Tables 3 to 5).

Now, mediating effect of SA, SS and SIP on the relationship between gratitude and transformational leadership is explored with Baron and Kenny's (1986) four-step approach, in a combination of "Sobel test" and "bootstrap of the indirect effect" as suggested by Zhao *et al.* (2010). As per Baron and Kenny (1986), the first three conditions for mediation are significant regression between an independent variable and dependent variable,

independent variable and mediator and mediator and dependent variable. In the fourth step, a mediator is controlled, and the independent variable is again regressed on the dependent variable. Complete mediation is present when the predictor variable no longer predicts criterion variable and partial mediation is present when the influence of the independent variable on the dependent variable is reduced on the controlling effect of the mediating variable.

Table 4 presents path coefficients and significance levels for the proposed mediation hypotheses using hierarchical regression. It is argued that Baron and Kenny's (1986) method provides a fair idea about mediation, but the statistical significance of the mediator could be examined using a Sobel test (Yay, 2017). It is also argued that both these methods assume a normal distribution of data. Thus, researchers advocate the bootstrapping process, which is resampling with a replacement for large times such as 5,000 or 10,000 times, to confirm the mediation effect. Bootstrapping method is increasingly gaining acceptance among researchers (Shrout and Bolger, 2002; Fritz *et al.*, 2012). Table 5 shows the result of the Sobel and bootstrapping method. Results of Tables 3 and 4 suggest that SA, SS and SIP partially mediate the relationship between gratitude and transformational leadership. Hence, *H8–H10* are supported.

Finally, the moderation effect of three dimensions of social intelligence was analyzed by calculating the interaction effect of an independent variable (gratitude) and proposed moderators. It is being suggested that significant interaction confirms the moderation effect. Results of Table 6 show that none of the three dimensions of social intelligence (SA, SS and social information processing) moderates the relationship between gratitude and leadership. Hence, *H11*, *H12* and *H13* are rejected. Findings are discussed in the next section of the paper.

Discussion

In one of the most popular articles, 'grateful leaders make the best leaders,' Nieuwhof (2017) discussed five essential reasons for leaders to cultivate and cherish the virtue of gratitude.

Variables	1	2	3	4	5
Gratitude	1				
Social awareness	0.507**	1			
Social skills	0.454**	0.652**	1		
Social information processing	0.376**	0.620**	0.753**	1	
Leadership	0.549**	0.629**	0.736**	0.654**	1

Notes: **Correlation is significant at the 0.01 level (two-tailed)
Source: Primary data

Table 2. Correlation matrix

Variables	1	2	3	4	5
Gratitude	0.877				
Social awareness	0.257	0.748			
Social skills	0.206	0.426	0.709		
Social information processing	0.141	0.384	0.567	0.736	
Leadership	0.301	0.395	0.541	0.427	0.778

Source: Primary data

Table 3. Discriminant validity results

Firstly, he argued that gratitude impacts the overall attitude of a human being. It makes a person more compassionate, kinder, less suspicious and less resentful. Then all these qualities are essential traits of a good leader. Secondly, a sense of abundance allows a grateful leader to see and use an opportunity where other misses. According to Nieuwhof, this abundance thinking develops an abundance mentality that enables leaders to realize immense prospects of growth and development. Thirdly, gratitude may fuel generosity among leaders, which wins the hearts and souls of the followers. Also, people desperately wish to work with a grateful leader. In other words, gratitude attracts people to work in a team for larger organizational goals. The author claims that employees are ready to accept lower wages or even pay cuts to work under a grateful senior. Then finally, gratitude neutralizes anger and jealousy of a leader. Then, a calm and compose leader can easily touch the skies of success.

Table 4. Mediation analysis (results of hierarchical regression)

Hypothesized relationships	IV to DV	IV to Med	M to DV (with mediator)	IV to DV	Hypotheses support
G→ SA →L	0.549*	0.507*	0.629*	0.472*	*H8*
G →SS →L	549*	0.454*	0.736*	0.499*	*H9*
G →SIP →L	0.549*	0.376*	0.654*	0.521*	*H10*

Notes: NS: Not significant; IV – Independent variable; DV – Dependent variable; M – Mediator; G – Gratitude; SA – Social awareness; SS – Social skills; SIP – Social information processing; L – Leadership
*$p < 0.05$
Source: Primary data

Table 5. Mediation analysis (results of sobel test and bootstrapping)

	Sobel test			Bootstrapping estimates		
Hypothesized relationships	*z*-value	Sig.	Effect	SE	LLCI	ULCI
G→SA→L	6.386	0.000*	0.1421	0.0284	0.0882	0.1992
G→SS→L	6.932	0.000*	0.1653	0.0312	0.1037	0.2254
G→SIP→L	5.559 0.000*		0.1163	0.0257	0.0658	0.1677

Notes: NS: Not significant; IV – Independent variable; DV – Dependent variable; M – Mediator; G – Gratitude; SA – Social awareness, SS – Social skills, SIP – Social information processing. *$p < 0.05$
Source: Primary data

Table 6. Moderation analysis

	Standardized coefficients		
Hypothesized relationships	β_1	β_2	β_3
$L = \beta_0 + \beta_1 G + \beta_2 SA + \beta_3 (G * SA)$	0.309(S)	0.468 (S)	−0.133(NS)
$L = \beta_0 + \beta_1 G + \beta_2 SS + \beta_3 (G * SS)$	0.273(S)	0.613 (S)	0.005(NS)
$L = \beta_0 + \beta_1 G + \beta_2 SIP + \beta_3 (G * SIP)$	0.334(S)	0.528(S)	−0.063(NS)

Notes: *$p < 0.05$; as interaction is insignificant so moderation does not exist; S: Significant, NS: Not significant, IV – Independent variable, DV – Dependent variable, M – Mediator, G – Gratitude, SA – Social awareness, SS – Social skills, SIP – Social information processing
Source: Primary data

Considering such expert opinions, objects of a broaden-and-build theory of positive emotions, attachment perspective of a social cognitive theory of gratitude, Fredrickson's theoretical predictions, previous literature and advice of practitioners, this study hypothesized that gratefulness influences leadership. It is revealed that gratitude is positively associated with leadership. The present study also suggests a statistically robust and positive relationship between gratitude and SA, SS and social information processing. Also, the mediation and moderation effect of three dimensions of social intelligence (SA, SS and SIP) are explored. It reports that all these three dimensions of social intelligence partially mediate the relationship between gratitude and transformational leadership. The findings suggest that cultivating gratitude broadens one's social intelligence, which builds leadership skills. These findings are consistent with the broaden-and-build theory of positive emotions and also with a prediction of Fredrickson.

SA refers to one's ability to emphatically sense social cues about another person's thoughts, aspirations and emotions, which helps them to understand others appropriately. Then, here gratitude hits the bull's eye. A grateful leader is presumed to analyze others' contributions correctly and then acknowledge their efforts through grateful and appreciative behaviors at an appropriate public forum. Leaders generally resort to the oral expression of thankfulness, appreciation note, reward, promotion and challenging assignment to express their gratitude toward their subordinates.

SS are competencies that promote interaction and communication with others. Then, researches suggested that managers who express gratitude make their subordinates more comfortable for voicing their concerns. Grateful people appear more warmth and responsive to others, and hence, they command their trust and faith (Fredrickson, 2004), which makes them easy to approach. This easiness and comfort of subordinates allows free flow of interaction between two parties and thereby facilitates SS. As far as SIP is concerned, it is one's ability to regulate negative emotions such as anxiety, nervousness, etc. in social situations. This study reports that gratitude influence the SIP capacity of leaders. The findings could be understood with the help of the "undoing effect" of positive emotions. Several studies, including Sansone *et al.* (2010), Allen (2018) and Jans-Beken (2018) reported the undoing effect of gratitude through neutralizing stress, anxiety, nervousness and depression.

The findings of the present study complement the results of previous studies. Garg and Gera (2019) also reported a positive association between gratitude and leadership. Afterward on, Garg (2020) advocated that gratitude is both necessary and sufficient conditions for the development of seven essential constructs of educational leadership. Waters and Stokes (2015) also reported that an effective school leader attends both emotion-gratitude and action-gratitude. The authors also supported that gratuitous behaviors of leaders reciprocate, and it is equally effective in good and challenging time. Howells (2013) discussed the importance of gratitude for school leadership in great length. The results of the present study support her qualitatively researched propositions.

Conclusion

Theoretical contribution of the study

The study makes a few significant contributions to the existing literature. Firstly, the current study aspires to address the scarcity of gratitude studies in organizational contexts. Fredrickson (2004), in her famous book *The Psychology of Gratitude,* suggested an exploration of gratitude for workplace betterment. She promoted future researchers to investigate the association between gratitude and employee performance, leadership skills, employee satisfaction level and profitability of the firm. Unfortunately, her call has not been

attended enough in past 15 years. Then, this is one of the few studies in the direction enlightened by Fredrickson. Secondly, findings of the study provide an alternate explanation of broaden-and-build theory. It could be inferred that gratitude as positive emotions broadens one's SS, which build resources in terms of leadership skills. Further, the study stresses the significance of gratitude in the context of leadership. Probably, this is the first empirical study to investigate the relationship between gratitude and leadership. Previous researchers have discussed the role of leaders' emotions in the staff's performance and productivity (Humphrey, 2002; Garg, 2017b). Combining the results of these studies with the findings of this study reveals that gratitude can be doubly beneficial for leaders. Firstly, gratitude can sharpen leadership skills and secondly, as emotion, gratitude can influence subordinates' performance. Future researchers can undoubtedly investigate the impact of grateful leaders on follower's performance. Positive emotions are contagious (Hatfield *et al.*, 1993), and feelings of organizational leaders are especially contagious (George, 1995). Thus, the gratuitous behavior of leaders may produce a cascading effect through eliciting a similar response from subordinates.

Fourthly, this resolves the "black-box" of a path through which the effect of gratitude transmits to leadership. The results show that SA, SS and SIP partially mediate the relationship between gratitude and leadership. In summary, gratitude influences social intelligence, which horns one's leadership skills. Again, this study is one of the first explorations to explore the mediating and moderating effect of social intelligence amid the relationship between gratitude and transformational leadership. Then finally, this study is conducted in Indian settings, and most of the previous studies on gratitude were conducted in western contexts. As discussed, the findings of this study are complementing the broaden-and-build theory and social attachment perspective of a social cognitive theory of emotions. These theories were developed in non-eastern conditions. Given vast cultural differences, this study provides an appropriate avenue of application of these theories in Indian states.

Practical implications of the study

The findings of this study provide several practical implications for practitioners too. Firstly, leadership skill trainers can help participants in cultivating gratitude to promote leadership skills. Trainers may use different gratitude practices and interventions including "gratitude journals" (Froh *et al.*, 2008; Froh *et al.*, 2007), "gratitude letters" (Seligman, 2002; Sergeant and Mongrain, 2011), "gratitude letters visits" "cultivating mindfulness" (Langer, 2014; Langer *et al.*, 2010), "gratitude collages" (Reivich, 2009), "sharing positive events" (Gable *et al.*, 2004), "acts of kindness" (Sheldon *et al.*, 2013) and "meditation and progressive muscle relaxation techniques" (Shapiro, Schwartz and Santerre, 2002). Secondly, as discussed, leaders' expression of gratefulness is not only beneficial for leaders; it also influences subordinates' performance too. Thus, organizations are suggested to motivate their line managers to cultivate the virtue of gratitude. This could quickly be done by imbibing gratuitous behavior in organizational culture. Again as suggested by previous researchers that positive emotions are contagious, gratuitous behavior by top management will motivate line managers to follow suit. Thirdly, findings suggest that social intelligence is a significant predictor of leadership. Although more researches are required for generalization, a recruiter may start discussions and debates on the inclusion of social intelligence test for the selection of managers.

Also, Fehr *et al.* (2017) proposed a multilevel model of gratitude in organizations. According to this model, gratitude could be observed at three levels in the workplace, namely, episodic gratitude ("feeling grateful in response to a particular experience"),

persistent gratitude ("a stable tendency to feel grateful within a particular context") and collective gratitude ("persistent gratitude that is shared by the members of an organization"). The authors also argued that these different forms of gratitude complement and support each other. If an organization facilitates institutionalization of gratitude at organizational level, then it facilitates cultivation of both episodic and persistent gratitude at individual level. Thus, the present study recommends organizational level gratitude programs such as "Formal appreciation program," "Thanks-giving day or week," and "Personalized gratuitous letters from managers and leaders." In recent commendable initiatives, contemporary organizations have started paid leaves to employees to serve the needy people. Under the Corporate Social Responsibility initiative, these companies allow a group of employees to submit a socially-beneficial project. If such a project is approved, the organization not only funds the project, but employees are also given paid leaves. Such schemes foster an organizational culture based on gratitude and trust.

Limitations, scope for future studies and conclusion

The present study provides an empirical model of gratitude, social intelligence, and transformational leadership. However, these contributions come in light of a few limitations. Firstly, this study is based on a cross-sectional research design, and also data is collected using self-reported measures. Thus, there are chances of common method variance (CMV). Although CMV was not reported in this study, a longitudinal study is required to conclude any causal and hypothesized relationships. Future researchers may conduct longitudinal studies to confirm the findings of this study. Secondly, the present study is based on a sample of 252 employees only. A larger sample comprising of respondents from diverse industries would be appropriate for generalization of the results. Future research should explore a larger sample spread over different companies. Thirdly, this study examined the mediating effect of social intelligence only. Future researchers may test a range of variables such as contentment, satisfaction perceived organizational support for possible mediation effect. Also, gratitude is measured with the help of a one-dimensional scale (GQ-6). Future researchers may use multidimensional scales like GRAT for better results. The present has considered consolidated values of gratitude and leadership measures, and future researchers may opt for in-depth analysis by considering different constructs of gratitude and leadership. It would be interesting to examine

In conclusion, the study provides empirical evidence supporting a positive association between gratitude and transformational leadership. Moreover, all three dimensions of social intelligence proved to be partially mediators of the relationship between gratitude and leadership. Besides, SS were also reported as moderator of a hypothesized relationship. As this study is a pioneer study in the field of gratitude, social intelligence and leadership, the results bring significant contributions that enhance understanding and support further research in this specific area of study.

References

Agran, M., Hughes, C., Thoma, C.A. and Scott, L.A. (2016), "Employment social skills: what skills are really valued?", *Career Development and Transition for Exceptional Individuals*, Vol. 39 No. 2, pp. 111-120.

Algoe, S.B. and Stanton, A.L. (2012), "Gratitude when it is needed most: social functions of gratitude in women with metastatic breast cancer", *Emotion*, Vol. 12 No. 1, pp. 163-168.

Alkozei, A., Smith, R. and Killgore, W.D. (2018), "Gratitude and subjective wellbeing: a proposal of two causal frameworks", *Journal of Happiness Studies*, Vol. 19, pp. 1519-1542.

Baron, R.M. and Kenny, D.A. (1986), "The moderator-mediator variable distinction in social psychological research: conceptual, strategic and statistical considerations", *Journal of Personality and Social Psychology*, Vol. 51 No. 6, pp. 1173-1182.

Bartlett, M.Y. and DeSteno, D. (2006), "Gratitude and prosocial behavior helping when it costs you", *Psychological Science*, Vol. 17 No. 4, pp. 319-325.

Bass, B.M. (1990), "From transactional to transformational leadership: learning to share the vision", *Organizational Dynamics*, Vol. 18 No. 3, pp. 19-31.

Boyatzis, R.E. and Saatcioglu, A. (2008), "A 20-year view of trying to develop emotional, social and cognitive intelligence competencies in graduate management education", *Journal of Management Development*, Vol. 27 No. 1, pp. 92-108.

Bryman, A. and Bell, E. (2014), *Research Methodology: business and Management Contexts*, *Oxford University Press Southern Africa*, *Cape Town*.

Cronbach, L.J. and Shavelson, R.J. (2004), "My current thoughts on coefficient alpha and successor procedures", *Educational and Psychological Measurement*, Vol. 64 No. 3, pp. 391-418.

Davies, M., Cooper, G., Kettler, R.J. and Elliott, S.N. (2015), "Developing social skills of students with additional needs within the context of the Australian curriculum", *Australasian Journal of Special Education*, Vol. 39 No. 1, pp. 37-55.

Dik, B.J., Duffy, R.D., Allan, B.A., O'Donnell, M.B., Shim, Y. and Steger, M.F. (2014), "Purpose and meaning in career development applications", *The Counseling Psychologist*, Vol. 43 No. 4, pp. 1-28, doi: 10.1177/0011000014546872.

Edmondson, A. (2002), "The local and variegated nature of learning in organizations: a grouplevel perspective", *Organization Science*, Vol. 13 No. 2, pp. 128-146.

Emmons, R.A. and Crumpler, C.A. (2000), "Gratitude as human strength: appraising the evidence", *Journal of Social and Clinical Psychology*, Vol. 19 No. 1, pp. 56-69, doi: 10.1521/jscp.2000. 19.1.56.

Emmons, R.A. and McCullough, M.E. (2003), "Counting blessings versus burdens: an experimental investigation of gratitude and subjective well-being in daily life", *Journal of Personality and Social Psychology*, Vol. 84 No. 2, pp. 377-389.

Fehr, R., Fulmer, A., Awtrey, E. and Miller, J.A. (2017), "The grateful workplace: a multilevel model of gratitude in organizations", *Academy of Management Review*, Vol. 42 No. 2, pp. 361-381, doi: 10.5465/amr.2014.0374.

Fornell, C. and Larcker, D.F. (1981), "Evaluating structural equation models with unobservable variables and measurement error", *Journal of Marketing Research*, Vol. 18 No. 1, pp. 39-50.

Fredrickson, B.L. (1998), "What good are the positive emotions?", *Review of General Psychology*, Vol. 2 No. 3, pp. 300-319.

Fredrickson, B.L. (2004), "The broaden-and-build theory of positive emotions", *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, Vol. 359 No. 1449, pp. 1367-1378.

Fritz, M.S., Taylor, A.B. and MacKinnon, D.P. (2012), "Explanation of two anomalous results in statistical mediation analysis", *Multivariate Behavioral Research*, Vol. 47 No. 1, pp. 61-87.

Froh, J.J., Miller, D.N. and Snyder, S.F. (2007), "Gratitude in children and adolescents: development, assessment, and school-based intervention", *School Psychology Forum*, Vol. 2, pp. 1-13.

Froh, J.J., Sefick, W.J. and Emmons, R.A. (2008), "Counting blessings in early adolescents: an experimental study of gratitude and subjective well-being", *Journal of School Psychology*, Vol. 46 No. 2, pp. 213-233.

Froh, J.J., Emmons, R.A., Card, N.A., Bono, G. and Wilson, J.A. (2011), "Gratitude and the reduced costs of materialism in adolescents", *Journal of Happiness Studies*, Vol. 12 No. 2, pp. 289-302.

Gable, S.L., Reis, H.T., Impett, E.A. and Asher, E.R. (2004), "What do you do when things go right? The intrapersonal and interpersonal benefits of sharing positive events", *Journal of Personality and Social Psychology*, Vol. 87 No. 2, pp. 228-245.

Garg, N. (2017a), "Workplace spirituality and employee well-being: an empirical exploration", *Journal of Human Values*, Vol. 23 No. 2, pp. 1-19.

Garg, N. (2017b), "Workplace spirituality and organizational commitment: an empirical exploration", *Purushartha- A Journal of Management, Ethics, and Spirituality*, Vol. 10 No. 2, pp. 8-24.

Garg, N. (2018), "Promoting organizational performance in indian insurance industry: the roles of workplace spirituality and organizational citizenship behaviour", *Global Business Review*, Vol. 1, doi: 10.1177/0972150918778983.

Garg, N. (2020), "Exploring role of gratitude in developing teacher leadership in Indian universities", *International Journal of Educational Management*, Vol. 34 No. 5, pp. 881-901, available at: www.emerald.com/insight/publication/issn/0951-354X

Garg, N. and Gera, S. (2019), "Gratitude and leadership in higher education institutions: exploring the mediating role of social intelligence among teachers", *Journal of Applied Research in Higher Education*, Vol. ahead-of-print No. ahead-of-print.

Garg, N., Punia, B.K. and Jain, A. (2019), "Workplace spirituality and job satisfaction: exploring mediating effect of organizational citizenship behaviour", *Vision: The Journal of Business Perspective*, Vol. 23 No. 3, pp. 287-296.

George, J.M. (1995), "Leader positive mood and group performance: the case of customer service", *Journal of Applied Social Psychology*, Vol. 25, pp. 778-794.

George, D. and Mallery, P. (2003), *SPSS for Windows Step by Step: A Simple Guide and Reference*, 4the ed., Allyn, Bacon, Boston, 11.0 update.

Giles, D. and Yates, R. (2014), "Enabling educational leaders: qualitatively surveying an organization's culture", *International Journal of Organizational Analysis*, Vol. 22 No. 1, pp. 94-106.

Goleman, D. (2006), *Social Intelligence: The New Science of Human Relationships*, Bantam Books.

Grant, A.M. and Gino, F. (2010), "A little thanks goes a long way: explaining why gratitude expressions motivate prosocial behaviour", *Journal of Personality and Social Psychology*, Vol. 98 No. 6, pp. 946-955.

Grant, A.M., Wrzesniewski, A. (2010), "I won't let you down . . . or will I? Core Self-evaluations, other-orientation, anticipated guilt and gratitude, and job performance", *Journal of Applied Psychology*, Vol. 95 No. 1, pp. 108-121.

Greer, C.R. and Plunkett, W.R. (2000), *Supervision: Diversity and Teams in the Workplace*, 9th ed., Prentice Hall, Upper Saddle River, NJ.

Hair, J.F., Black, W.C., Babin, B.J. and Anderson, R.E. (2010), *Multivariate Data Analysis*, 7th ed., Prentice Hall, Upper Saddle River, NJ.

Hatfield, E. Cacioppo, J.T. and Rapson, R.L. (1993), "Emotional contagion", Vol. 2, No. 3, pp. 96-100, available at: https://doi.org/10.1111%2F1467-8721.ep10770953

Howells, K. (2013), "The role of gratitude in helping school leaders strengthen relationships", *Perspectives on Educational Leadership*, Vol. 6 No. 1, pp. 1-3.

Hu, X. and Kaplan, S. (2014), "Is "feeling good" good enough? Differentiating discrete positive emotions at work", *Journal of Organizational Behavior*, Vol. 36 No. 1, pp. 39-58.

Humphrey, R.H. (2002), "The many faces of emotional leadership", *Leadership Quarterly*, Vol. 13, No. 5, pp. 493-504.

Jans-Beken, L., Lataster, J., Peels, D., Lechner, L. and Jacobs, N. (2018), "Gratitude, psychopathology and subjective well-being: results from a 7.5-month prospective general population study", *Journal of Happiness Studies*, Vol. 19, pp. 1673-1689.

Kaplan, S., Bradley-Geist, J.C., Ahmad, A., Anderson, A., Hargrove, A.K. and Lindsey, A. (2014), "Test of two positive psychology interventions to increase employee well-being", *Journal of Business and Psychology*, Vol. 29 No. 3, pp. 367-380.

Kline, R.B. (2009), *Becoming a Behavioural Science Researcher: A Guide to Producing Research That Matters*, Guildford Press New York, NY, Chapter 7, Measurement.

Kong, F., Ding, K. and Zhao, J. (2015), "The relationships among gratitude, self-esteem, social support and life satisfaction among undergraduate students", *Journal of Happiness Studies*, Vol. 16 No. 2, pp. 477-489.

Langer, E.J. (2014), "Mindfulness forward and back", *Wiley Online Library*, available at: https://doi.org/10.1002/9781118294895.ch1

Langer, E., Djikic, M., Pirson, M., Madenci, A. and Donohue, R. (2010), "Believing is seeing using mindlessness (mindfully) to improve visual acuity", *Psychological Science*, Vol. 21, pp. 661-666.

Malhotra, N.K. and Dash, S. (2018), "Marketing research an applied orientation", Seventh Edition Revised.

Marlowe, H.A. (1986), "Social intelligence: evidence for multidimensionality and construct independence", *Journal of Educational Psychology*, Vol. 78 No. 1, pp. 52-58.

McCullough, M.E., Emmons, R.A. and Tsang, J.A. (2002), "The grateful disposition: a conceptual and empirical topography", *Journal of Personality and Social Psychology*, Vol. 82 No. 1, pp. 112-127.

McCullough, M.E., Emmons, R.A., Kilpatrick, S.D. and Larson, D.B. (2001), "Is gratitude a moral affect?", *Psychological Bulletin*, Vol. 127 No. 2, pp. 249-266.

Nieuwhof, C. (2017), "5 Crucial ways your gratitude impacts your attitude as leader", available at: https://careynieuwhof.com/5-ways-your-gratitude-impacts-your-attitude-as-a-leader/ (accessed 12 November 2019).

Northouse, G. (2007), *Leadership Theory and Practice*, 3rd ed., Sage Publications, Thousand Oaks, CA.

Peterson, B.E. and Stewart, A.J. (1996), "Antecedents and contexts of generativity motivation at mid-life", *Psychology and Aging*, Vol. 11, pp. 21-33.

Podsakoff, P.M., MacKenzie, S.B. and Podsakoff, N.P. (2012), "Sources of method bias in social science research and recommendations on how to control it", *Annual Review of Psychology*, Vol. 63 No. 1, pp. 539-569.

Reivich, K. (2009), "Fishful thinking: Cultivating gratitude in youth", *National Association of School Psychologists*, Vol. 38 No. 3, available at: www.nasponline.org/publications/cq/index.aspx

Rosenberg, E.L. (1998), "Levels of analysis and the organization of affect", *Review of General Psychology*, Vol. 2 No. 3, pp. 247-270.

Rowold, J. (2004), "MLQ-5X for sport coaches", German translation and adaptation of Bass & Avolio's Multifactor Leadership Questionnaire Redwood City: Mind Garden.

Sansone, R.A. and Sansone, L.A. (2010), "Gratitude and well-being: the benefits of appreciation", *Psychiatry*, Vol. 7 No. 11, pp. 18-22.

Seligman, M.E.P. (2002), *Authentic Happiness: Using the New Positive Psychology to Realize Your Potential for Lasting Fulfillment*, *Free Press*, *New York, NY*.

Sergeant, S. and Mongrain, M. (2011), "Are positive psychology exercises helpful for people with depressive personality styles?", *The Journal of Positive Psychology*, Vol. 6 No. 4, pp. 260-272, doi: 10.1080/17439760.2011.577089.

Shapiro, S.L., Schwartz, G.E.R. and Santerre, C. (2002), "Meditation and positive psychology", in Snyder, C.R. and Lopez, S. (Eds), *Handbook of Positive Psychology*, Oxford University Press, London, pp. 632-645.

Sheldon, K.M., Boehm, J.K. and Lyubomirsky, S. (2013), "Variety is the spice of happiness: the hedonic adaptation prevention (HAP) model", in Boniwell I. and David S. (Eds), *Oxford Handbook of Happiness*, Oxford University Press, Oxford, pp. 901-914.

Shrout, P.E. and Bolger, N. (2002), "Mediation in experimental and nonexperimental studies: new procedures and recommendations", *Psychological Methods*, Vol. 7 No. 4, pp. 422-445.

Silvera, D.H., Martinussen, M. and Dahl, T.I. (2001), "The tromsø social intelligence scale, a self-report measure of social intelligence", *Scandinavian Journal of Psychology*, Vol. 42 No. 4, pp. 313-319.

Spence, J.R., Brown, D.J. and Keeping, L.M., L., H. (2013), "Helpful today, but not tomorrow? Feeling grateful as a predictor of daily organizational citizenship behaviors", *Pers. Psychol*, Vol. 67, pp. 705-738.

Sun, P. and Kong, F. (2013), "Affective mediators of the influence of gratitude on life satisfaction in late adolescence", *Social Indicator Research*, Vol. 114 No. 3, pp. 1361-1369.

Uhder, J., Watkins, P.C. and Hammamoto, D. (2010), "Would the humble please stand: Can self-reported humility be valid?", Paper presented to the annual convention of the American Psychological Association, San Diego, CA.

Waters, L. (2012), "Predicting job satisfaction: contributions of individual gratitude and institutionalized gratitude", *Psychology*, Vol. 03 No. 12, pp. 1174-1176.

Waters, L. and Stokes, H. (2015), "Positive education for school leaders: exploring the effects of emotion-gratitude and action-gratitude", *The Educational and Developmental Psychologist*, Vol. 32 No. 1, pp. 1-22.

Watkins, P.C., Uhder, J. and Pichinevskiy, S. (2014), "Grateful recounting enchances subjective well-being: the importance of grateful processing", *The Journal of Positive Psychology*, Vol. 10 No. 2, pp. 91-98.

Watkins, P.C., Woodward, K., Stone, T. and Kolts, R.L. (2003), "Gratitude and happiness: development of a measure of gratitude, and relationships with subjective well-being", *Social Behavior and Personality: An International Journal*, Vol. 31 No. 5, pp. 431-451.

Wood, A.M., Froh, J.J. and Geraghty, A.W.A. (2010), "Gratitude and well-being: a review and theoretical integration", *Clinical Psychology Review*, Vol. 30 No. 7.

Yay, M. (2017), "The mediation analysis with the sobel test and the percentile bootstrap", *International Journal of Management and Applied Science (IJMAS)*, Vol. 3 No. 2, pp. 123-125.

Yüksel, A. and Oguz Duran, N. (2012), "Turkish adaptation of the gratitude questionnaire", *Eurasian Journal of Educational Research*, Vol. 46, pp. 199-216.

Zhao, X., Lynch, J.G. and Chen, Q. (2010), "Reconsidering Baron and Kenny: myths and truths about mediation analysis", *Journal of Consumer Research*, Vol. 37 No. 2, pp. 197-206.

Further reading

Al-Hashimi, M. and Al-Barri, Z. (2017), "Gratitude intervention at work: its impact on well-being, job satisfaction and organisational commitment", *International Journal Advances in Social Science and Humanities*, Vol. 5 No. 3, pp. 01-10.

Bartlett, M.Y., Condon, P., Cruz, J., Baumann, J. and Desteno, D. (2012), "Gratitude: prompting behaviors that build relationships", *Cognition and Emotion*, Vol. 26 No. 1, pp. 2-13.

Bono, G., Emmons, R.E. and McCullough, M.E. (2004), "Gratitude in practice and the practice of gratitude", *Positive Psychology in Practice*, Vol. 1, pp. 464-481.

Bothma, C.F.C. and Roodt, G. (2013), "The validation of the turnover intention scale", *SA Journal of Human Resource Management*, Vol. 11 No. 1.

Chen, X.P., Hui, C. and Sego, D.J. (1998), "The role of organizational citizenship behavior in turnover: Conceptualization and preliminary tests of key hypotheses", *Journal of Applied Psychology*, Vol. 83 No. 6, pp. 922-931.

Coyne, I. and Ong, T. (2007), "Organizational citizenship behavior and turnover intention: a cross-cultural study", *The International Journal of Human Resource Management*, Vol. 18 No. 6, pp. 1085-1097.

Du Plooy, J. and Roodt, G. (2010), "Work engagement, burnout, and related constructs as predictors of turnover intentions", *SA Journal of Industrial Psychology*, Vol. 36 No. 1.

Emmons, R.A. and Mishra, A. (2010), "Why gratitude enhances well-being: what we know", *What We Need to Know," Designing Positive Psychology: Taking Stock and Moving Forward*, *Oxford University Press*, *New York, NY*, pp. 248-262.

Emmons, R.A. and Stern, R. (2013), "Gratitude as a psychotherapeutic intervention", *Journal of Clinical Psychology*, Vol. 69 No. 8, pp. 846-855.

Erdil, O. and Müceldili, B. (2014), "The effects of envy on job engagement and turnover intention", *Procedia - Social and Behavioral Sciences*, Vol. 150, pp. 447-454.

Fabio, A.D., Palazzeschi, L. and Bucci, O. (2017), "Gratitude in organizations: a contribution for healthy organizational contexts", *Frontiers in Psychology*, Vol. 8, pp. 1-6.

Fehr, R., Fulmer, A., Awtrey, E. and Miller, J.A. (2014), "The grateful workplace: a multilevel model of gratitude in organizations", *Academy of Management Review*, Vol. 1

Fitzgerald, P. (1998), "Gratitude and justice", *Ethics*, Vol. 109 No. 1, pp. 119-153.

Foundation, J.T., Good, G., May, U.C.B. and Allen, S. (2018), "The science of gratitude", *The University of California at Berkeley Wellness Letter*, Vol. 32 No. 6.

Froman, L. (2009), *Positive Psychology in the Workplace*, Springer, New York, NY, pp. 59-69.

Garg, N. (2017c), "Workplace spirituality and organizational performance in indian context: mediating effect of organizational commitment, work motivation, and employee engagement", *South Asian Journal of Human Resources Management*, Vol. 4 No. 2, pp. 191-211.

Griffeth, R., Hom, P.W. and Gaertner, S. (2000), "A meta-analysis of antecedents and correlates of employee turnover: update, moderator tests, and research implications for the next millennium", *Journal of Management*, Vol. 26 No. 3, pp. 463-488.

Lambert, N.M., Fincham, F.D., Stillman, T.F. and Dean, L.R. (2009), "More gratitude, less materialism: the mediating role of life satisfaction", *The Journal of Positive Psychology*, Vol. 4 No. 1, pp. 32-42.

Layous, K., Sweeny, K., Armenta, C., Na, S., Choi, I. and Lyubomirsky, S. (2017), "The proximal experience of gratitude", *Plos One*, Vol. 12 No. 7, pp. 1-26.

Lin, C.P. (2008), "Clarifying the relationship between organizational citizenship behaviors, gender, and knowledge sharing in workplace organizations in Taiwan", *Journal of Business and Psychology*, Vol. 22 No. 3, pp. 241-250.

Lin, C.C. (2013), "A higher-Order gratitude uniquely predicts subjective well-being: incremental validity above the personality and a single gratitude", *Social Indicators Research*, Vol. 119 No. 2, pp. 909-924.

McCullough, M.E., Tsang, J.A. and Emmons, R.A. (2004), "Gratitude in intermediate affective terrain: links of grateful moods to individual differences and daily emotional experience", *Journal of Personality and Social Psychology*, Vol. 86 No. 2, pp. 295-309.

Podsakoff, P.M., MacKenzie, S.B., Paine, J.B. and Bachrach, D.G. (2000), "Organizational citizenship behaviors: a critical review of the theoretical and empirical literature and suggestions for future research", *Journal of Management*, Vol. 26 No. 3, pp. 513-563.

Podsakoff, N.P., Whiting, S.W., Podsakoff, P.M. and Blume, B.D. (2009), "Individual- and organizational-level consequences of organizational citizenship behaviors: a meta-analysis", *Journal of Applied Psychology*, Vol. 94 No. 1, pp. 122-141.

Podsakoff, N.P., Podsakoff, P.M., Mackenzie, S.B., Maynes, T.D. and Spoelma, T.M. (2014), "Consequences of unit-level organizational citizenship behaviors: a review and recommendations for future research", *Journal of Organizational Behavior*, Vol. 35, pp. 87-119.

Shanker, M. (2018), "Organizational citizenship behavior about employees' intention to stay in indian organizations", *Business Process Management Journal*, Vol. 24 No. 6.

Solom, R., Watkins, P.C., McCurrach, D. and Scheibe, D. (2016), "Thieves of thankfulness: traits that inhibit gratitude", *The Journal of Positive Psychology*, Vol. 12 No. 2, pp. 120-129.

Spence, J.R. (2013), "Helpful today but not tomorrow? Feeling grateful as a predictor of daily organizational citizenship behaviours", *Personnel Psychology*, No. 67, pp. 705-738.

Spence, J.R. (2014), "Helpful today but not tomorrow? Feeling grateful as a predictor of daily organizational citizenship behaviours", *Personnel Psychology*, No. 67, pp. 705-738.

Wood, A.M., Joseph, S. and Linley, P.A. (2007), "Coping style as a psychological resource of grateful people", *Journal of Social and Clinical Psychology*, Vol. 26 No. 9, pp. 1076-1093.

Appendix. Questionnaire

Gratitude scale

(1) I have so much in life to be thankful for.
(2) If I had to list everything that I felt grateful for, it would be a very long list.
(3) When I look at the world, I do not see much to be grateful for.*
(4) I am grateful to a wide variety of people.
(5) As I get older I find myself more able to appreciate the people, events, and situations that have been part of my life history.
(6) Long amounts of time can go by before I feel grateful to something or someone.

* *Items 3 and 6 are reverse-scored.

Social intelligence scale

(1) I can predict other people's behavior.
(2) I often feel that it is difficult to understand other's choices.
(3) I know how my actions will make others feel.
(4) I often feel uncertain around new people who I do not know.
(5) People often surprise me with the things they do.
(6) I understand other people's feelings.
(7) I fit in easily in social situations.
(8) Other people become angry with me without me being able to explain why.
(9) I understand other's wishes.
(10) I am good at entering new situations and meeting people for the first time.
(11) It seems as though people are often angry or irritated when I say what I think.
(12) I have a hard time getting along with other people.
(13) I find people unpredictable.
(14) I can often understand what others are trying to accomplish without the need for them to say anything.
(15) It takes a long time for me to get to know others well.
(16) I have often hurt others without realizing it.
(17) I can predict how others will react to my behavior.
(18) I am good at getting on good terms with new people.
(19) I can often understand what others really mean through their expression, body language etc.
(20) I frequently have problems finding good conversation topics.
(21) I am often surprised by others' reactions to what I do.

Leadership scale

(1) I make others feel good to be around me.
(2) I express with a few simple words what we could and should do.
(3) I enable others to think about old problems in new ways.
(4) I help others develop themselves.

(5) I tell others what to do if they want to be rewarded for their work.
(6) I am satisfied when others meet agreed-upon standards.
(7) I am content to let others continue working in the same way as always.
(8) Others have complete faith in me.
(9) I provide appealing images about what we can do.
(10) I provide others with new ways of looking at puzzling things.
(11) I let others know how I think they are doing.
(12) I provide recognition/rewards when others reach their goals.
(13) As long as things are working, I do not try to change anything.
(14) Whatever others want to do is O.K. with me.
(15) Others are proud to be associated with me.
(16) I help others find meaning in their work.
(17) I get others to rethink ideas that they had never questioned before.
(18) I give personal attention to others who seem rejected.
(19) I call attention to what others can get for what they accomplish.
(20) I tell others the standards they have to know to carry out their work.
(21) I ask no more of others than what is absolutely essential.

Corresponding author
Naval Garg can be contacted at: naval.garg@dtu.ac.in

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2999267, IEEE Journal of Emerging and Selected Topics in Power Electronics

Implementation Strategy for Resource Optimization of FPGA-Based Adaptive Finite Control Set-MPC using XSG for a VSI System

Vijay Kumar Singh, Ravi Nath Tripathi, and Tsuyoshi Hanamoto

***Abstract*— Finite control set-model predictive control (FCS-MPC) for power electronic systems has been implemented using field-programmable gate array (FPGA) considering an alternative solution to handle the computational burden of the control algorithm. However, FPGA resource utilization is a concern for optimum design and development of the FCS-MPC. This paper presents, an analytical approach based implementation strategies using the Xilinx system generator (XSG) considering the real-time operation of the FCS-MPC for a three-phase voltage source inverter (VSI). Moreover, an adaptive prediction model is proposed to adapt to the change in load current. The percentage THD, dynamic response, and steady-state error are evaluated for system performance analysis. The implementation strategies are compared based on FPGA resource utilization for the FCS-MPC in stationary *αβ* and rotating *dq* frames considering fixed, approximated and adaptive values of a coefficient used for current prediction. The look-up tables (LUTs), flip-flops and DSP slices have opted as utilization indices of FPGA resources for comparative analysis. The optimum design based controller model is used for FPGA-based experimental system implementation. The simulation and experimental results are used for the performance evaluation of the FCS-MPC in both *αβ* and *dq* frames. Moreover, change in switching state constraint is included in the cost function and controller performance is evaluated corresponding to the weighting factor for approximated and adaptive coefficient values.**

***Index Terms*— finite control set-model predictive control, field-programmable gate array, model-based design, voltage source inverter, Xilinx system generator.**

I. INTRODUCTION

MODEL predictive control (MPC) possesses appealing characteristics such as explicit use of a system model, flexibility of simultaneous handling of multiple constraints and easy inclusion of nonlinearities [1]–[4]. These distinctive features of MPC have created considerable attention towards researcher and MPC has been applied to a wide variety of power converter and drive applications. For instance, two-level voltage source inverter (VSI) [5], [6], three-level neutral-point clamped converter (NPC) [7], active front end rectifier [8], [9], cascaded H-Bridge inverter [10], three-phase direct matrix converters [11], [12], predictive control for UPS applications [13], predictive torque control (PTC) and field-oriented control (FOC) of an induction machine [2], [14], just to name a few.

The fundamental principle of the MPC is to have a plant or system model that is used for the prediction of the future behavior of control parameters and the control action is selected based on an optimization function or cost function in each sampling interval. Finite control set-model predictive control (FCS-MPC) is one of the classifications of MPC that uses the discrete-time model of a power converter. The cost function is computed based on the finite number of switching states of the power converter and the control action is directly applied to the converter. As compared to the generalized MPC method, the computational requirement is reduced in the FCS-MPC by using the discretized power converter model and a small prediction horizon (N_P = 1). The computations required during the implementation of predictive controls for power converters are accomplished by the microprocessor and digital signal processor (DSP) [3], [4], [14].

The prediction and cost function computations required for the algorithm in the practical implementation should strictly complete within a given sampling interval. However, the computational task is a concern especially for complex converter systems and creates a delay issue that deteriorates the quality of waveforms [5], [15]. This issue has been addressed and some methodologies for delay compensation have been presented to compute the optimum switching state within the specified sampling interval [16], [17]. However, delay compensation techniques have to deal with the increased computational burden and average switching frequency [16].

Field programmable gate array (FPGA) is a pragmatic choice for the implementation of MPC because of parallel processing capability [18]–[24]. In [18], a comparative analysis among various digital control platforms is presented for the implementation of model predictive control including FPGA. Moreover, the different approach is adopted for FPGA-based system implementation. In [19], an FPGA platform of NI-CRIO

The manuscript was submitted on November 18, 2019 for review.

V. K. Singh is with the Department of Life Science and Systems Engineering, Kyushu Institute of Technology, Wakamatsu-ku, Kitakyushu-shi, Fukuoka 808-0196 Japan (e-mail: vijay.ec1985@gmail.com).

R. N. Tripathi is with the Next Generation Power Electronics Research Center, Kyushu Institute of Technology, Wakamatsu-ku, Kitakyushu-shi, Fukuoka 808-0196 Japan (e-mail: ravi1989tripathi@gmail.com).

T. Hanamoto is with the Department of Life Science and Systems Engineering, Kyushu Institute of Technology, Wakamatsu-ku, Kitakyushu-shi, Fukuoka 808-0196 Japan (e-mail: hanamoto@life.kyutech.ac.jp).

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2020 at 05:20:42 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2999267, IEEE Journal of Emerging and Selected Topics in Power Electronics

reconfigurable system is used to implement MPC for a back-to-back converter. The FPGA implementation of MPC is performed using a model-based design (MBD) through MATLAB/Simulink in [20], [21] and FPGA code was obtained using hardware description language (HDL) coder functionality of Simulink. In [22], a high-speed FCS-MPC implementation is presented with a control algorithm coded in C++ using PROTOIP toolbox and HDL code was generated through Xilinx Vivado HLS. Xilinx system generator (XSG) as a digital simulator adopting the MBD platform was used to implement an FPGA-based predictive current controller in [23]–[26]. XSG is an integrated platform with MATLAB/Simulink and facilitates the HDL code generation for the developed system.

In addition to the implementation approach, resource utilization is another key aspect of FPGA-based system control implementation. The system control implementation based on FPGA resource utilization is compared in [22] and [25] for Xilinx FPGA. In [22], the prediction horizon is considered for the comparison of FPGA resource utilization. In [25], semi-parallel and fully serial implementation approaches for MPC are adopted and resource utilization is compared for the same. The look-up tables (LUTs), flip-flops and DSP slices are the main digital logic indices considered for resource utilization in the aforementioned papers.

There is a possibility to adopt different implementation strategies for an FPGA implementation to enhance resource utilization. In this way, the MBD approach can be advantageous by allowing different implementation strategies under the same environment and to compare the FPGA resource utilization. Moreover, the MBD approach provides better system visualization and easy debugging that is appealing for rapid controller prototyping.

This paper presents analytical implementation strategies using an MBD approach considering the reduction in the FPGA resource utilization for fully parallel implementation of the FCS-MPC algorithm. The predictive model coefficient computed using load parameters is one of the significant factors that influence system performance. Furthermore, an adaptive predictive model is proposed to compensate for the effect of change in load parameters by using simple mathematical relation of the three-phase inverter output voltage, dc voltage supply, and modulation index. The unity and non-unity modulation index is also included to investigate the proposed adaptive model design considering system performance.

A comparative analysis of FPGA resource utilization using analytical implementation strategies is considered corresponding to the predictive model coefficient of the FCS-MPC for a three-phase VSI system with RL load. Moreover, resource utilization for controller implementation using stationary $\alpha\beta$ and rotating dq reference frames is used for in-depth analysis. In general, the cost function is formulated based on stationary $\alpha\beta$ coordinates [27], [28]. Considering current control with the FCS-MPC, the sinusoidal future current references need be predicted accurately using an extrapolation method, however, the extrapolated reference causes unwanted oscillations that influence transient response [7], [9], [27]. Considering the above issue, the cost function can be designed using a rotating dq frame that does not require the extrapolation method due to continuous current references [27]. However, the comparative analysis based on design constraints and performance indices is required to be addressed for two frames. Furthermore, the computational burden, as well as the FPGA resource requirement corresponding to each coordinate system, is crucial for practical system implementation.

The FPGA-based implementation of the FCS-MPC is performed using the MBD approach provided by the XSG integrated to MATLAB/Simulink. The system performance is analyzed based on the implementation of the FCS-MPC using optimum FPGA resource utilization in $\alpha\beta$ and dq frames for the adaptive predictive model (unity and non-unity modulation index). The dynamic behavior and steady-state error are considered for analysis of system performance. Furthermore, a constraint of change in switching state [2], [29] is considered to analyze the performance of average switching frequency reduction and the current harmonics for the adaptive predictive model.

The paper is organized as follows: Section II describes the algorithm of the FCS-MPC for both stationary $\alpha\beta$ as well as rotating dq frames considering the discrete-time mathematical model of the three-phase VSI system. In Section III, the model-based design and development of the controller in the digital simulator are explained. Section IV covers the different implementation strategies considered for the development of the FCS-MPC algorithm and the comparative FPGA resource utilization. The simulation and experimental results are presented with detailed discussion in Section V. Finally, appropriate conclusions are drawn in Section VII.

II. Finite Control Set-MPC

The power circuit of a three-phase VSI with *RL* load is shown in Fig. 1 where R is the load resistance and L is the load inductance. The switching states of the converter are determined by the switching signals S_a, S_b, and S_c as shown in Table I. Considering all possible switching combinations of the switching signals, eight switching states $\boldsymbol{S}$ ($\boldsymbol{S}_0$–$\boldsymbol{S}_7$) and hence, eight voltage vectors $\boldsymbol{v}_i$ ($\boldsymbol{v}_0$–$\boldsymbol{v}_7$) are obtained as shown in Table II.

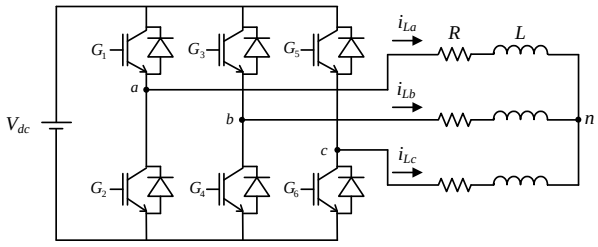


Fig. 1. Power circuit of three-phase VSI with RL load.

TABLE I
Gating Signals of Inverter

Leg 'a', S_a	Leg 'b', S_b	Leg 'c', S_c
G_1 ON, 1 G_2 OFF, 0	G_3 ON, 1 G_4 OFF, 0	G_5 ON, 1 G_6 OFF, 0
G_1 OFF, 0 G_2 ON, 1	G_3 OFF, 0 G_4 ON, 1	G_5 OFF, 0 G_6 ON, 1

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2020 at 05:20:42 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2999267, IEEE Journal of Emerging and Selected Topics in Power Electronics

TABLE II
VOLTAGE VECTORS AND SWITCHING STATES

Switching states	Voltage vectors
$\boldsymbol{S} = [S_a \ S_b \ S_c]$	$\boldsymbol{v_i} = [v_{i\alpha} \ v_{i\beta}]$
$\boldsymbol{S}_0 = [0 \ 0 \ 0]$	$\boldsymbol{v}_0 = [0, 0]$
$\boldsymbol{S}_1 = [1 \ 0 \ 0]$	$\boldsymbol{v}_1 = [2V_{dc}/3, 0]$
$\boldsymbol{S}_2 = [1 \ 1 \ 0]$	$\boldsymbol{v}_2 = [V_{dc}/3, \sqrt{3}V_{dc}/3]$
$\boldsymbol{S}_3 = [0 \ 1 \ 0]$	$\boldsymbol{v}_3 = [-V_{dc}/3, \sqrt{3}\ V_{dc}/3]$
$\boldsymbol{S}_4 = [0 \ 1 \ 1]$	$\boldsymbol{v}_4 = [-2V_{dc}/3, 0]$
$\boldsymbol{S}_5 = [0 \ 0 \ 1]$	$\boldsymbol{v}_5 = [-V_{dc}/3, -\sqrt{3}V_{dc}/3]$
$\boldsymbol{S}_6 = [1 \ 0 \ 1]$	$\boldsymbol{v}_6 = [V_{dc}/3, -\sqrt{3}V_{dc}/3]$
$\boldsymbol{S}_7 = [1 \ 1 \ 1]$	$\boldsymbol{v}_7 = [0, 0]$

The FCS-MPC based on the current control objective uses a discrete-time model of the load current for the formulation of the control algorithm and deals only with a limited number of possible switching states of the power converter. The discrete-time model can be obtained from the dynamics of the load current. The dynamics of the load current $\boldsymbol{i}_L$ (i_{La}, i_{Lb}, i_{Lc}) can be described by the vector differential equation as below:

$$\boldsymbol{v_i} = R\boldsymbol{i_L} + L\frac{d\boldsymbol{i_L}}{dt} \tag{1}$$

A. Discrete-time Predictive Model

The discrete-time model of the load current dynamics (1) for a sampling time T_S represents the predictive model. It will be used to predict the future value of load currents considering each voltage vector listed in Table II. The discrete-time model can be obtained by an approximation of load current derivative $d\boldsymbol{i}_L/dt$ using the forward Euler discretization method given as:

$$\frac{d\boldsymbol{i_L}}{dt} \approx \frac{\boldsymbol{i_L}(k+1) - \boldsymbol{i_L}(k)}{T_S} \tag{2}$$

After substituting (2) into (1), an expression of predicted future load current at sampling interval k+1 is obtained for each of the seven different voltage vectors $\boldsymbol{v_i}(k)$ in $\alpha\beta$ frame as:

$$\begin{aligned} i_{L\alpha}^{p}(k+1) &= k_1 i_{L\alpha}(k) + k_2 v_{i\alpha}(k) \\ i_{L\beta}^{p}(k+1) &= k_1 i_{L\beta}(k) + k_2 v_{i\beta}(k) \end{aligned} \tag{3}$$

$$\text{where } k_1 = \left(1 - \frac{RT_S}{L}\right),\ k_2 = \frac{T_S}{L}$$

$i_{L\alpha}(k)$, $i_{L\beta}(k)$ and $i^p{}_{L\alpha}(k+1)$, $i^p{}_{L\beta}(k+1)$ denote the measured load currents at instant k and the predicted future load currents at instant k+1, respectively. The coordinate transformations from abc to $\alpha\beta$ are computed using mathematical relation defined by Clarke transformation as

$$\begin{bmatrix} x_\alpha \\ x_\beta \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{3} & -1/\sqrt{3} \end{bmatrix} \begin{bmatrix} x_a \\ x_b \\ x_c \end{bmatrix} \tag{4}$$

where x can be any quantity voltage or current.

The voltages and currents in dq coordinates can be transformed from $\alpha\beta$ components using Park transformation relation given as:

$$\begin{bmatrix} x_d \\ x_q \end{bmatrix} = \begin{bmatrix} \cos\theta^* & \sin\theta^* \\ -\sin\theta^* & \cos\theta^* \end{bmatrix} \begin{bmatrix} x_\alpha \\ x_\beta \end{bmatrix} \tag{5}$$

where θ^* is the reference phase angle for the coordinate conversion. The predicted load currents in rotating dq-frame can also be obtained using forward Euler discretization considering feed-forward terms for the decoupling of the d and q components of the current as given below:

$$\begin{aligned} i_{Ld}^{p}(k+1) &= k_1 i_{Ld}(k) + k_2 \left\{ v_{id}(k) + k_3 i_{Lq}(k) \right\} \\ i_{Lq}^{p}(k+1) &= k_1 i_{Lq}(k) + k_2 \left\{ v_{iq}(k) - k_3 i_{Ld}(k) \right\} \end{aligned} \tag{6}$$

$$\text{where } k_1 = \left(1 - \frac{RT_S}{L}\right),\ k_2 = \frac{T_S}{L},\ k_3 = \omega^* L$$

ω^* is an angular frequency of the current reference and calculated as $\omega^* = d\theta^*/dt$.

B. Adaptive Predictive Model

The predictive model depends on the coefficients k_1, k_2, k_3, for current prediction and computed using the load parameters (R, L). Any change in load parameters disturbs the accuracy of the predictive model, which ultimately affects the performance of the FCS-MPC. The value of a coefficient can be updated to compensate for the effect of change in load parameters. In this paper, an approach to update the coefficient k_1 (adaptive k_1) corresponding to the change in load resistance R is considered by assuming unchanged load inductance L.

According to the phase voltage and line current relationship, the load impedance Z can be calculated as given below:

$$Z = \frac{v_{ph,rms}}{i_{L,rms}} \tag{7}$$

where $v_{ph,rms}$ is the rms values of voltage in any phase and $i_{l,rms}$ is the rms value of the load current. The $v_{ph,rms}$ can be simply obtained using the relation between the phase voltage and a dc supply to an inverter given as:

$$V_{dc} = \frac{2\sqrt{2}v_{ph,rms}}{m} \tag{8}$$

where m is a modulation index.

By substituting the value of $v_{ph,rms}$ from (8) into (7), the value of Z can be updated as follows

$$Z = \frac{mV_{dc}}{2\sqrt{2}i_{L,rms}} \tag{9}$$

The load impedance Z contains resistance (R) and inductive reactance (X_L) as given below

$$Z = \sqrt{R^2 + (X_L)^2},\ \text{where } X_L = \omega L \tag{10}$$

R can be computed by putting the value of Z from (9) to (10) and represented as:

$$R = \sqrt{\left(\frac{mV_{dc}}{2\sqrt{2}i_{L,rms}}\right)^2 - (X_L)^2} \tag{11}$$

R can be updated by (11) according to the change in load current. Further, the coefficient k_1 used for current prediction in (3) and (6) can be automatically updated by putting the value of

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2020 at 05:20:42 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2999267, IEEE Journal of Emerging and Selected Topics in Power Electronics

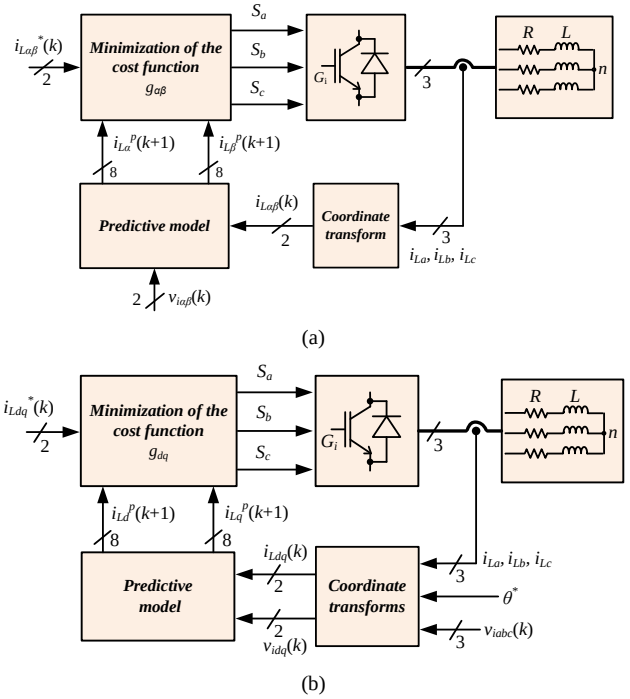


Fig. 2. Schematic diagram of load current control using the FCS-MPC in (a) $\alpha\beta$-frame and, (b) dq-frame.

the updated R from (11) and an expression for adaptive k_1 can be written as follows:

$$k_1 = \left(1 - \frac{T_S}{L}\sqrt{\left(\frac{mV_{dc}}{2\sqrt{2}i_{L,rms}}\right)^2 - \left(X_L\right)^2}\right) \tag{12}$$

Considering any system parameters, $\left(\frac{mV_{dc}}{2\sqrt{2}i_{L,rms}}\right)^2 \gg \left(X_L\right)^2$.

Thus, the expression for adaptive k_1 is approximated as:

$$k_1 \approx \left(1 - \frac{mV_{dc}T_S}{2\sqrt{2}Li_{L,rms}}\right) \tag{13}$$

The values of V_{dc}, T_S, and L are known and $i_{l,rms}$ can be obtained from the reference load current.

C. Cost Function

To select an optimum switching state according to the optimum voltage vector in each sampling and to minimize the current error for the next sampling interval, selection criteria is required to be defined considering control objectives.

The simple selection criteria i.e cost function can be expressed in orthogonal $\alpha\beta$ coordinates as follows:

$$g_{\alpha\beta} = \left|i^*_{L\alpha}\left(k+1\right) - i^p_{L\alpha}\left(k+1\right)\right| + \left|i^*_{L\beta}\left(k+1\right) - i^p_{L\beta}\left(k+1\right)\right| \tag{14}$$

where $i^*_{L\alpha}$ and $i^*_{L\beta}$ are the real and imaginary components of the reference current. The future reference current value required by (14) has to be predicted using the Lagrange extrapolation method discussed in [5]. However, for sufficient small T_S, a simple approximation $i^*_L(k+1) \approx i^*_L(k)$ can be used and no extrapolation is required. The same approximation is considered in this paper.

The cost function for the control of dq coordinate is formulated as:

$$g_{dq} = \left|i^*_{Ld}\left(k+1\right) - i^p_{Ld}\left(k+1\right)\right| + \left|i^*_{Lq}\left(k+1\right) - i^p_{Lq}\left(k+1\right)\right| \tag{15}$$

where i^*_{Ld} and i^*_{Lq} are the reference currents for d and q components respectively. The block diagram of the load current control using the FCS-MPC with stationary $\alpha\beta$ and rotating dq reference frames are depicted in Fig. 2(a) and Fig. 2(b) respectively.

The cost functions expressed in (14) and (15) are based on the primary control objective that is current control. Furthermore, the cost function with a constraint of change in switching state of the inverter is considered to compare the system performance for the FCS-MPC. The inclusion of this constraint reduces the number of commutations of power switches that leads to a reduction in switching frequency. Present switching state $\boldsymbol{S}(k)$ is compared with previously applied switching state $\boldsymbol{S}(k\text{-}1)$ to ensure fewer switching commutations for each phase. The cost function with the constraint is formulated as

$$g = g_{\alpha\beta/dq} + \lambda \cdot p \tag{16}$$

where λ is the weighting factor and p is the number of switching commutations defined as

$$p = \sum_x \left|S_x\left(k\right) - S_x\left(k-1\right)\right|, \quad x = a,b,c \tag{17}$$

III. Model-Based Design of the FCS-MPC

A low-level digital code (VHDL/Verilog) of the desired control system is required to program an FPGA device for the FPGA-based system implementation. The developments in the system implementation facilitate the generation of low-level code for FPGAs using the functionality of a model-based design approach [20], [23]. The XSG, an MBD platform, provides the Xilinx blockset for the MATLAB/Simulink to develop a virtual model of digital circuits specifically for Xilinx FPGAs. Xilinx blockset consists of the fundamental mathematical blocks, a block for the inclusion of low-level VHDL code (Blackbox) and a block for the inclusion of MATLAB code (MCode). During implementation, the Blackbox is used for the analog-to-digital converter (ADC) and digital-to-analog (DAC) interfaces.

The modeling of the FCS-MPC is performed based on the schematic diagram shown in Fig. 3. This is a fully parallel implementation approach and the control algorithm is divided

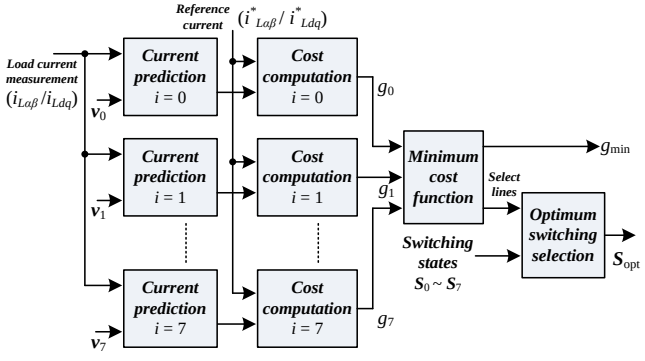


Fig. 3. Block diagram for parallel implementation of the FCS-MPC algorithm using model-based design approach.

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2020 at 05:20:42 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2999267, IEEE Journal of Emerging and Selected Topics in Power Electronics

into three steps: prediction of future load currents, computation of cost functions, and selection of optimum switching state. The cost function without constraint is opted to demonstrate the modeling steps and FPGA resource utilization based on implementation strategies considering equal resource requirements by the constraint during the FCS-MPC implementation with $\alpha\beta$ as well as dq-frames.

The prediction of future load currents is performed firstly using (3) and (6) corresponding to the FCS-MPC in $\alpha\beta$-frame and dq-frame respectively. Further, the cost functions are computed using (14) and (15) for both the frames described in section II. The future current predictions and cost functions are computed on the parallel basis for each inverter voltage vectors ($\boldsymbol{v}_0$–$\boldsymbol{v}_7$) defined for corresponding switching states (S_a, S_b, S_c) as given in Table II.

The modeling for the prediction and cost computation can be performed using only fundamental mathematical blocks based on (3), (14) and (6), (15) for $\alpha\beta$ and dq-frame respectively. However, there is a possibility to perform the same operation using other blocks of the Xilinx blockset. The following strategies are considered for the modeling of the FCS-MPC algorithm for both $\alpha\beta$ and dq frame:

- Fundamental mathematical blocks for the current prediction and cost function computation
- MCode block for the current prediction and cost function computation
- Approximation for the coefficient k_1
- Adaptive k_1

IV. FPGA Implementation Strategy and Resource Utilization

A. Implementation Strategy

The implementation strategy is crucial for the time required for computation and resource utilization. There is a possibility to reduce the required FPGA resources significantly by using an appropriate implementation strategy that may create the possibility of system implementation using smaller FPGAs. The following are the objectives of analytical implementation strategies mentioned in Section III:

- FPGA resource: minimization of resource utilization (LUTs, flip-flops, DSP slices) required for fully parallel implementation.
- Usability: the implementation strategy with minimum resource utilization can be used for semi-parallel/sequential implementation.
- Trade-off: maintaining appropriate control quality and computational speed with reduced resource utilization.

Besides, the optimum strategy can be beneficial to develop various controllers for the FPGA-based system implementation. The implementation strategies are described as follows:

1) Fundamental mathematical blocks

In this strategy, the future current predictions and cost function computations are performed using fundamental mathematical XSG blocks: AddSub, CMult, Mult, Absolute. AddSub block performs the addition and subtraction for two inputs, CMult block performs the function of gain multiplier for a single input, Mult block performs the function of multiplication for two inputs and Absolute block performs the function of absolute value computation for real and imaginary quantities in (14) and (15). The coefficients (k_1, k_2, k_3) used for future current predictions are provided by Xilinx Constant blocks with an appropriate number of bits and binary point settings for fixed-point number representation.

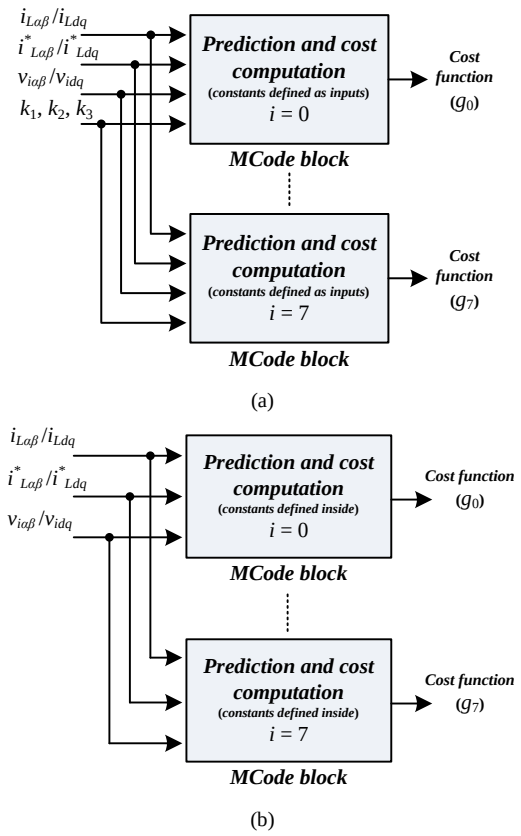


Fig. 4. Development methodologies for the FCS-MPC in XSG using MCode block (a) case-I, (b) case-II.

2) MCode blocks

In this strategy, the future current predictions and cost function computations are performed together through a single MCode block corresponding to each voltage vector using MATLAB code. This strategy is further subdivided into two cases based on the coefficients (k_1, k_2, k_3):

Case-I: coefficients defined as inputs to MCode blocks

Case-II: coefficients defined inside MCode blocks

The development methodologies are presented in Fig. 4 for case-I (Fig. 4(a)) and case-II (Fig. 4(b)). An additional case (case-III) is considered for the resource utilization in the current prediction using adaptive k_1 where other coefficients defined inside the MCode block like case-II.

The MATLAB code for the prediction (Ipred_real and Ipred_img) and cost computation using current errors (x and y) in $\alpha\beta$-frame for case-I and case-II is as follows:

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2020 at 05:20:42 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2999267, IEEE Journal of Emerging and Selected Topics in Power Electronics

MATLAB code: **Case-I**

```
function [cost] = mpc_case-I (Im_real, Im_img,
Iref_real, Iref_img, k1, k2, v_real, v_img)
Ipred_real = k1*Im_real + k2*v_real;
Ipred_img = k1*Im_img + k2*v_img;
x = Iref_real – Ipred_real;
y = Iref_img – Ipred_img;
    if (x<0)
        mod_x = – x;
    else
        mod_x = x;
    end
    if (y<0)
        mod_y = – y;
    else
        mod_y = y;
    end
cost = mod_x + mod_y;
end
```

MATLAB code: **Case-II**

```
function [cost] = mpc_case-II (Im_real, Im_img,
Iref_real, Iref_img, v_real, v_img)
k1 = xfix({xlSigned, 10, 8, xlRound, xlWrap}, 0.95);
k2 = xfix({xlSigned, 13, 13, xlRound, xlWrap},
0.005);
Ipred_real = k1*Im_real + k2*v_real;
Ipred_img = k1*Im_img + k2*v_img;
x = Iref_real – Ipred_real;
y = Iref_img – Ipred_img;
    if (x<0)
        mod_x = – x;
    else
        mod_x = x;
    end
    if (y<0)
        mod_y = – y;
    else
        mod_y = y;
    end
cost = mod_x + mod_y;
end
```

The values of k_1, k_2 are defined inside the MATLAB code in case-II using a command xfix that is used for the fixed-point number representation. Similarly, the MATLAB codes for the case-I and case-II of the FCS-MPC in *dq*-frame are designed using the coefficients k_1, k_2, and k_3 according to the codes designed for the FCS-MPC in $\alpha\beta$-frame mentioned above. An additional case (case-III) is considered for the resource utilization in the current prediction using adaptive k_1 where other coefficients defined inside the MCode block like case-II.

B. Resource Utilization

Every FPGA has a set of programmable logic, I/O, and memory resources. The FPGA resources consist of configurable logic blocks (CLBs) which are the main constituent to implement code on the FPGA. A CLB element comprises of a pair of slices and a slice contains LUTs and flip-flops [30], [31]. According to the family of the FPGA chip, the number of LUTs and flip-flops in a single slice are different. For instance, there are four LUTs and eight flip-flops in a slice considering Zynq-7000 series FPGAs [31]. In addition, there is a major component of the FPGA resources, called DSP slices that are mainly used to implement digital signal processing functions. The main constituents of a DSP slice are signed multipliers, adder/accumulator, arithmetic logic unit (ALU) logic functions and many more [32], [33].

The FPGA resource utilization considering implementation strategies mentioned in the previous section are listed in Table III and IV for the FCS-MPC in $\alpha\beta$-frame and *dq*-frame respectively. To select the optimum design strategy, only two logic indices LUTs and DSP slices need attention because the number of Flip-flops are the same considering $\alpha\beta$-frame as well as *dq*-frames.

1) Coefficient k_1 (k_1 = 0.95, k_1 = 1)

The coefficient k_1 is used in future current prediction using (3) and (6) as mentioned in Section II. The value of k_1 depends on the load parameters (R, L) and sampling time (T_S). The exact value of k_1 is 0.95 based on the parameters (R = 10 Ω, L = 10 mH and T_S = 50 μs) considered for simulation and experimental validations. In the case of k_1 = 0.95, the number of bits 10 and binary points 8 are specified and provides the coefficient value of 0.9492187 that utilizes significant FPGA resources. The approximation (k_1 = 1) can be opted considering the reduction in FPGA resource utilization due to the omission of the required bits and the multiplier used to multiply the exact value of k_1. Consequently, the required bits and the multipliers directly reduce the LUTs and DSP slices, respectively. However, the

TABLE III
FPGA RESOURCE UTILIZATION FOR THE DEVELOPMENT OF THE FCS-MPC IN $\alpha\beta$ - FRAME

Logic utilization indices	*Fundamental Mathematical Blocks*				*MCode Block*					
					Case-I		*Case-II*		*Case-III*	
	k_1 = 0.95	k_1 = 1	*Adaptive* k_1		k_1 = 0.95	k_1 = 1	k_1 = 0.95	k_1 = 1	*Adaptive* k_1	
			m = 1	m = 0.95, 0.9					m = 1	m = 0.95, 0.9
LUTs (53200)	5732 (11%)	4799 (9%)	6546 (12%)	7580 (14%)	4364 (8%)	3804 (7%)	4896 (9%)	3794 (7%)	5488 (10%)	6792 (14%)
Flip-flops (106400)	549 (1%)	549 (1%)	1551 (2%)	1820 (2%)	549 (1%)	549 (1%)	549 (1%)	549 (1%)	1551 (2%)	1820 (2%)
DSP slices (220)	21 (10%)	19 (9%)	26 (12%)	28 (13%)	25 (11%)	19 (9%)	19 (9%)	19 (9%)	27 (12%)	29 (13%)

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2020 at 05:20:42 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2999267, IEEE Journal of Emerging and Selected Topics in Power Electronics

TABLE IV
FPGA RESOURCE UTILIZATION FOR THE DEVELOPMENT OF THE FCS-MPC IN *dq* - FRAME

Logic utilization indices	*Fundamental Mathematical Blocks*				*MCode Block*					
					Case-I		*Case-II*		*Case-III*	
	$k_1 = 0.95$	$k_1 = 1$	*Adaptive* k_1		$k_1 = 0.95$	$k_1 = 1$	$k_1 = 0.95$	$k_1 = 1$	*Adaptive* k_1	
			$m = 1$	$m = 0.95, 0.9$					$m = 1$	$m = 0.95, 0.9$
LUTs (53200)	5017 (9%)	5004 (9%)	5489 (10%)	5784 (11%)	8534 (16%)	8304 (16%)	9967 (19%)	8072 (15%)	9133 (17%)	10646 (20%)
Flip-flops (106400)	804 (1%)	804 (1%)	1806 (2%)	2845 (3%)	804 (1%)	804 (1%)	804 (1%)	804 (1%)	1806 (2%)	2845 (3%)
DSP slices (220)	66 (30%)	62 (28%)	66 (30%)	70 (32%)	86 (39%)	74 (34%)	24 (11%)	24 (11%)	32 (15%)	36 (16%)

performance with $k_1 = 1$ is required to be analyzed through simulation as well as the experiment.

In case-II of MCode block, the equal number of DSP slices are required for both the subcases $k_1 = 0.95$ and 1, however, the required number of LUTs for $k_1 = 0.95$ is slightly higher as required for $k_1 = 1$. In the case of the FPGA implementation using controller development in $\alpha\beta$-frame, both cases (case-I and II) of MCode with $k_1 = 1$ are the optimum solutions, however, considering better controller performance based on current prediction using an exact value of k_1 with the cost of slightly increased LUTs, case-II of MCode block with $k_1 = 0.95$ can be the optimum choice.

On the other hand, for the controller development in *dq*-frame, case-II of MCode block is an optimum strategy due to the least percentage utilization of DSP slices among all the individual resource indices. Although the LUTs requirement has reduced significantly in the case of the development using fundamental mathematical blocks, the required DSP slices have increased approximately 3 times. Furthermore, the FPGA resource utilization for the development of the FCS-MPC in *dq*-frame is even higher as compared to $\alpha\beta$-frame due to the additional coordinate transformation ($\alpha\beta$ to *dq*) for voltage vector including measured current, reference phase angle θ^* generation using CORDIC SINCOS and an extra effort for computation of feed-forward term used for decoupling.

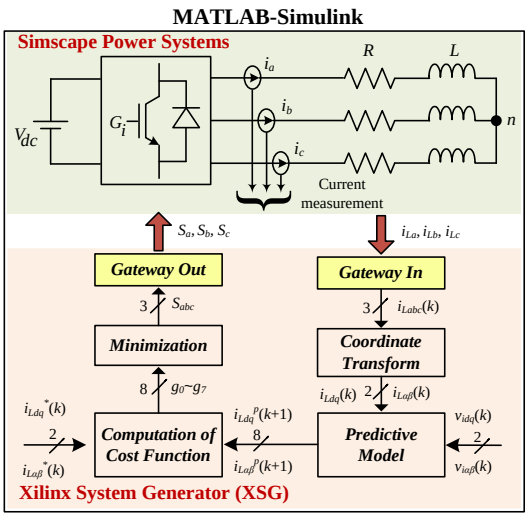


Fig. 5. Block diagram representation of methodology for simulation.

2) Adaptive k_1 (m = 1, 0.95, 0.9)

The adaptive k_1 is used for future current prediction using (3), (6) and (13) as mentioned in Section II. The adaptive k_1 depends on the modulation index (m) for a fixed dc supply. The unity value (m = 1) and non-unity value (m = 0.95, 0.9) are opted for system performance analysis as well as resource utilization.

In the case of the controller development using adaptive k_1, resource utilization is significantly higher than the optimum design cases (MCode block: case-II) in both $\alpha\beta$-frame as well as *dq*-frames. However, considering the updated predictive model corresponding to the load changes, the controller performance needs to be investigated for the controller development using adaptive k_1. Moreover, a trade-off between controller performance and the FPGA resource utilization can be a better choice to decide an appropriate implementation strategy corresponding to the system application.

An optimum implementation strategy corresponding to the minimum resource utilization is further utilized for the prediction and cost computation in both the frames. After the completion of current prediction and cost computation, the minimum cost function g_{min} and an optimum switching state $\boldsymbol{S}_{opt}$ are selected using a pipelining method for all three implementation strategies. The logic for the selection of g_{min} and $\boldsymbol{S}_{opt}$ are developed in XSG using a multiplexer (Mux block) and comparator (Relational block) based on the XSG diagram depicted in [26].

V. RESULTS AND DISCUSSION

Simulation methodology in Fig. 5 is represented using a block diagram to express the development approach of the controller. The power circuit of three-phase VSI is developed in MATLAB/Simulink using simscape toolbox, whereas the controller is developed in the digital environment of XSG using Xilinx toolbox. Due to the non-identical platforms, a signal from simscape to XSG is interfaced using Xilinx "Gateway In" block and a signal from XSG to simscape is interfaced using Xilinx "Gateway Out" block.

The FPGA architecture of the FCS-MPC with the schematic block diagram representation of the experimental setup in Fig. 6 is developed for a complete understanding of the FPGA-based implementation of the system. An adequate latency is provided to blocks used to predict load current and to compute cost

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2020 at 05:20:42 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2999267, IEEE Journal of Emerging and Selected Topics in Power Electronics

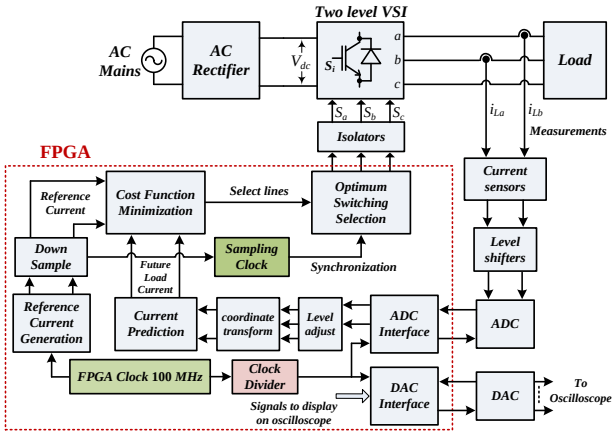


Fig. 6. Schematic diagram of FPGA architecture for experimental setup.

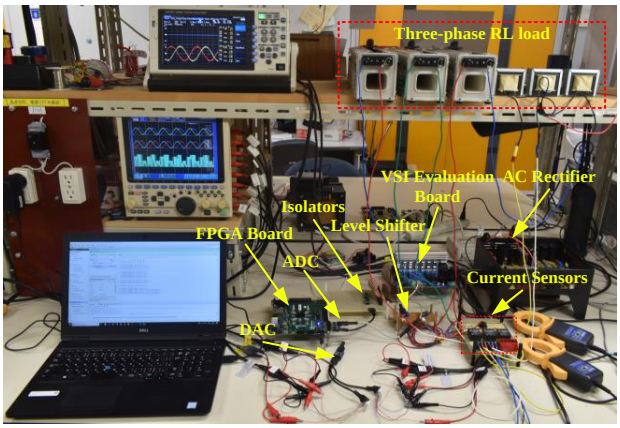


Fig. 7. Laboratory prototype of the experimental setup.

function for proper completion of the computational task by the corresponding block. To ensure the selection of the optimum switching state in the given sampling interval, the sampling clock is generated as a synchronizer. The sampling clock is used as an enabling signal and given to the multiplexer that is used for the selection of optimum switching state. Before sending switching signals directly to the VSI, the isolator circuit is used for electrical isolation of the FPGA board and inverter board considering safety precautions of the FPGA board (Zynq-XC7020). The FPGA code was generated through the modeled controller in XSG and programmed using dedicated software (Xilinx Vivado Design Suite) for the real-time implementation of the VSI system with the FCS-MPC. The laboratory prototype of the experimental setup is shown in Fig. 7. The experimental system components used for the development of laboratory prototype are listed in Table V.

TABLE V
DISCRIPTION OF COMPONENTS SPECIFICATIONS FOR THE EXPERIMENTAL SYSTEM

S.No.	Components	Specifications
1.	Three-phase VSI	STEVAL-IHM023V3, 1 kW
2.	DC supply	THDSHVMTRPFCKIT
3.	Current transducer	LA 25-NP
4.	Op-amp IC for level shifter	LM385N
5.	Isolator IC	ADuM3440
6.	ADC	Digilent Pmod AD1
7.	DAC	Digilent Pmod DA4
8.	FPGA board	Zedboard Zynq Evaluation and Development Kit

TABLE VI
PARAMETERS FOR SYSTEM IMPLEMENTATION

Parameter	Symbol	Value
Supply dc voltage	V_{dc}	145 V
Load resistance	R	10 Ω
Load inductance	L	10 mH
System frequency	f	50 Hz
Sampling frequency	f_S	20 kHz
FPGA clock frequency	f_{CLK}	100 MHz

The performance of the FCS-MPC in $\alpha\beta$ and dq frames was experimentally evaluated and the experimental data are obtained through HIOKI 8855 Memory Hicorder. Further, the data are plotted with the help of MATLAB plotting tool for demonstration and analysis of experimental results. The system parameters considered for the simulation and the experimental validation are listed in Table VI. The validation and analysis of FPGA-based FCS-MPC are performed using the implementation strategy having the optimum resource utilization.

A. *Effect of Predictive Model Coefficient*

The performance of the FCS-MPC in $\alpha\beta$ and dq frames is analyzed for the predictive model with exact (k_1 = 0.95), approximated (k_1 = 1) as well as the adaptive values of coefficient k_1 considering unity (m = 1) and non-unity (m = 0.95, 0.9) values of 'm'. Moreover, the simulation and experiment were carried out considering the cost function without constraint for reference

TABLE VII
PERFORMANCE OF THE FCS-MPC IN $\alpha\beta$ - FRAME

Indices		i_L = 4 A					i_L = 2.5 A				
		k_1 = 0.95	k_1 = 1	adaptive k_1			k_1 = 0.95	k_1 = 1	adaptive k_1		
				m=1	m=0.95	m=0.9			m=1	m=0.95	m=0.9
Simulation	THD (%)	3.58	3.64	3.46	3.66	3.68	5.27	5.37	5.12	5.22	5.36
	Average f_{sw} (Hz)	3637	3663	3677	3640	3713	3147	3057	3123	3023	3053
Experiment	THD (%)	4.9	5.2	4.6	4.7	4.9	8.5	9.1	7.5	7.7	8.1
	Average f_{sw} (Hz)	2556	2489	2445	2430	2533	2325	2150	2350	2378	2393

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2020 at 05:20:42 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2999267, IEEE Journal of Emerging and Selected Topics in Power Electronics

TABLE VIII
PERFORMANCE OF THE FCS-MPC IN *dq* - FRAME

Indices		i_L = 4 A					i_L = 2.5 A				
		k_1 = 0.95	k_1 = 1	adaptive k_1			k_1 = 0.95	k_1 = 1	adaptive k_1		
				m=1	m=0.95	m=0.9			m=1	m=0.95	m=0.9
Simulation	THD (%)	3.85	4.07	3.75	3.79	4.02	6.09	6.54	5.65	5.74	5.87
	Average f_{sw} (Hz)	3650	3607	3713	3693	3727	3027	2967	3127	3113	3027
Experiment	THD (%)	4.8	5.3	4.7	5.0	5.2	8.4	9.2	7.9	8.2	8.5
	Average f_{sw} (Hz)	2586	2524	2465	2438	2457	2365	2120	2410	2388	2347

currents with low load ($\boldsymbol{i}_L$ = 2.5A) and high load ($\boldsymbol{i}_L$ = 4A) conditions. To investigate the quality of the load current and the switching stress on the power devices, the harmonic content in the load current corresponding to the average switching frequency (f_{sw}) is also analyzed considering both the frames. The results obtained during the simulation as well as an experiment are listed in Table VII and VIII for the FCS-MPC in *αβ* and *dq* frames respectively. The harmonic content in the load current during the experiment was measured through HIOKI 3390 power analyzer and current clamp sensor.

1) Coefficient k_1 (k_1 = 0.95, k_1 = 1)

During the simulation, there is a slight difference in percentage THD for the FCS-MPC in *αβ* and *dq* frames corresponding to the coefficient k_1, however, the performance is almost similar during the experiment for both frames. The percentage THD is less in the case of high load with the cost of slightly higher switching stress as compared to low load during the simulation. However, a significantly high current THDs are observed during the experiment for the low load as compared to high load condition. The percentage THD of the load currents in the experiment is higher as compared to the simulation that is because of the simplified model of the VSI and the load used for the simulation study.

2) Adaptive k_1 (m = 1, 0.95, 0.9)

In this case, there is a slight difference in percentage THD for the simulation as well as the experiment in *αβ* and *dq* frames. This difference is more pronounced for low load current as well as non-unity values of '*m*'. In addition, percentage THD for adaptive k_1 with m = 1 is slightly better as compared to m = 0.95 and 0.9. Compared to the other cases of coefficient k_1 (k_1 = 1, 0.95), the controller performance is slightly better for current prediction with adaptive k_1 in both low load as well as high load conditions for unity value of '*m*', however, implementation with adaptive k_1 requires highest FPGA resources amongst the cases of k_1.

B. Dynamic Performance (k_1 = 0.95, 1; *adaptive* k_1: m = 1, 0.95, 0.9)

To investigate the dynamic performance of the FCS-MPC for the three-phase VSI system, the step change in the reference current is considered for current tracking during the transients. The comparative analysis for the system implementation using k_1 = 0.95, 1 and adaptive k_1 (m = 1, 0.95, 0.9) is performed considering simulation as well as experimental results. The tracking performance of the load current is demonstrated for the change in reference current from 2.5A to 4A at instant 0.062s and from 4A to 2.5A at instant 0.14s.

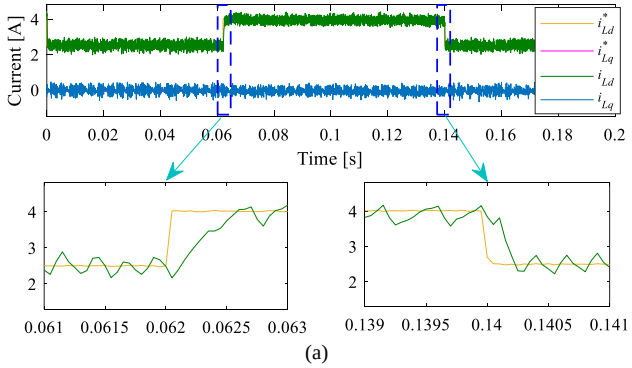
The dynamic performance of FCS-MPC in *αβ* and *dq* frames is represented in Table IX and X considering the settling time during transients for the cases of the coefficient k_1. The settling time for a step change at 0.062s ($\boldsymbol{i}_L$ = 2.5A to 4A) is significantly lower in *dq*-frame as compared to *αβ*-frame. Moreover, the proposed adaptive k_1 case demonstrates the fastest settling time amongst in simulation as well as experiment. The settling time for a step change at 0.14s ($\boldsymbol{i}_L$ = 4A to 2.5A) is slightly better for

TABLE IX
DYNAMIC RESPONSE: SETTLING TIME FOR STEP CHANGE AT 0.062S

Coefficient k_1	FCS-MPC in *αβ*-frame		FCS-MPC in *dq*-frame	
	Sim.	Exp.	Sim.	Exp.
k_1 = 0.95	375 µs	430 µs	350 µs	400 µs
k_1 = 1	400 µs	520 µs	270 µs	350 µs
k_1 adaptive: m = 1	300 µs	370 µs	250 µs	300 µs
k_1 adaptive: m = 0.95	350 µs	410 µs	270 µs	320 µs
k_1 adaptive: m = 0.9	350 µs	430 µs	280 µs	330 µs

TABLE X
DYNAMIC RESPONSE: SETTLING TIME FOR STEP CHANGE AT 0.14S

Coefficient k_1	FCS-MPC in *αβ*-frame		FCS-MPC in *dq*-frame	
	Sim.	Exp.	Sim.	Exp.
k_1 = 0.95	150 µs	250 µs	150 µs	230 µs
k_1 = 1	200 µs	310 µs	170 µs	250 µs
k_1 adaptive: m = 1	150 µs	240 µs	150 µs	220 µs
k_1 adaptive: m = 0.95	200 µs	300 µs	150 µs	240 µs
k_1 adaptive: m = 0.9	250 µs	320 µs	200 µs	300 µs



(a)

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2020 at 05:20:42 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2999267, IEEE Journal of Emerging and Selected Topics in Power Electronics

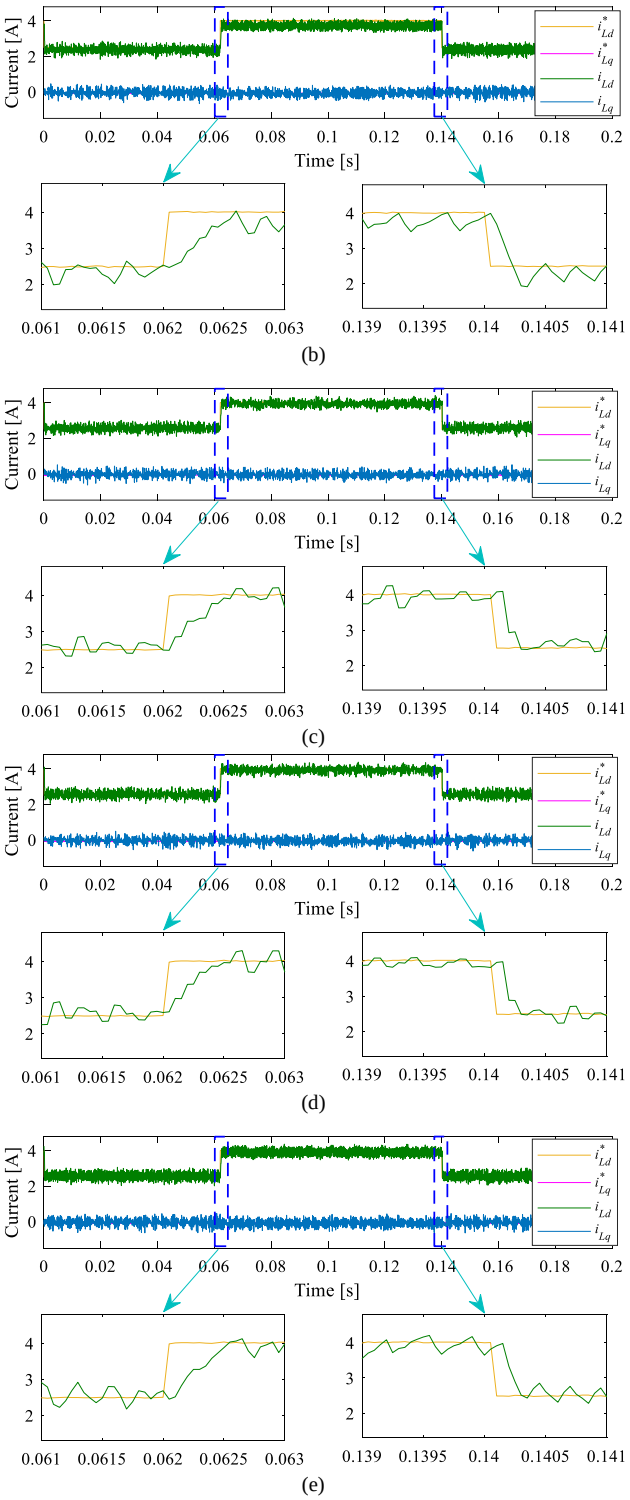


Fig. 8. Experimental results of load current dynamic performance for the FCS-MPC in *dq*-frame (a) coefficient k_1: k_1 = 0.95 (fixed), (b) coefficient k_1: k_1 = 1 (approximated), (c) adaptive k_1: m = 1, (d) adaptive k_1: m = 0.95, and (e) adaptive k_1: m = 0.9.

dq-frame as well as adaptive k_1 case as compared to *αβ*-frame and other cases of k_1.

The current tracking in Fig. 8 are depicted for the controller performance in *dq*-frame for coefficient k_1 = 0.95, 1 as well as adaptive k_1 with m = 1, 0.95, 0.9. In *dq*-frame, the step changes are applied in the *d*-axis component of the reference current, keeping *q*-axis reference current zero. The time axis is enlarged to demonstrate the current tracking during the transient caused by step changes in reference current. The transient response is almost the same for the current prediction using k_1 = 0.95 and 1. In addition, the transient response for adaptive k_1 with m = 1 is almost the same for m = 0.95 and 0.9 and slightly better as compared to k_1 = 0.95 and 1. However, the current tracking performance of adaptive k_1 with m = 1 is best amongst the aforementioned cases.

C. *Steady-state Error*

The steady-state error (SSE) is analyzed through the simulation as well as experiment for load current tracking considering adaptive k_1 (m = 1, 0.95, 0.9) and coefficient k_1 (k_1 = 1, 0.95). The comparative SSE analysis is performed for the $\boldsymbol{i}_L$ = 4A and $\boldsymbol{i}_L$ = 2.5A in Fig. 9 and Fig. 10 respectively using bar graphs for the FCS-MPC in both frames. The comparative analysis demonstrates a similar trend for the $\boldsymbol{i}_L$ = 4A and $\boldsymbol{i}_L$ = 2.5A in simulation as well as experimental results. The SSE

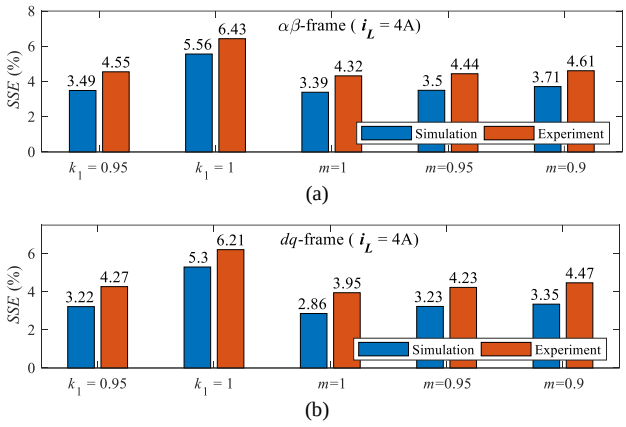


Fig. 9. Steady-state error performance for $\boldsymbol{i}_L$ = 4A with the current prediction using coefficient k_1: k_1 = 0.95 (fixed), k_1 = 1 (approximated) and adaptive k_1: m = 1, 0.95, 0.9 through the simulation as well as experiment for the FCS-MPC in (a) *αβ*-frame, and (b) *dq*-frame.

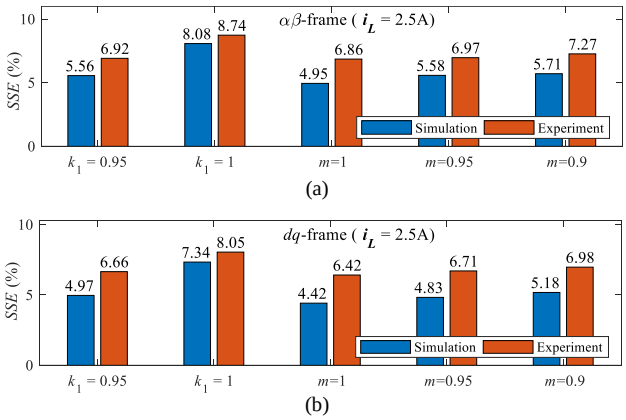


Fig. 10. Steady-state error performance for $\boldsymbol{i}_L$ = 2.5A with the current prediction using coefficient k_1: k_1 = 0.95 (fixed), k_1 = 1 (approximated) and adaptive k_1: m = 1, 0.95, 0.9 through the simulation as well as experiment for the FCS-MPC in (a) *αβ*-frame, and (b) *dq*-frame.

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2020 at 05:20:42 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2999267, IEEE Journal of Emerging and Selected Topics in Power Electronics

performance for adaptive k_1 with m = 1 is better as compared to other cases and it is more pronounced for the low load current $\boldsymbol{i}_L$ = 2.5A. Moreover, the SSE performance in *dq*-frame is better as compared to *αβ*-frame for the considered cases of coefficient k_1 as well as both load current conditions.

D. Switching State Constraint

The system performance is investigated for the cost function with the switching state constraint (16) as discussed in section II during the simulation and experiment. The constraint of change in switching state provides a reduction in the switching frequency corresponding to the variation in weighting factor λ. The system performance with varying λ is examined for average switching frequency f_{sw} as well as percentage THD in the load current considering the case of best system performance (adaptive k_1: m = 1) and poor performance (coefficient k_1 = 1).

1) For condition $\boldsymbol{i}_L$ = 4A (adaptive k_1: m = 1 and k_1 = 1)

The performance of the FCS-MPC in *αβ*, as well as *dq* frames for $\boldsymbol{i}_L$ = 4A, are depicted in Fig. 11 and Fig. 12 for the simulation and experiment results respectively. The average f_{sw} decreases corresponding to an increase in λ, consequently, the harmonic

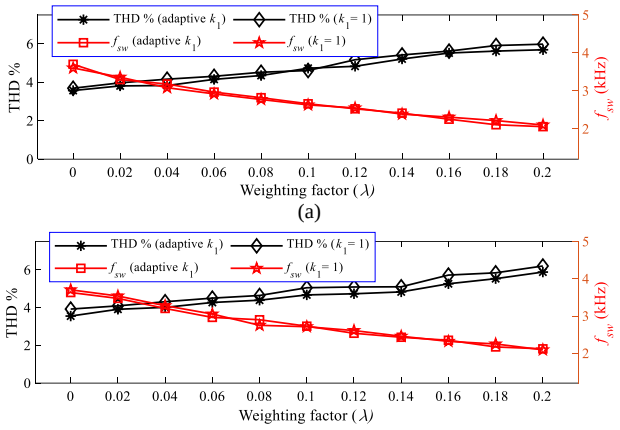


Fig. 11. Average switching frequency (f_{sw}) and percentage THD in load current for $\boldsymbol{i}_L$ = 4A during simulation with the FCS-MPC in (a) *αβ*-frame, and (b) *dq*-frame.

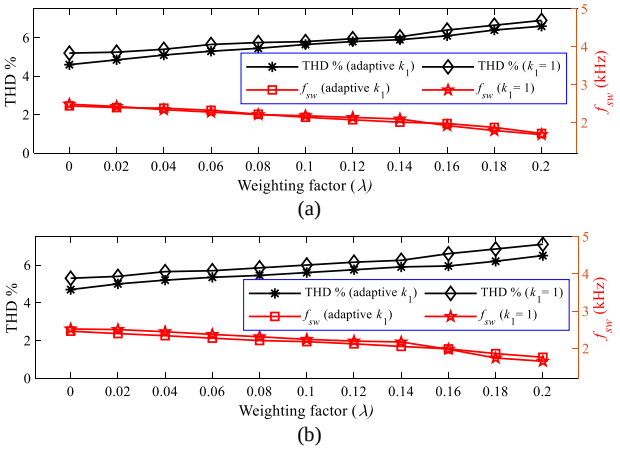


Fig. 12. Average switching frequency (f_{sw}) and percentage THD in load current for $\boldsymbol{i}_L$ = 4A during experiment with the FCS-MPC in (a) *αβ*-frame, and (b) *dq*-frame.

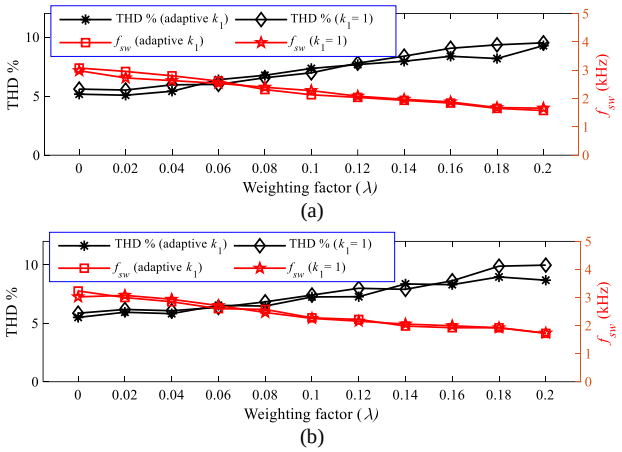


Fig. 13. Average switching frequency (f_{sw}) and percentage THD in load current for $\boldsymbol{i}_L$ = 2.5 A during simulation with the FCS-MPC in (a) *αβ*-frame, and (b) *dq*-frame.

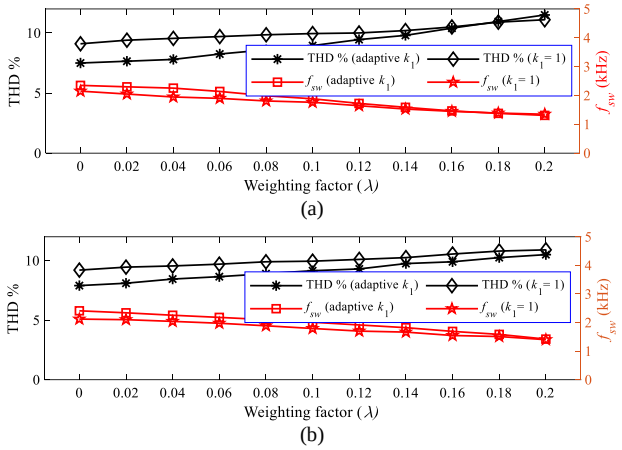


Fig. 14. Average switching frequency (f_{sw}) and percentage THD in load current for $\boldsymbol{i}_L$ = 2.5 A during experiment with the FCS-MPC in (a) *αβ*-frame, and (b) *dq*-frame.

content in the load current increases. A slightly higher percentage THD is observed corresponding to λ for k_1 = 1 as compared to the adaptive k_1 during the simulation as well as the experiment. However, THD is almost the same especially for the value of λ = 0.1 to 0.14 in the simulation as well as the experiment for both the frames.

2) For condition $\boldsymbol{i}_L$ = 2.5A: Adaptive k_1 (m = 1) and k_1 = 1:

Similar to the previous case, the performance in both the frames for $\boldsymbol{i}_L$ = 2.5A is presented in Fig. 13 and Fig. 14 for simulation and experiment respectively corresponding to the same range of λ.

In the case of the simulation, the gap between THD lines for adaptive k_1 and k_1 = 1 is negligible corresponding to variation in λ. However, the gap between THD lines decreases with the increase in λ for both the frames in the case of the experiment. And, similar to the case of $\boldsymbol{i}_L$ = 4A, the THD is almost the same for value λ = 0.1 to 0.18. This trend of THD is authenticated through the average f_{sw} corresponding to change in λ and THD increases with the reduction in average f_{sw}.

The controller performance is significantly changed

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2020 at 05:20:42 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2999267, IEEE Journal of Emerging and Selected Topics in Power Electronics

TABLE XI
FPGA RESOURCE UTILIZATION FOR THE DEVELOPMENT OF THE FCS-MPC WITH SWITCHING CONSTRAINT

Logic utilization indices	FCS-MPC in $\alpha\beta$-frame		FCS-MPC in dq-frame	
	adaptive k_1: $m = 1$	$k_1 = 1$	adaptive k_1: $m = 1$	$k_1 = 1$
LUTs (53200)	5763 (11%)	4074 (8%)	9405 (18%)	8352 (16%)
Flip-flops (106400)	2617 (2%)	1005 (1%)	2262 (2%)	1260 (1%)
DSP slices (220)	38 (17%)	26 (12%)	43 (20%)	31 (14%)

considering the percentage THD in load current for current prediction using fixed, adaptive and approximated value of coefficient k_1 as summarized in Table VII and Table VIII. In addition, the performance is dependent on constraint for switching frequency reduction that ultimately results in higher percentage THD. However, the switching constraint almost compensates for the effect of current prediction using the approximated value of coefficient k_1 for the higher value of λ.

3) FPGA resources

The FPGA resource utilization in the case of the FCS-MPC development with the change in switching state constraint using adaptive k_1 (MCode block: case-III) as well as $k_1 = 1$ (MCode block: case-II) is also listed in Table XI for both the frames. The FPGA resource utilization is minimum for the implementation strategy using MCode block: case-II, $k_1 = 1$. So, a trade-off between the FPGA resource utilization and controller performance can be opted considering the requirement of the system application using an appropriate implementation strategy.

VI. CONCLUSION

This paper presented analytical implementation strategies for the development of FCS-MPC using a fully parallel approach considering conventional current prediction and proposed adaptive current prediction for $\alpha\beta$ and dq coordinates. A significant reduction is achieved (around 65% in DSP Slices) in overall FPGA resource requirement in dq coordinate for system development through MCode block case-II (coefficients defined inside the MATLAB code) of Xilinx blockset. Especially, system development through an approximated value of a coefficient ($k_1 = 1$) provides a further reduction in FPGA resources. Furthermore, resources are optimized using case-II for adaptive current prediction (around 50% reduced DSP Slices) for unity and non-unity modulation index. However, a trade-off between the controller performance and optimum resource utilization is vital for the cost-performance ratio.

The FCS-MPC performance for percentage THD, dynamic behavior, and SSE is investigated for conventional current prediction and proposed adaptive current prediction in $\alpha\beta$ and dq frames using implementation strategy with optimum FPGA resources. The system performance for adaptive k_1 with $m = 1$ is significantly better for all performance indices under high as well as low current operation in both the frames. The performance for coefficient k_1 ($k_1 = 1$) is poor amongst all the case of current predictions and become quite poorer for the low current operation. Furthermore, change in switching state constraint considered for the in-depth analysis lead to an interesting conclusion that percentage THD performance for coefficient $k_1 = 1$ and adaptive k_1 with $m = 1$ become almost similar corresponding to the change in weighting factor under high as well as low current operation. However, for this case, the SSE performance is not investigated and considered as possible future work. This overall system implementation and performance analysis will be vital for achieving a trade-off between controller performance and FPGA resource utilization. The effect of inductive load parameter variation on controller performance is a possible future work considering constraint as well for FPGA-based FCS-MPC implementation using optimum resources.

REFERENCES

[1] S. Kouro, P. Cortes, R. Vargas, U. Ammann and J. Rodriguez, "Model Predictive Control – a Simple and Powerful Method to Control Power Converters," *IEEE Trans. Ind. Electron.*, vol. 56, no. 6, pp. 1826–1838, Jun. 2009.

[2] J. Rodríguez and P. Cortes, Predictive Control of Power Converters and Electrical Drives, 1st ed. New York: *Wiley-IEEE Press*, 2012.

[3] S. Kouro, M. A. Perez, J. Rodriguez, A. M. Llor and H. A. Young, "Model Predictive Control: MPC's Role in the Evolution of Power Electronics," *IEEE Ind. Electron. Mag*., vol. 8, no. 1, pp. 16–31, Mar. 2014.

[4] S. Vazquez, J. I. Leon, L. G. Franquelo, J. Rodriguez, H. A. Young, A. Marquez, and P. Zanchetta, "Model predictive control: A review of its applications in power electronics," *IEEE Ind. Electron. Mag*., vol. 9, no. 4, pp. 8–21, Dec. 2015.

[5] J. Rodríguez, J. Pontt, C. Silva, P. Correa, P. Lezana, P. Cortés, and U. Ammann, "Predictive current control of a voltage source inverter," *IEEE Trans. Ind. Electron.*, vol. 54, no. 1, pp. 495–503, Feb. 2007.

[6] P. Nikhil, K. Sonam, M. Monika, and S. Wagh, "Finite control set model predictive control for two level inverter with fixed switching frequency," in *2018 SICE International Symposium on Control Systems* (SICE ISCS), Mar 2018, pp. 74–81.

[7] F. Donoso, A. Mora, R. Cárdenas, A. Angulo, D. Sáez, and M. Rivera, "Finite-set model-predictive control strategies for a 3L-NPC inverter operating with fixed switching frequency," *IEEE Trans. Ind. Electron.*, vol. 65, pp. 3954–3965, May 2018.

[8] P. Cortes, J. Rodriguez, P. Antoniewicz, and M. Kazmierkowski, "Direct power control of an AFE using predictive control," *IEEE Trans. Power Electron.*, vol. 23, no. 5, pp. 2516–2523, 2008.

[9] J. Rodriguez, M. P. Kazmierkowski, J. R. Espinoza, P. Zanchetta, H. Abu-Rub, H. A. Young, and C. A. Rojas, "State of the art of finite control set model predictive control in power electronics," *IEEE Trans. Ind. Informat.*, vol. 9, no. 2, pp. 1003–1016, May 2013.

[10] C. Qi, X. Chen, P. Tu, and P. Wang, "Cell-by-cell-based finite-control-set model predictive control for a single-phase cascaded H-bridge rectifier," *IEEE Trans. Power Electron.*, vol. 33, no. 2, pp. 1654–1665, Feb. 2018.

[11] M. Rivera, C. A. Rojas, J. Rodriguez, P. Wheeler, B. Wu, and J. R. Espinoza, "Predictive current control with input filter resonance mitigation for a direct matrix converter," *IEEE Trans. Power Electron.*, vol. 26, no. 10, pp. 2794–2803, Oct. 2011.

[12] L. Wang, H.Dan, Y. Zhao, Q. Zhu, T. Peng, Y. Sun and P. Wheeler "A finite control set model predictive control method for matrix converter with zero common-mode voltage," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 6, no. 1, pp. 327–338, Mar. 2018.

[13] P. Cortes, G. Ortiz, J. Yuz, J. Rodríguez, S. Vazquez, and L. Franquelo, "Model predictive control of an inverter with output LC filter for UPS applications," *IEEE Trans. Ind. Electron.*, vol. 56, no. 6, pp. 1875–1883, 2009.

[14] S. Vazquez, J. Rodriguez, M. Rivera, L. G. Franquelo and M. Norambuena, "Model Predictive Control for Power Converters and Drives: Advances and Trends," *IEEE Trans. Ind. Electron.*, vol. 64, no. 2, pp. 935–947, Feb. 2017.

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2020 at 05:20:42 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JESTPE.2020.2999267, IEEE Journal of Emerging and Selected Topics in Power Electronics

[15] P. Cortes, M. P. Kazmierkowski, R. M. Kennel, D. E. Quevedo, and J. Rodriguez, "Predictive control in power electronics and drives," *IEEE Trans. Ind. Electron.*, vol. 55, no. 12, pp. 4312–4324, Dec. 2008.
[16] P. Cortes, J. Rodriguez, C. Silva, and A. Flores, "Delay compensation in model predictive current control of a three-phase inverter," *IEEE Trans. Ind. Electron.*, vol. 59, pp. 1323-1325, 2012.
[17] Y. Yang, H. Wen, and D. Li, "A fast and fixed switching frequency model predictive control with delay compensation for three-phase inverters," *IEEE Access*, 5:17904–17913, 2017.
[18] O. Gulbudak, E. Santi, "FPGA-based model predictive controller for direct matrix converter," *IEEE Trans. Ind. Electron.*, vol. 63, no. 7, pp. 4560–4570, Jul. 2016.
[19] Z. Zhang, F. Wang, T. Sun, J. Rodríguez, guez, and R. Kennel, "FPGA-Based Experimental Investigation of a Quasi-Centralized Model Predictive Control for Back-to-Back Converters," *IEEE Trans. Power Electron.*, vol. 31, pp. 662-674, Jan. 2016.
[20] Z. Ma, S. Saeidi, and R. Kennel, "FPGA implementation of model predictive control with constant switching frequency for PMSM drives," *IEEE Trans. Ind. Informat.*, vol. 10, no. 4, pp. 2055–2063, Nov. 2014.
[21] S. Wendel, A. Dietz, and R. Kennel, "FPGA based finite-set model predictive current control for small PMSM drives with efficient resource streaming," in *2017 IEEE International Symposium on Predictive Control of Electrical Drives and Power Electronics* (PRECEDE), pp. 66–71, Sept. 2017.
[22] B. Stellato, T. Geyer, and P. J. Goulart, "High-Speed Finite Control Set Model Predictive Control for Power Electronics," *IEEE Trans. Power Electron.*, vol. 32, pp. 4007-4020, 2017.
[23] V. K. Singh, R. Tripathi and T. Hanamoto, "FPGA-Based Implementation of Finite Set-MPC for a VSI System using XSG-Based Modeling," *Energies*, vol. 13, no. 1, 260, January 2020.
[24] T. J. Vyncke, S. Thielemans, and J. A. Melkebeek, "Finite-set model based predictive control for flying-capacitor converters: Cost function design and efficient FPGA implementation," *IEEE Trans. Ind. Informat.*, vol. 9, no. 2, pp. 1113–1121, May 2013.
[25] P. M. Sanchez, O. Machado, E. J. B. Peña, F. J. Rodriguez, and F. J. Meca, "FPGA-based implementation of a predictive current controller for power converters," *IEEE Trans. Ind. Informat.*, vol. 9, no. 3, pp. 1312-1321, Aug. 2013.
[26] V. K. Singh, R. N. Tripathi and T. Hanamoto, "Xilinx System Generator Based Modeling of Finite State MPC", *9th International Power Electronics Conference* (IPEC 2018-ECCE Asia), May 20-24, 2018.
[27] H. A. Young, M. A. Perez, J. Rodriguez, and H. Abu-Rub, "Assessing finite-control-set model predictive control: A comparison with a linear current controller in two-level voltage source inverters," *IEEE Ind. Electron. Mag.*, vol. 8, no. 1, pp. 44–52, Mar. 2014.
[28] C. Xia, T. Liu, T. Shi and Z. Song, "A Simplified Finite-Control-Set Model-Predictive Control for Power Converters," *IEEE Trans. Ind. Informat.*, vol. 10, no. 2, pp. 991–1002, May. 2014.
[29] R. Vargas, P. Cortes, U. Ammann, J. Rodriguez, and J. Pontt, "Predictive control of a three-phase neutral point-clamped inverter," *IEEE Trans. Ind. Electron.*, vol. 54, no. 5, pp. 2697–2705, Oct. 2007.
[30] E. Monmasson, L. Idkhajine, M. W. Naouar, "FPGA-based controllers", *IEEE Ind. Electron. Mag.*, vol. 5, no. 1, pp. 14–26, April 2011.
[31] Zynq-7000 SoC: *Technical Reference Manual*, July 1, 2018.
[32] Zynq-7000 SoC: *Data Sheet Overview*, July 2, 2018.
[33] 7 Series DSP48E1 Slice: *User Guide*, March 27, 2018.



Vijay Kumar Singh received his B.Tech. degree from the Department of Electronics and Communication Engineering, United College of Engineering and Research, Allahabad (under UPTU Lucknow), Uttar Pradesh, India, in 2008 and M. Tech. degree (Power Electronics and ASIC Design) from the Department of Electrical Engineering, Motilal Nehru National Institute of Technology, Allahabad, Uttar Pradesh, India in 2012. He completed Ph.D. degree course from the Department of Life Science and Systems Engineering, Kyushu Institute of Technology, Kitakyushu, Fukuoka, Japan in 2020. His research interests include model predictive control, design, modeling, and development of control techniques for power electronic converters.



Ravi Nath Tripathi received his B.Tech (Electrical and Electronics Engineering) degree in 2011 from Graphic Era Institute of Technology (GEIT), India and M.Tech (Power Systems) degree in 2013 from Delhi Technological University (formerly Delhi College of Engineering), Delhi, India. He completed doctoral degree course in 2017 from Kyushu Institute of Technology. He has worked as Post-Doctoral Researcher on NEDO project from 2017-2020 and currently working as a researcher in Next Generation Power Electronics Research Center, Kyushu Institute of Technology, Japan. His research interest is design and development of Gate Drive Control for power semiconductor devices, design and development of controller for power electronic converters, Grid-integration of Solar PV system and motor drives.



Tsuyoshi Hanamoto was born in 1961. He received an M.S degree from Kyushu Institute of Technology, Japan in 1986 and then joined Kobe Works of Kobe Steel, Ltd. In 1990, he joined the Center for Cooperative Research of Kyushu Institute of Technology. From 1997 to 2000 he was with Department of Electrical Engineering and since April 2000, he has been with the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, where he is presently a Professor. His research interests include motor control, power electronics, and FPGA-based system development.

2168-6777 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 06,2020 at 05:20:42 UTC from IEEE Xplore. Restrictions apply.



Taylor & Francis
Taylor & Francis Group

Transactions of the IMF
The International Journal of Surface Engineering and Coatings

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ytim20

Modelling and optimisation of magnetic abrasive finishing process based on a non-orthogonal array with ANN-GA approach

S. Ahmad , R. M. Singari & R.S. Mishra

To cite this article: S. Ahmad , R. M. Singari & R.S. Mishra (2020) Modelling and optimisation of magnetic abrasive finishing process based on a non-orthogonal array with ANN-GA approach, Transactions of the IMF, 98:4, 186-198, DOI: 10.1080/00202967.2020.1776966

To link to this article: https://doi.org/10.1080/00202967.2020.1776966

Published online: 17 Jul 2020.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ytim20

TRANSACTIONS OF THE IMF
2020, VOL. 98, NO. 4, 186–198
https://doi.org/10.1080/00202967.2020.1776966

Check for updates

Modelling and optimisation of magnetic abrasive finishing process based on a non-orthogonal array with ANN-GA approach

S. Ahmad, R. M. Singari and R.S. Mishra

Department of Mechanical, Production & Industrial, Automobile Engineering, Delhi Technological University, New Delhi, India

ABSTRACT

Magnetic abrasive finishing (MAF) is an advanced precise finishing method that achieves micro-level to nano-level surface roughness. In industries, MAF is highly recommended where zero or negligible post-process surface defects are an obligatory requirement. In the same context, process optimisation is essential for making it commercially viable. This study presents an artificial neural network and genetic algorithm (ANN-GA), a robust modelling and optimisation tool (applicable to any sort of data set orthogonal array design or non-orthogonal array design) that is applied to scrutinise and improve the performance of the magnetic abrasive finishing of stainless steel SS302. In addition, the results from ANN-GA modelling and optimisation have been compared with conclusions drawn from conventionally used Taguchi-ANOVA analysis. An L_{27} non-orthogonal array design has been opted for as per machining set-up restriction. Abrasive size, voltage, machining gap, and rotational speed were the design variables considered in the present research work. It was found that the parametric design used in this study provides a straightforward, methodical, and proficient method of modelling and optimisation of change of surface roughness or finishing behaviour during the MAF process. Modelling and optimisation done with ANN-GA show a maximum value of $(\Delta R_a)_{max}$ equal to 0.256 µm, which is 7% better than the result obtained from Taguchi-ANOVA analysis.

ARTICLE HISTORY
Received 13 February 2020
Accepted 22 April 2020

KEYWORDS
Magnetic abrasive finishing; non-orthogonal array; artificial neural network; genetic algorithm; ANOVA

1. Introduction

Expeditious advancements of latest materials with properties like high hot hardness, toughness, fragility, strength-weight ratio, etc. are becoming worthy of attention. These significant properties of such advanced materials have made them more demandable in various industrial and research applications. The processing of these materials by traditional manufacturing processes always has been a challenge. Also, the quality of surface finish always has a consequential effect on the functionality of any mechanical component.[1,2] For example, frictional power losses, wear, tear, and component fatigue life have a great dependency on surface quality and surface integrity.[3]

In recent times, a few non-traditional finishing and polishing processes have evolved. These processes are capable enough to process advanced material such as NiMoC alloys, silicon wafers, carbide, and CBN cutting tools, etc. as well, as they can obtain micro to a nano-level range of surface finish.[4–6] Abrasive flow machining (AFM), chemo-mechanical polishing (CMP), magnetic float polishing (MFP), magneto-rheological finishing (MRF), and magnetic abrasive finishing (MAF), etc., are some such non-traditional finishing and polishing processes.[3,7–10]

In the MAF process, application of gentle forces by using a flexible magnetic abrasives brush (FMAB) for removing surface unevenness has been a really innovative idea. During MAF, the magnetic forces allow the FMAB to shear off the material from the surface in the form of micro-chips.[11–15] Many authors have studied various control factors, and optimisation techniques to accomplish the best surface finish on different advanced materials using the MAF manufacturing processes. However, these optimisation techniques lack intelligent iterative optimisation methodology for better performance of the process.

Yamaguchi *et al.*[6] used diamond paste for the finishing of an uncoated carbide tool (UCT) surface using MAF to get better tool wear characteristics. Finishing time, pole tip feed, spindle speed, machining gap, steel grits mean-diameter, diamond abrasive, lubricant, were chosen as control factors. It was reported that roughness less than 25 nm could be achieved. Tool life of MAF processed UCT had been doubled compared to UCT not processed with MAF, when it was used in turning of titanium (Grade 5) rods. It is vital for coated tool finishing to use different sizes of the abrasive particle simultaneously adjusting magnetic force for desired finishing and less material removal.[16] Amineh *et al.*[1] utilised MAF to remove the recast layer formed by wire electro-discharge machining (WEDM). The MAF was able to shear off the recast layer generated during WEDM without any micro-cracking. It was found that increasing abrasive particles causes higher recast layer removal, and improved the surface roughness. In another study, Amnieh *et al.*[17] examined the effectiveness of finishing parameters such as tool rotational speed, abrasive mesh number, and finishing time. Jiang Guo *et al.*[2] found that the MAF was an effective finishing operation for a flat workpiece of polyamide 12 (PA12) 3-D printed parts under constant magnetic field. The process leaves no visible chemical change within the 10 µm subsurface level. Jiang Guo *et al.*[18] investigated the tribological performance of RSA-905 after finishing the surface by magnetic field-assisted finishing (MFAF). Magnetic abrasive (SiC, Al_2O_3, Carbonyl Iron Powder) was used for finishing and achieved the desired finish without any post-process subsurface defect. Tribological performance was improved, and

CONTACT R. M. Singari ranganath@dce.ac.in Department of Mechanical, Production & Industrial, Automobile Engineering, Delhi Technological University, Shahbad Daulatpur, Bawana Road, New Delhi 110042, India

© 2020 Institute of Materials Finishing Published by Taylor & Francis on behalf of the Institute

surface residual stresses were released. However, a small decrement in hardness was noticed. Yuewu Gao *et al.*[19] studied the influence of various process variables on the polish quality and compared the surface roughness of Cu-27400 and SS316 flat workpiece. An Al_2O_3 based atomised magnetic abrasive tool was selected for processing under the control of a permanent magnet (Nd-Fe-B). Shadab Ahmad *et al.*[20] examined the consequence of sintered magnetic abrasive (SMA) on the MAF process. The process parameters considered were voltage, machining gap, and rotational speed. There were three experimental arrays designed according to the L_9 orthogonal array. The Taguchi-ANOVA method was used to find the optimised surface roughness change and corresponding values of the parameter with their quantitative effect. The most influential parameter was voltage followed by machining gap and rotational speed, respectively and it was found that the abrasive size has a significant role in the MAF process.[20,21]

The application of different optimisation methods like Taguchi, Response Surface, Fuzzy Logic, Genetic Algorithm, Moth Flame Algorithm, and Particle Swarm Algorithm for the effective utilisation of the MAF process, is increasing.[22,23] The influence of voltage, machining gap, abrasive size, and rotational speed on surface roughness during the MAF process was studied by employing various methods by many researchers. Table 1 shows the finishing parameters and optimisation techniques employed in the research.

The above literature survey shows that many researchers have analysed the effects of process parameters and optimised the process with different techniques. In addition, very few researchers have employed an iterative optimisation method for more effective utilisation of the MAF process. Artificial neural networks (ANNs) have solved various parametric based problems in manufacturing processes, like pattern finding and learning in a set of data, developing the ability of pattern classification, forecasting, clustering, or modelling.[31–33] Genetic algorithms (GAs) can perform a global search, optimisation, and generalisation.[34,35] The neural network solves the problems where modelling by mapping input patterns to output results is important. However, to perform computational optimisation, tools like GAs are used and need a single stature of merit as a fitness function.

This current research study explains the development and application of an artificial neural network and genetic algorithm (ANN-GA) as a modelling and optimisation tool that is used to improve the performance of the magnetic abrasive finishing of SS302. In addition, the results from ANN-GA modelling and optimisation have been compared with conclusions drawn from conventionally used Taguchi-ANOVA analysis. The Taguchi-ANOVA method is used to analyse the data for comparison.

2. Magnetic abrasive finishing (MAF)

Magnetic Abrasive Finishing (MAF) is an advanced finishing operation to produce an efficient and excellent surface finish quality on a cylindrical surface, flat surface, and externally and internally in tubes.[36,37] This process has the flexibility to work without the restriction of workpiece material characteristics (physical and mechanical properties). The workpiece material could be electrically conductive or non-conductive, magnetic or non-magnetic.[5,16,36] Thus, there is no restriction if material is made of ferrous or non-ferrous. Conventionally

Table 1. Finishing parameters and optimisation techniques in previous studies.

S. No.	Authors	Workpiece/ Abrasive	Type of Magnetic Source	Voltage or current supply	Machining Gap /mm	Rotational Speed /rpm	Abrasive Size	Surface Roughness Range or effect	Remarks
1.	T. C. Kanish *et al.*[22]	SS316L/ SiC	–	15–20 V	1.5-2	270–540	400–1200 mesh	35–50%	Taguchi optimisation and fuzzy modelling used to predict SR
2.	Prateek Kala *et al.*[24]	Copper/Alumina	Permanent	-	1.5–2.5	200–400	800 mesh	–	Taguchi optimisation and analysis of finishing force and torque involved in DDMAF done
3.	K. B. Judal *et al.*[25]	AISI-304/ SiC	DC Supply	0.5-2.5 A	1–3	150–1200	10–180 µm	(0.28 µm, 0.32 µm) – 0.039 µm	RSM optimisation techniques were used. Rotational speed and electrolytic current lead to significant improvement in MRR and SR.
4.	Yi-Hsun Lee *et al.*[26]	SUS304/ SiC	Permanent	-	1–2	100–1000	1000–8000 mesh	0.13-0.03 µm	Taguchi optimisation done and 77% improvement found in SR
5.	Amer A Moosa[27]	Brass/Quartz	DC supply	1.5–3.5 A	1–2	175–525	150 µm	–	ANFIS used for mapping and optimisation of experimental results
6.	Prateek Kala *et al.*[28]	Copper Alloy C70600/ Al_2O_3	DC supply	70–100 V	2	112–224	600–1200 mesh	62% (MAP), 76% (UAMAP)	Taguchi method applied for process optimisation, and this study concludes that UAMAP is an improved technique over MAP.
7.	A. Cheng Wang *et al.*[29]	SKD11/ AL_2O_3, SiC	-	1–3 A	–	700–1300	4000–8000 mesh	0.029 µm	Claimed that MFGA is better than MAF, because of gel-based MA. Optimisation of the process done by the Taguchi method.
8.	Nitesh Sihag *et al.*[30]	Tungsten Alloy/ Al_2O_3	Permanent	–	2	150	1200 mesh	0.34–17 µm	86% improvement in SR with chemical oxidation RSM and ANOVA was used to design the experiments and analyse the results

SR (Surface Roughness); MRR (Material Removal Rate); DDMAF (Double Disk Magnetic Abrasive Finishing); RSM (Response Surface Methodology); ANFIS (Adaptive Neuro-Fuzzy Inference System); MAP (Magnetic Abrasive Polishing); UAMAP (Ultrasonic Assisted Magnetic Abrasive Polishing); MFGA (Magnetic Finishing with Gel Abrasives); MA (Magnetic Abrasives); MAF (Magnetic Abrasive Finishing); ANOVA (Analysis Of Variance).

in MAF, magnetic abrasives particles (MAPs) composed of ferromagnetic material with abrasive particles are employed as cutting tools to shear-off the material from the target surface. But this material removal is greatly affected by finishing pressure, which works on the influence of a magnetic field across the machining gap.[38] The schematics of the MAF process are illustrated in Figure 1 representing the equipotential lines across which magnetic abrasive is arranged on the work surface to remove material.[20,38] The rotational motion causes the cutting force parallel to the targeted surface, and abrasive penetration inside the workpiece is due to the normal force generated by the magnetic field intensity.[39] MAPs become arranged with each other magnetically along a line of magnetic force.[9] A Fishbone diagram representing all the factors that affect the MAF process is shown in Figure 2.

3. Materials and method

3.1 Material and tools

Stainless steel grade 202 was the workpiece material for the study of this particular finishing operation. Sintered magnetic abrasive (SMA) was working as many tiny cutting tools with random cutting edges in simultaneous action. These abrasives performed shearing operation to remove the material from the targeted surface in the form of microchips.[40] The constituent of SMA was alumina powder, and cast iron powder sintered together in 40/60 by weight ratio, its preparation techniques have been discussed in the literature.[20,21] An electromagnet (32 mm pole diameter) mounted on a lathe tool post connected to a DC Supply produces a static magnetic field, as shown in Figure 3. Various independent control factors influence the performance of the MAF process, including abrasive size, voltage, machining gap, and rotational speed.

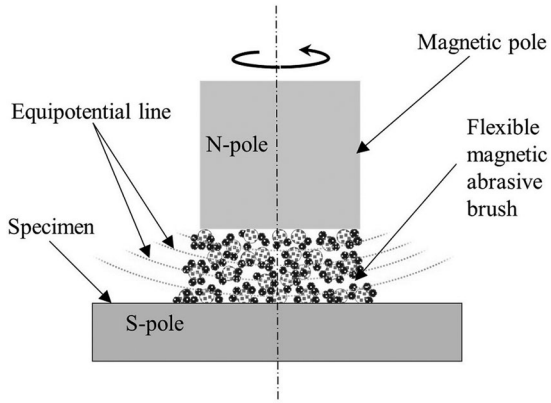


Figure 1. Schematic of the MAF process.

3.2 Experimental detail

The Taguchi technique is a prevailing tool for designing, arrangement, and collection of experimental data in a controlled way. It also tells us the effects of process variables over some particular variables for the design of efficient systems, and simultaneously, it reveals the effects of uncontrolled factors.[41,42]

The L_9 orthogonal array for three levels and three factors was used by Ahmad *et al.*[20]; each row of each array (three different arrays set for three different abrasives sizes, for investigation of a good range of abrasive size) represents the trial situation with a combination of factors. Columns were corresponding to each control parameter or process variable in the array, and there were nine combinations, as it is in the L_9 orthogonal array. The response parameter was Change in Surface Roughness (ΔR_a). This technique has been effectively utilised by numerous researchers to understand the finishing impacts during the MAF process.[4,20,21] Nevertheless, the satisfactory decision of the researched parameters (control factors) is fundamental.

The observations in this examination were led according to a non-orthogonal array[43,44] as per machining set-up restrictions. The determination of the non-orthogonal array depends on relying on the prerequisite that did not fit by the orthogonal array. Table 2 shows the control factors, their units, definition, and levels. It is notable that the control factors level is non-uniform for abrasive size because of sieve size restriction. The experimental data[20] presented in Table 3 have been used to develop the ANN model. The following observations were set in a fashion using an L_{27} non-orthogonal array with four inputs and one output. The input parameters

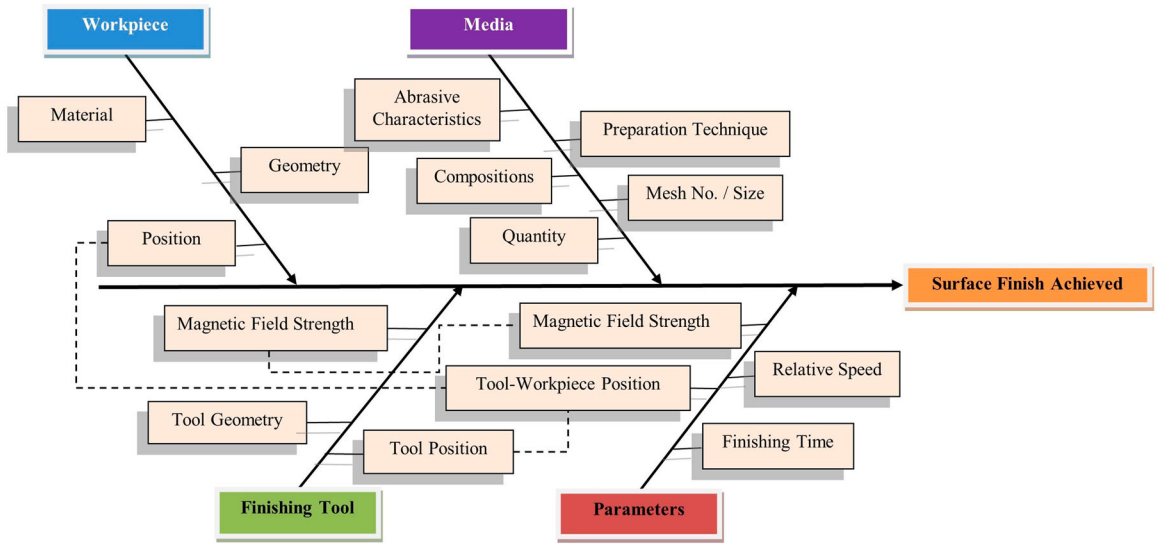


Figure 2. Fishbone diagram for MAF process.

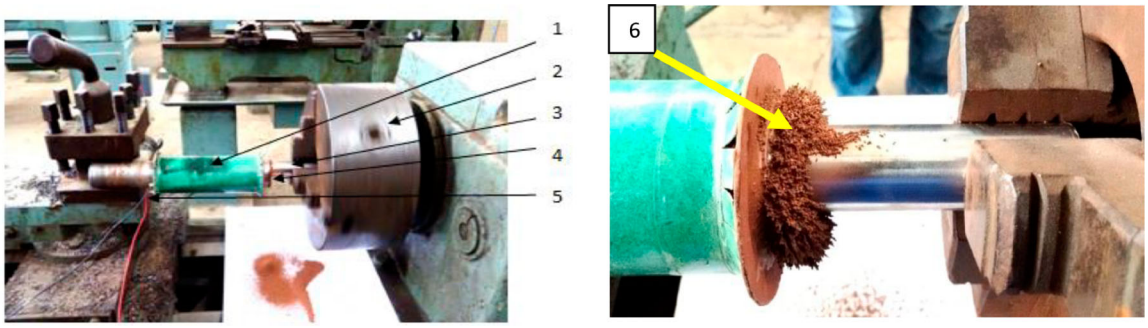


Figure 3. Images of the MAF setup: (1) Electromagnet (2) Chuck, (3) Workpiece, (4) Flexible Magnetic Abrasive Brush (FMAB), and (5) D.C. Power supply (6) FMAB in a closer view.

Table 2. Control factors, units, definition, and levels.

Control factor	Unit	Definition	Level 1	Level 2	Level 3
Abrasive Size (*d*)	µm	The average size of abrasive particles	90	150	300
Voltage (*V*)	Volts	Voltage applied to electromagnet to alter magnetic intensity	6	12	18
Machining Gap (*x*)	mm	The gap between the electromagnetic tool and workpiece	1	1.5	2
Rotational speed (*S*)	rpm	The relative rotational motion between tool and workpiece	60	90	120

Time 20 min for each run.

were abrasive size, voltage, machining gap, and rotational speed. The Change in Surface Roughness (ΔR_a) was an output/response parameter, measured in microns (µm).

In the Taguchi method, the signal-to-noise (S/N) ratio response has been formed corresponding to each observation, and 'larger-the-better' quality characteristic was targeted in this assessment, as more difference in surface roughness required. The control factors and the process response were taken as inputs to calculate the S/N ratio. Moreover, ANOVA analysis was carried out to scrutinise which controlling factor was statistically significant, so the best combination of the control factors, i.e. optimal combination, can be forecasted.[42] The S/N ratio (larger the better quality) for the change of surface roughness given by the Taguchi method is given below:

$$S/N = -10\log\frac{1}{n}\left(\frac{1}{y_1^2}+\frac{1}{y_2^2}+\cdots+\frac{1}{y_n^2}\right)$$

where $y_1, y_2, \ldots, y_n$ are the responses of the change of surface roughness and n is the number of observations. For each level, the average of selected characteristics of the factor has been shown in Table 3 (see *S/N* ratio). The response, *S/N* ratio, merges the reiterations, and the impact of noise levels into one whole coherent information. ANOVA for the *S/N* ratio was carried out for the classification of the significant factor.[45] The mean-response graphs were plotted, whereas the percentage contribution of factors was resolved by ANOVA.

4. Results and discussion

The prime motive of this study was to calculate the maximum Change in Surface Roughness $(\Delta R_a)_{max}$ and optimal values of factors correspondingly that have the most significant influence. Experimental data, which connect the influence of abrasive size, voltage, machining gap, and rotational speed, are shown in Table 3. These factors have a vital impact on the MAF process.

4.1 Analysis of variance (ANOVA)

For the computation of the process quality and its characteristics, the experimental data has been converted into S/N ratio. Based on S/N ratio, a ranking of control factors have been shown in Table 4. It is clear from Table 4 that the prevailing factor that affects the ΔR_a is voltage, and then machining gap, followed by abrasive size and rotational speed, respectively. Also, the variations in S/N ratio are shown in the main effects plot (Figure 4) where the most optimum result by Taguchi analysis is observed as abrasive size, voltage, machining gap, and rotational speed at 150 µm, 18 V, 1 mm, and 90 rpm respectively.

The analysis of the experimental data was carried out using ANOVA. The ANOVA analyses of the ΔR_a for all the analysed factors are shown in Table 5. The factors, having *P*-value less than 0.05 show a statistically considerable effect on the process performance. Also, Table 5 shows the percentage contribution of all factors in the overall result.

By scrutinising the results from Table 5, it can be noticed that voltage has the biggest influence on ΔR_a (% contribution = 41.60%), so much so, that this factor should be the centre of attention. Besides voltage, the change in surface roughness is also affected by machining gap (% contribution = 27.98%), abrasive size (% contribution = 10.95%), and rotational speed (% contribution = 9.39%).

The calculated S/N ratio was satisfactory because of the points on the normal probability plot in Figure 5 showing the approximate linear pattern with normal distribution having small disintegration.[19,45] Also, frequency versus residual plot Figure 5 confirms better the data variation as there were no outliers, and the negatively skewed histogram shows the initial observed S/N ratio has more impacts on results. The plot of residuals versus fitted value shows a random pattern (illustrated in Figure 5), which means residuals have approximately constant variance, and the residuals versus order plot shows that the data collected can be used to find the non-random error.

4.2 Interaction plot and the effect of factors on the change in surface roughness

The effect of the various process factors (Data means) on ΔR_a is shown in Figure 6. Factors corresponding to near-

Table 3. Experimental design.

Exp. Run	Abrasive Size/ µm	Voltage/ V	Machining Gap/ mm	Rotational Speed/ rpm	Experimental ΔR_a (A)	S/N ratio	ANN Predicted ΔR_a (B)	% Error ($\|A-B\|/A*100$)
1	90	6	1	60	0.13	−17.7211	0.1300	0
2	90	6	1.5	90	0.11	−19.1721	0.1099	0.0909
3	90	6	2	120	0.11	−19.1721	0.1101	0.0909
4	90	12	1	90	0.21	−13.5556	0.2103	0.1428
5	90	12	1.5	120	0.14	−17.0774	0.1396	0.2857
6	90	12	2	60	0.11	−19.1721	0.1097	0.2727
7	90	18	1	120	0.23	−12.7654	0.2300	0
8	90	18	1.5	90	0.17	−15.3910	0.1702	0.1176
9	90	18	2	60	0.15	−16.4782	0.1508	0.5333
10	150	6	1	60	0.15	−16.4782	0.1498	0.1333
11	150	6	1.5	90	0.14	−17.0774	0.1401	0.0714
12	150	6	2	120	0.13	−17.7211	0.1294	0.4615
13	150	12	1	90	0.24	−12.3958	0.2397	0.1250
14	150	12	1.5	120	0.15	−16.4782	0.1494	0.4000
15	150	12	2	60	0.13	−17.7211	0.1300	0
16	150	18	1	120	0.25	−12.0412	0.2497	0.1200
17	150	18	1.5	90	0.19	−14.4249	0.1900	0
18	150	18	2	60	0.18	−14.8945	0.1791	0.5000
19	300	6	1	60	0.11	−19.1721	0.1073	2.4540
20	300	6	1.5	90	0.10	−20.0000	0.1002	0.2000
21	300	6	2	120	0.12	−18.4164	0.1198	0.1666
22	300	12	1	90	0.17	−15.3910	0.1707	0.4117
23	300	12	1.5	120	0.16	−15.9176	0.1654	3.3750
24	300	12	2	60	0.10	−20.0000	0.1015	1.5000
25	300	18	1	120	0.21	−13.5556	0.2097	0.1428
26	300	18	1.5	90	0.16	−15.9176	0.1597	0.1875
27	300	18	2	60	0.14	−17.0774	0.1402	0.1428

Table 4. Response table for S/N Ratios – Larger is better.

Level	Abrasive size	Voltage	Machining gap	Rotational speed
1	−16.72	−18.33	−14.79	−17.63
2	−15.47	−16.41	−16.83	−15.93
3	−17.27	−14.73	−17.85	−15.91
Delta	1.80	3.60	3.06	1.73
Rank	3	1	2	4

horizontal line show insignificant effect, whereas factors corresponding to the highest inclined lines have the most significant effect. The most effective factor on ΔR_a was Voltage as its increment was highest, while the other factors show a lesser impact. ΔR_a increases with the increase of Voltage and Rotational Speed. Simultaneously, ΔR_a decreases with the increase of Machining Gap and Abrasive Size. The highest ΔR_a was found for the highest Voltage and the highest Rotational Speed.

4.3 Linear regression model

The linear dependence between the value of controlling factors (Abrasive Size, Voltage, Machining Gap, and Rotational Speed) and ΔR_a was obtained. Using MINITAB 17 software, the linear regression equation was derived from set values of abrasive size, voltage, machining gap, and rotational speed and ANOVA. However, a linear regression model was obtained as following to predict ΔR_a. A linear dependence between parameters and the unknown quantity is provided by the

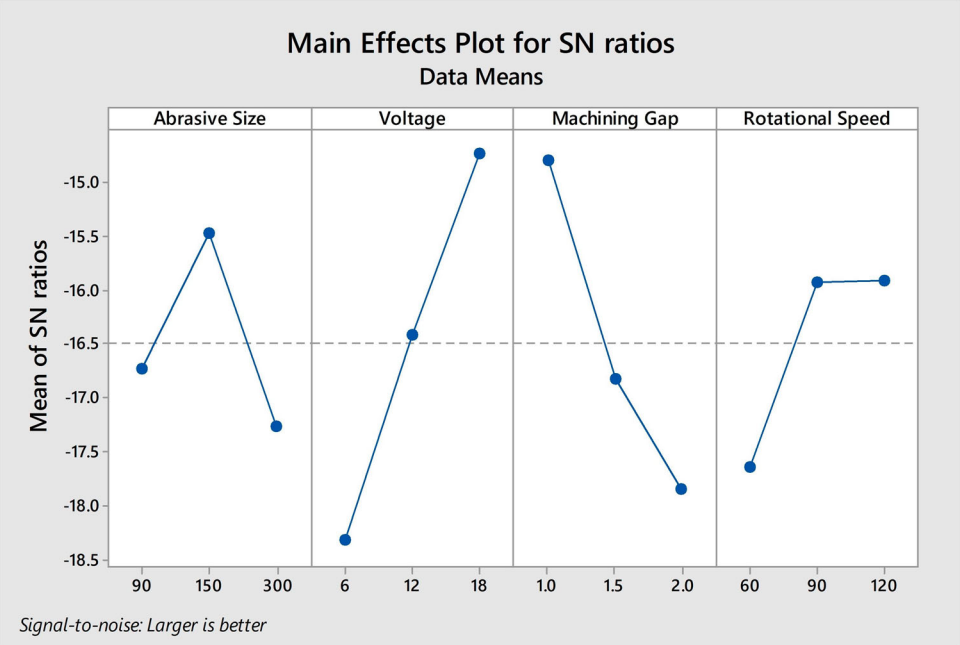


Figure 4. Main effects plot for S/N Ratio for ΔR_a.

Table 5. Analysis of variance for S/N ratio for change in surface roughness

Source	DF	Adj SS	Adj MS	*F*-value	*P*-value	% contribution
Abrasive size	2	15.350	7.6748	14.44	0.000	10.95
Voltage	2	58.344	29.1718	54.88	0.000	41.60
Machining gap	2	39.225	19.6126	36.89	0.000	27.98
Rotational speed	2	13.165	6.5826	12.38	0.000	9.39
Error	18	9.569	0.5316			6.80
Total	26	140.238				

developed model.

$$\text{Experimental } \Delta R_a(A) = 0.1511 - 0.000080d + 0.005370X_1 - 0.0549X_2 + 0.000403X_3 \quad (1)$$

$\Delta R_a = 0.23938$ (at abrasive size $d = 150$ µm, voltage $X_1 = 18$ V, machining gap $X_2 = 1$ mm, rotational speed $X_3 = 90$ rpm)

The developed linear regression equation with the help of MINITAB 15 for the ΔR_a (particularly for 150-micron size abrasive particle) is as below.[20]

$$\Delta R_a = 0.1850 + 0.00556X_1 - 0.0646X_2 + 0.00021X_3 \quad (2)$$

$\Delta R_a = 0.21713$ (at $d = 150$ µm, $X_1 = 18$ V, $X_2 = 1$ mm, $X_3 = 90$ rpm)

Based on the above Equations (1) and (2), a difference has been shown between non-orthogonal array and an orthogonal array, and it applied for the chosen controlling factors and their values. However, from Equations (1) and (2), coefficients showed that each factor has individual impacts, on ΔR_a which were dissimilar. The ΔR_a increases with increasing

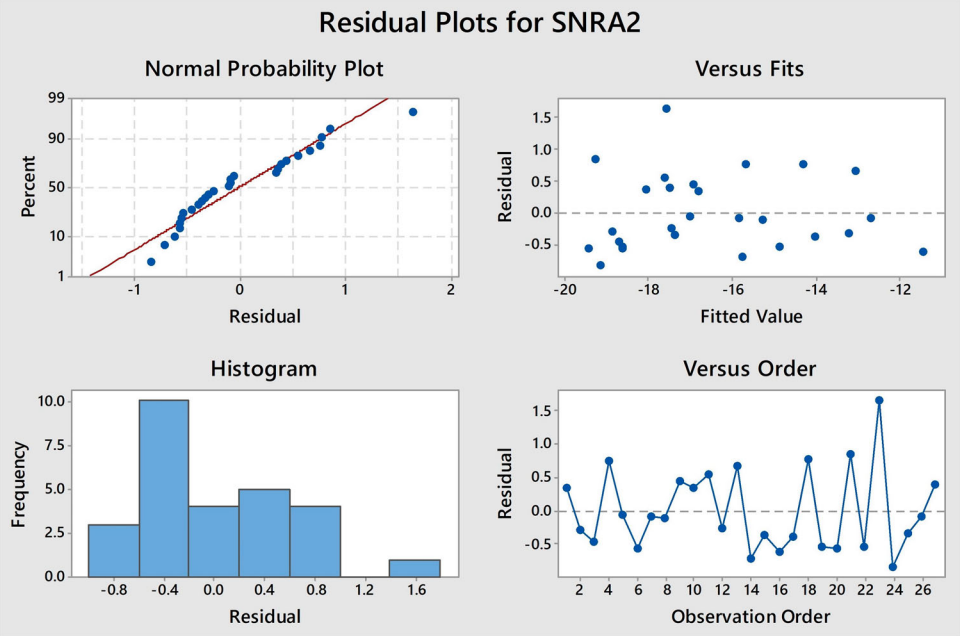


Figure 5. Residual plot for S/N Ratio for the Change in Surface Roughness (ΔR_a).

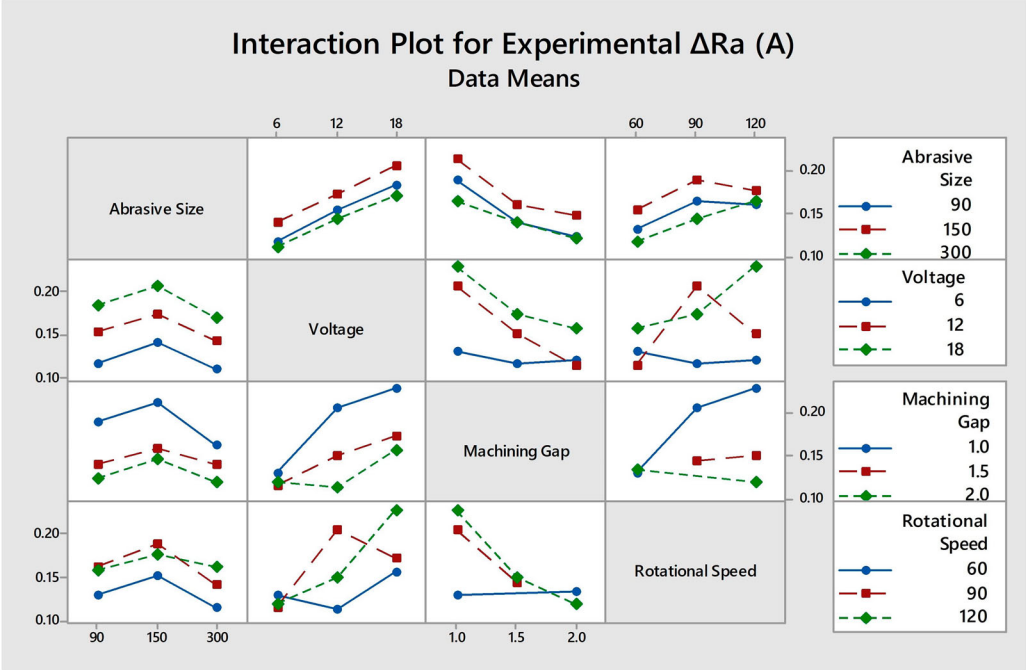


Figure 6. Interaction plot for ΔR_a.

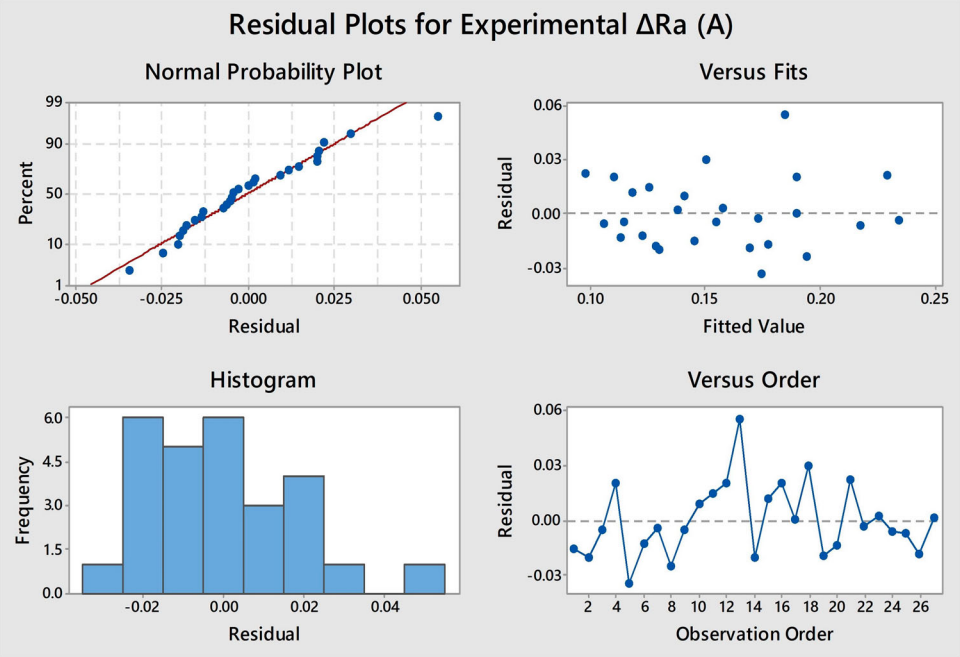


Figure 7. Residual plot for the change of surface roughness (ΔR_a).

Voltage and Rotational Speed, and decreases with increasing Machining Gap and Abrasive Size. The model formulated for prediction of ΔR_a, represented by Equation (1) was satisfactory because of the points on the normal probability plot in Figure 7 showing the approximate linear pattern with normal distribution having small disintegration.[44,45] In Figure 7 the plot of residuals versus fitted value shows a random pattern, which suggests that residuals have approximately constant variance, and the residuals versus order plot shows that the data collected can be used to find the non-random error. Also, the frequency versus residual plot in Figure 7 confirms the finer data variation as there were no outliers; however, the histogram is negatively skewed showing that the initial observations have more weight on the overall results.

4.4 Artificial Neural Network (ANN)

Artificial Neural Networks (ANN), normally referred to as Neural Networks, are analogous to biological neural networks. These networks have been used to approximate a function (in mathematical model form) that depends on a huge number of inputs. Due to the highly complex nature of calculations involved during the development of the network, software is normally used.[46] For data training and testing, MATLAB R2015a software has been used in this work.

Neural networks contain an input layer which is composed of neurons corresponding to the various process parameters. The experimental or realtime data (x_i) related to these parameters are subsequently fed into the neurons of the input layer. Input data are transferred to the hidden layer and output layer by multiplying weights (w_{ij}) between two neurons and adding using a summation function as shown in Equation (3). The feed-forward ANN model developed for modelling of ΔR_a (response or output) contains 4 different input layers of neurons equal to four input parameters (in this case, abrasive size, voltage, machining gap, and rotational speed). The developed ANN architecture was 4-10-8-1, i.e. four inputs, two hidden layers (10 and 8 neurons, respectively), and one output, as illustrated in Figure 8. This ANN architecture (4-10-8-1) has been found more appropriate for the modelling of data for which error is nearly zero as shown in Figure 9(c). For the present data, the ANN training algorithm was '*trainlm*' (trainlm is often the fastest backpropagation algorithm in the toolbox). The activation transfer functions (mathematical function that converts and processes the input data of each

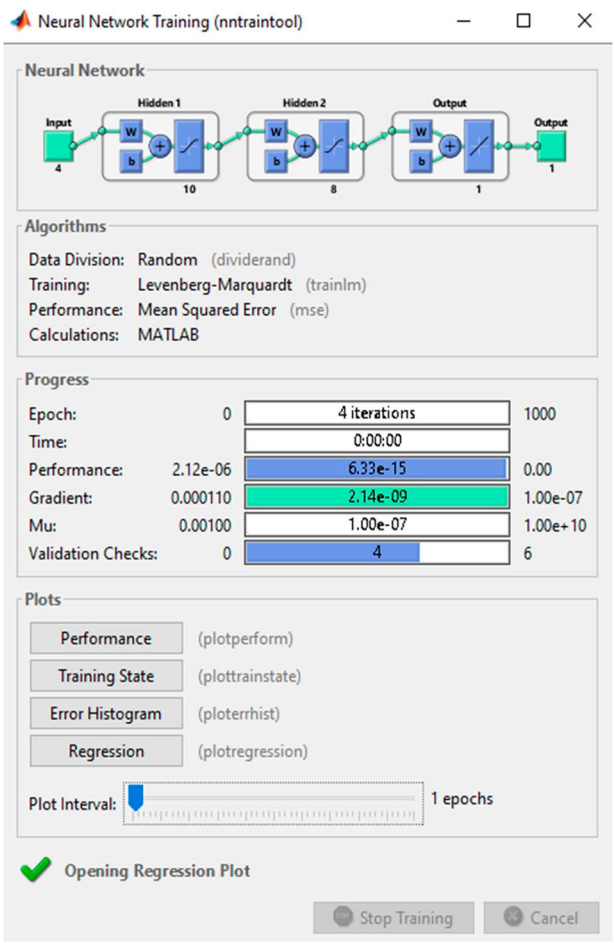


Figure 8. Neural network training.

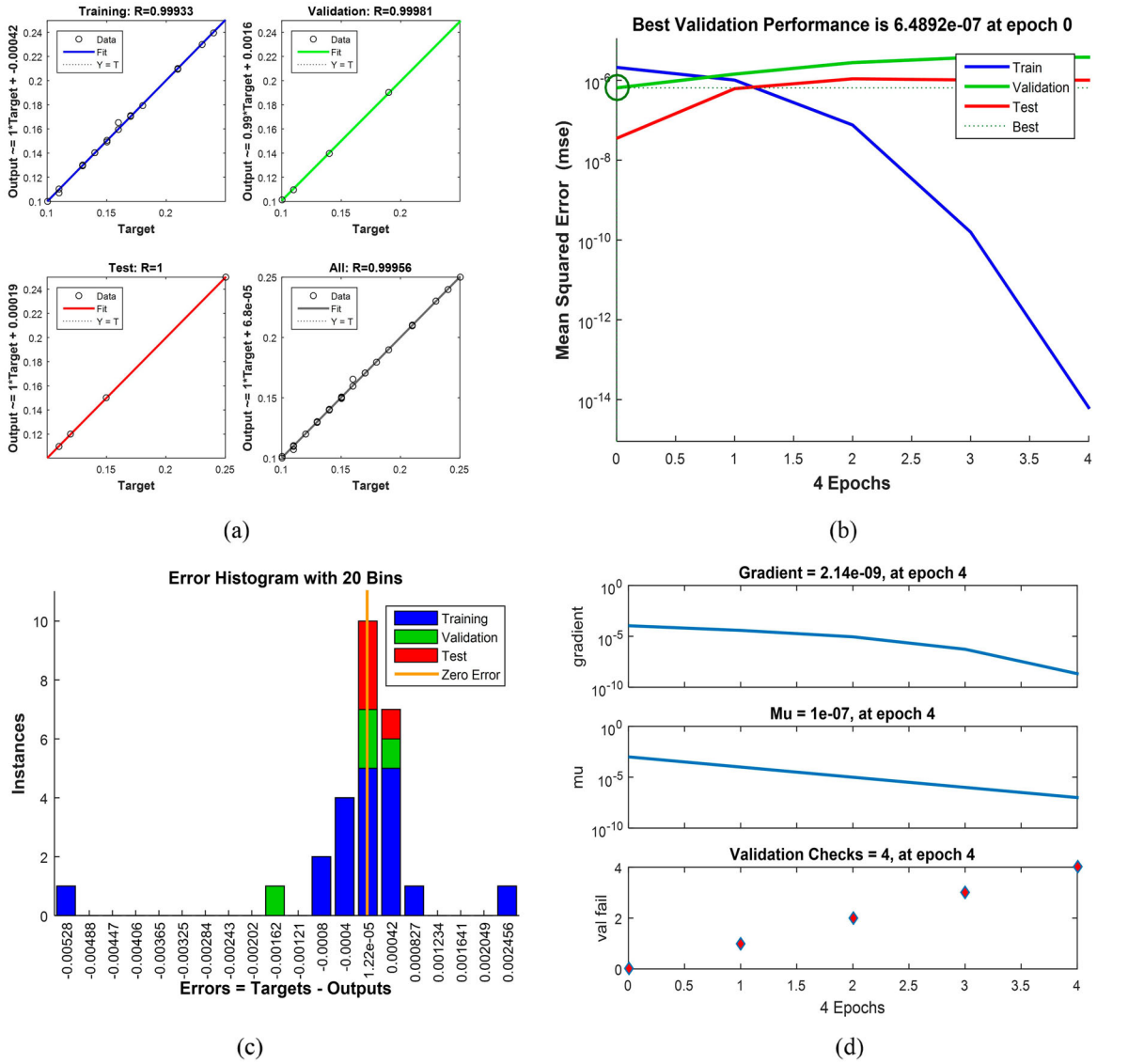


Figure 9 Training plots of neural network for ΔR_a: (a) regression Performance for trained ANN (b) variation of error with epochs (c) Error histogram plot (d) training state of ANN.

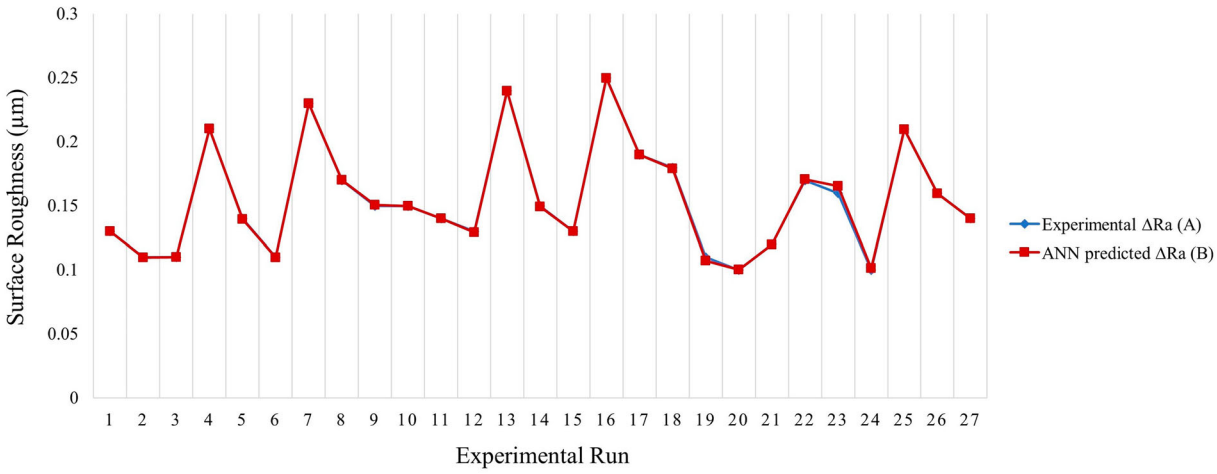


Figure 10 Comparison for surface roughness.

layer to outputs) '*logarithmic sigmoid logsig*' and the linear transfer function '*purelin*' were used as they offer better predictions.[31,32] This training algorithm updates weights and biases in pursuance of '*Levenberg–Marquardt*' optimisation (used to solve non-linear least squares problems). For the training, testing and validation of ANN, 70%, 15%, and 15% data are used, respectively. Regression coefficients can be acquired for the training, validation, and testing after ANN

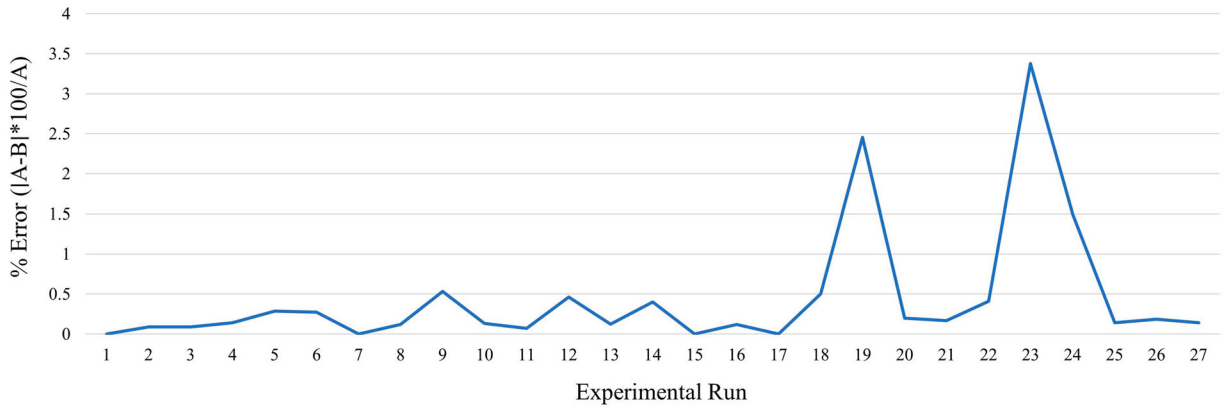


Figure 11 Percentage error plot between experimental and ANN predicted result.

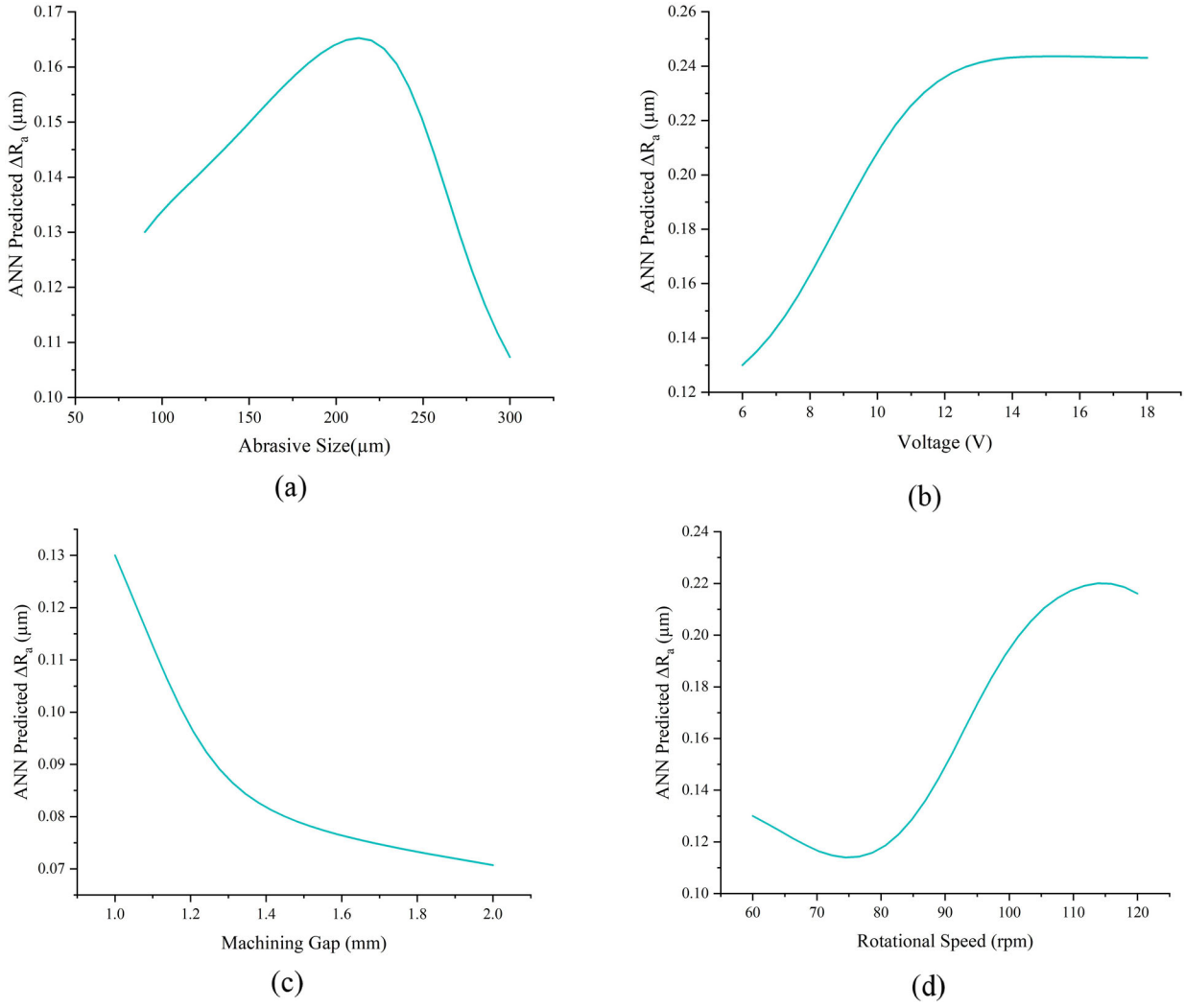


Figure 12. ANN predicted the individual effect of process variables (a) effect of abrasive size on ΔR_a (b) effect of voltage on ΔR_a (c) effect of machining gap on ΔR_a (d) effect of rotational speed on ΔR_a.

training.[32,33,46]

$$\text{Output} = f\left(\sum_{i=1}^{n} b + x_i w_i\right) \tag{3}$$

where, b is bias, experimental or realtime data (x_i), corresponding weights (w_{ij}), and f is activation function.

The trained ANN performance is shown in Figure 9(a), which points to the rationally good performance of trained ANN because the validation and test curves are perfectly on line; the error histogram also confirms this fact in Figure 9(c). It cannot be ignored that the best validation performance has attained zero epoch (which shows the number of passes the ANN training has performed over the entire training dataset) with the value of $6.4892*10^{-7}$, where the training persists until the 4th epoch, Figure 9 (b), showing the best validation performance. Figure 9(d) shows the training state of ANN, where the gradient

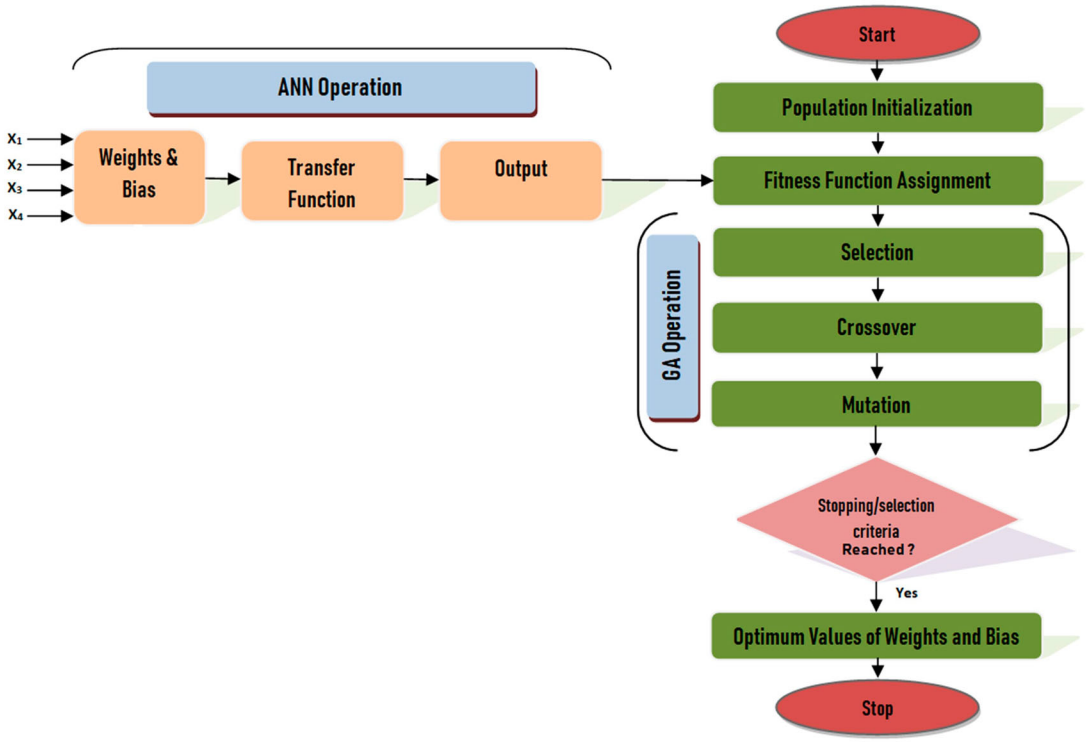


Figure 13. Flow chart of ANN-GA model.

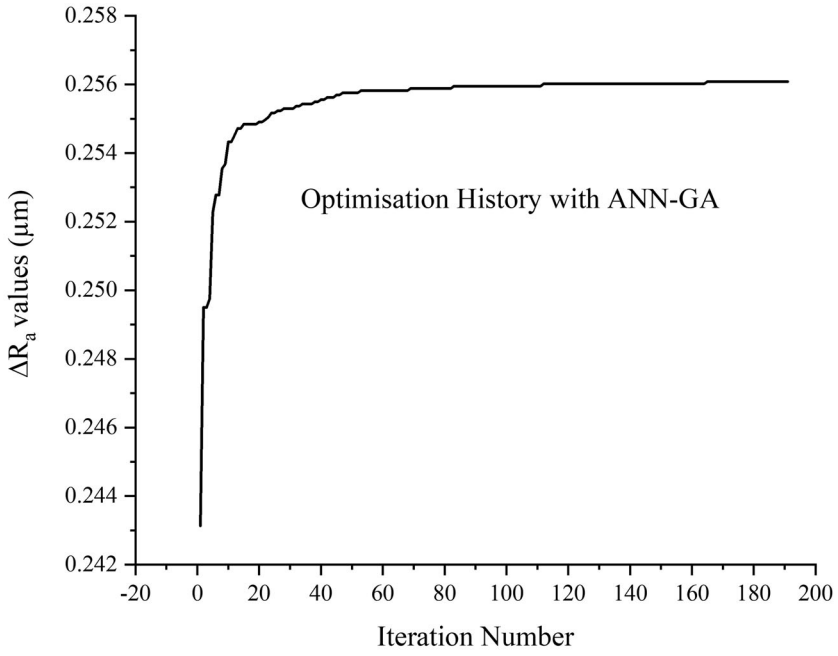


Figure 14. Optimisation history plot.

coefficient's final value is $2.14*10^{-9}$, approximately zero at the 4th epoch. Also, a diminishing value of gradient can be seen with the increase of epochs number. From Figure 10, it is clear that the ANN predicted and experimental values of ΔR_a between initial, and final states conform with each other to a very high degree, with the maximum error equal to 3.43%, and the other errors are negligible. Figure 11 shows the percentage error plot. Hence, it is certain that the developed ANN model has effectively learned the relationship between the input values and ΔR_a values, and thus can be used for maximising the change in surface roughness.

4.5 The individual effect of process parameters predicted by ANN Training

ANN training has predicted the relationship between the ΔR_a with controlling factors by varying the value of one factor, keeping constant value of other factors.

A relationship between the ΔR_a with abrasive size is shown in Figure 12(a). Initially, the trend is increasing until a certain limit, then it decreases. Due to the weaker magnetic bond formation on the edge of an irregularity,[39] the abrasive particles which are tinier than the peaks of irregularities get detached and cannot perform the shearing action satisfactorily. The

Table 6. Parameters of selection

Population size	Population type	Crossover fraction	Mutation rate	Max. generations	Initial population	Initial scores
50	Double vector	0.8	0.01	191	50	50

increasing size of the abrasive particles improves the shearing action and starts removing irregularities. The particles of size bigger than the irregularities shear-off the irregularities of the surface with a ploughing action. These ploughing actions further cause higher surface roughness by removing the material from the root.[47]

Since voltage increase causes an increase in the strength and contact area of the magnetic brush with the target surface, this escalates the normal magnetic force (F_n); consequently, the cutting force (F_c) increases.[4] This increased F_c helps the magnetic abrasive particles to remove irregularities of the surface more efficiently. Hence, the increase in voltage causes an increase in ΔR_a and the same trend has been predicted by ANN as illustrated in Figure 12(b).

When the face of the electromagnet comes closer to the targeted surface, then dense electromagnetic flux causes a higher magnetic force which controls the magnetic abrasive particles to perform more intensely, and vice-versa when further away. Thus, ΔR_a decreases when the machining gap increases, the relationship illustrated in Figure 12(c).

Figure 12(d) shows an increase in ΔR_a when rotational speed in most of the ranges increases; initially it decreases slightly because the normal magnetic force marginally decreases with increase in speed.[4,48]

4.6 Optimisation

The optimal selection of parameters has been carried out to increase the utility and commercial viability of manufacturing processes. Genetic Algorithms (GAs) solve optimisation problems that depend on several variables. Using an input-output pattern of data from the MAF process, by combining the capability of ANN and GA, can solve both the modelling and the optimisation problems.[46]

The basic optimisation problem for this ANN-GA approach is as in the following Equation (4):[49]

$$\text{Optimise } y = f(x, w) \tag{4}$$

where y represents the response or outcome of GA iteration, such as a change in surface roughness; x is a variable (instantaneous process factor during iteration) in the neural network, and w represents weights that evolve in the network training process. $f(.)$ is a fitness function that depends on the nature of input–output relation.

For the accomplishment of the target, a two-phase process has been applied. These two phases are modelling phase and optimisation phase. Conventionally, the standard optimisation problem is to be defined as an objective function. In this surface finishing operation, the difference between the initial and final roughness has to be maximum to achieve minimum surface roughness.

Thus, the present problem is the maximisation of ΔR_a and the objective function can be expressed as:

$$\left.\begin{array}{l}\text{Find values of:(V, x, S, d)for which, } \Delta R_a\text{is maximum}\\ \text{Subjected to range of magnetic abrasive finishing parameters:}\\ 6\text{ Volts} \le \text{V} \le 18\text{ Volts}\\ 1\text{ mm} \le \times \le 2\text{ mm}\\ 90\text{ rpm} \le \text{S} \le 120\text{ rpm}\\ 90\,\mu\text{m} \le \text{d} \le 300\,\mu\text{m}\end{array}\right\} \tag{5}$$

where ΔR_a is change in the surface roughness, 'V' is voltage, 'x' is maching gap, 'S' is rotational speed and 'd' is abrasive size. ΔR_a is a function of the affecting parameters, trained by neurals network.

An effective global optimisation method genetic algorithm (GA) can be used to solve the optimisation problem expressed by Equation (4). GA is a search algorithm, analogous to the process of natural selection that belongs to the class of evolutionary algorithms (EA). Genetic algorithms optimise the problems using methods motivated by the phenomenon of reproduction, mutation, crossover (recombination), and selection.[50] Manufacturing problems that contain data sets as information and defining a continuous and smooth mathematical formula is not possible for 'optimisation objective function,' these can be modelled by ANN and optimised by GA. Figure 13. shows a methodology for the ANN-GA operation.

The decisive parameters of GA optimisation operations are the population size, crossover fraction mutation rate, iterations number (i.e. generations). The opted initial population size was 50, the crossover rate of 0.8, the mutation rate of 0.1, and a number of iterations of 191 were utilised; the optimisation was terminated, and the optimisation history plot up to 191 iterations is shown in Figure 14. Parameters of selection for GA operation are illustrated in Table 6.

The optimisation problem in Equation (4) has been solved with the constraint (lower and upper limits of controlling factors) to search the optimum values of finishing parameters. The contribution of the optimisation process in maximising the ΔR_a is evident when compared with the Taguchi-ANOVA optimised result. From Table 7, ANN-GA shows $(\Delta R_a)_{max}$ of 0.256 µm, whereas Taguchi and ANOVA analysis find $(\Delta R_a)_{max}$ of 0.239 µm.

Table 7. Comparison of results obtained from the Taguchi method and ANN-GA Optimisation

Control factor	Taguchi predicted	ANN-GA Predicted
Abrasive Size/µm	150	105.158
Voltage/V	18	17.999
Machining Gap/mm	1	1.001
Rotational Speed/rpm	90	91.890
Change in Surface Roughness/µm	0.239	0.256

5. Conclusions

Based upon the above results and discussion the following conclusions can be made-

1. ANN-GA is a robust modelling and optimisation tool for any sort of data set (orthogonal array design or non-orthogonal array design) appropriate to scrutinising the finishing of SS302, and it was noticed that the parametric design used in this study provides a straightforward, methodical, and proficient method of modelling and

optimisation of change of surface roughness or finishing behaviour during the MAF process.
2. Modelling and optimisation done with ANN-GA show a maximum value of $(\Delta R_a)_{max}$ equal to 0.25610 µm, which is 7% better than the result obtained from Taguchi-ANOVA analysis.
3. From the regression analysis, ΔR_a increases with the increase of Voltage and Rotational Speed and decreases with increasing Machining Gap and Abrasive Size, whereas the ANN modelling predicts the same results except for the effect of abrasive size, i.e. ΔR_a increases then decreases over the range, and this argument is justified by the optimum value of abrasive size (between the lower and upper limits) obtained from both the optimisation techniques.
4. Voltage has the highest influence on ΔR_a, i.e. 41.60%, followed by Machining Gap (27.98%), while Abrasive Size (10.95%) and Rotational Speed (9.39%) show almost four times lower effect.
5. An ANN model for predicting ΔR_a was developed, and a linear regression equation was developed, where ΔR_a depends on abrasive size, voltage, machining gap, and rotational speed at a confidence level of 99.5%. By comparing the results obtained from ANN methods, and ANOVA, it can be concluded that the ANN model is capable of predicting with very high accuracy and is more efficient.

Acknowledgment and Funding

Authors offer the most sincere gratitude to the Council of Scientific and Industrial Research (CSIR), Government of India, for financial support in the form of Senior Research Fellowships.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

1. S. Khalaj Amineh, A. Fadaei Tehrani and A. Mohammadi: Improving the surface quality in wire electrical discharge machined specimens by removing the recast layer using magnetic abrasive finishing method. *Int. J. Adv. Manuf. Technol.*, 2012, **66**, 1793–1803. doi:10.1007/s00170-012-4459-7.
2. J. Guo, J. Bai, K. Liu and J. Wei: Surface quality improvement of selective laser sintered polyamide 12 by precision grinding and magnetic field-assisted finishing. *Mater. Des.*, 2018, **138**, 39–45. doi:10.1016/j.matdes.2017.10.048.
3. A. K. Singh, S. Jha and P. M. Pandey: Parametric analysis of an improved ball end magnetorheological finishing process. *Proc. Inst. Mech. Eng. Part B J. Eng. Manuf.*, 2012, **226**, 1550–1563. doi:10.1177/0954405412453805.
4. D. K. Singh, V. K. Jain and V. Raghuram: Parametric study of magnetic abrasive finishing process. *J. Mater. Process. Technol.*, 2004, **149**, 22–29. doi:10.1016/j.jmatprotec.2003.10.030.
5. M. Mosavat and A. Rahimi: Numerical-experimental study on polishing of silicon wafer using magnetic abrasive finishing process. *Wear*, 2019, **424–425**, 143–150. doi:10.1016/j.wear.2019.02.007.
6. H. Yamaguchi, A. K. Srivastava, M. A. Tan, R. E. Riveros and F. Hashimoto: Magnetic abrasive finishing of cutting tools for machining of titanium alloys. *CIRP Ann.*, 2012, **61**, 311–314. doi:10.1016/j.cirp.2012.03.066.
7. V. K. Jain: Magnetic field assisted abrasive based micro-/nano-finishing. *J. Mater. Process. Technol.*, 2009, **209**, 6022–6038. doi:10.1016/j.jmatprotec.2009.08.015.
8. V. K. Jain: Abrasive-based nano-finishing techniques: an overview. *Mach. Sci. Technol.*, 2008, **12**, 257–294. doi:10.1080/10910340802278133.
9. N. Sihag, P. Kala and P. M. Pandey: Chemo assisted magnetic abrasive finishing: experimental investigations. *Procedia CIRP*, 2015, **26**, 539–543. doi:10.1016/j.procir.2014.07.067.
10. S. R. Bhagavatula and R. Komanduri: On chemomechanical polishing of Si_3N_4 with Cr_2O_3. *Philos. Mag. A*, 1996, **74**, 1003–1017. doi:10.1080/01418619608242173.
11. A. Misra, P. M. Pandey and U. S. Dixit: Modeling of material removal in ultrasonic assisted magnetic abrasive finishing process. *Int. J. Mech. Sci.*, 2017, **131–132**, 853–867. doi:10.1016/j.ijmecsci.2017.07.023.
12. S. A. Sirwal and A. K. Singh: Analysis of the surface roughness for novel magnetorheological finishing of a typical blind hole workpiece. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.*, 2019, **233**, 1541–1561. doi:10.1177/0954406218776036.
13. V. K. Jain: 'Advanced machining processes', 2013, New Delhi: Allied Publishers.
14. S. Yang and W. Li: 'Surface finishing theory and new technology', 2018, Berlin, Heidelberg: Springer Berlin Heidelberg.
15. D. K. Singh, V. K. Jain, V. Raghuram and R. Komanduri: Analysis of surface texture generated by a flexible magnetic abrasive brush. *Wear*, 2005, **259**, 1254–1261. doi:10.1016/j.wear.2005.02.030.
16. H. Yamaguchi, A. K. Srivastava, M. Tan and F. Hashimoto: Magnetic abrasive finishing of cutting tools for high-speed machining of titanium alloys. *CIRP J. Manuf. Sci. Technol.*, 2014, **7**, 299–304. doi:10.1016/j.cirpj.2014.08.002.
17. S. K. Amnieh, P. Mosaddegh and A. F. Tehrani: Study on magnetic abrasive finishing of spiral grooves inside of aluminum cylinders. *Int. J. Adv. Manuf. Technol.*, 2017, **91**, 2885–2894. doi:10.1007/s00170-016-9970-9.
18. J. Guo, H. Wang, M. Goh and K. Liu: Investigation on Surface Integrity of rapidly solidified aluminum RSA 905 by magnetic field-assisted finishing. *Micromachines (Basel)*, 2018, **9**, 146. doi:10.3390/mi9040146.
19. Y. Gao, Y. Zhao, G. Zhang, G. Zhang and F. Yin: Polishing of paramagnetic materials using atomized magnetic abrasive powder. *Mater. Manuf. Process*, 2019, **34**, 604–611. doi:10.1080/10426914.2018.1532087.
20. S. Ahmad, S. Gangwar, P. C. Yadav and D. K. Singh: Optimisation of process parameters affecting surface roughness in magnetic abrasive finishing process. *Mater. Manuf. Process*, 2017, **32**, 1723–1729. doi:10.1080/10426914.2017.1279307.
21. C.-T. Lin, L.-D. Yang and H.-M. Chow: Study of magnetic abrasive finishing in free-form surface operations using the Taguchi method. *Int. J. Adv. Manuf. Technol.*, 2007, **34**, 122–130. doi:10.1007/s00170-006-0573-8.
22. T. C. Kanish, P. Kuppan, S. Narayanan and S. D. Ashok: A fuzzy logic based model to predict the improvement in surface roughness in magnetic field assisted abrasive finishing. *Procedia Eng.*, 2014, **97**, 1948–1956. doi:10.1016/j.proeng.2014.12.349.
23. R. K. Singh, S. Gangwar, D. K. Singh and V. K. Pathak: A novel hybridization of artificial neural network and moth-flame optimisation (ANN–MFO) for multi-objective optimisation in magnetic abrasive finishing of aluminium 6060. *J. Braz. Soc. Mech. Sci. Eng.*, 2019, **41**, 270. doi:10.1007/s40430-019-1778-8.
24. P. Kala and P. M. Pandey: Experimental study on finishing forces in double disk magnetic abrasive finishing process while finishing paramagnetic workpiece. *Procedia Mater. Sci.*, 2014, **5**, 1677–1684. doi:10.1016/j.mspro.2014.07.356.
25. K. B. Judal, V. Yadava and D. Pathak: Experimental investigation of vibration assisted cylindrical–magnetic abrasive finishing of aluminum workpiece. *Mater. Manuf. Process.*, 2013, **28**, 1196–1202. doi:10.1080/10426914.2013.811725.
26. Y.-H. Lee, K.-L. Wu, J.-H. Jhou, Y.-H. Tsai and B.-H. Yan: Two-dimensional vibration-assisted magnetic abrasive finishing of stainless steel SUS304. *Int. J. Adv. Manuf. Technol.*, 2013, **69**, 2723–2733. doi:10.1007/s00170-013-5242-0.
27. A. A. Moosa: Utilizing a magnetic abrasive finishing technique (MAF) via adaptive nero fuzzy (ANFIS). *Am. J. Mater. Eng. Technol.*, 2013, **1**, 49–53. doi:10.12691/materials-1-3-5.
28. P. Kala, S. Kumar and P. M. Pandey: Polishing of copper alloy using double disk ultrasonic assisted magnetic abrasive polishing. *Mater. Manuf. Process*, 2013, **28**, 200–206. doi:10.1080/10426914.2012.746704.
29. A. C. Wang, L. Tsai, C. H. Liu, K. Z. Liang and S. J. Lee: Elucidating the optimal parameters in magnetic finishing with gel abrasive. *Mater. Manuf. Process*, 2011, **26**, 786–791. doi:10.1080/10426914.2010.505620.
30. N. Sihag, P. Kala and P. M. Pandey: Analysis of surface finish improvement during ultrasonic assisted magnetic abrasive finishing on

chemically treated tungsten substrate. *Procedia Manuf.*, 2017, **10**, 136–146. doi:10.1016/j.promfg.2017.07.040.
31. E. Deniz, O. Aydogmus and Z. Aydogmus: Implementation of ANN-based selective harmonic elimination PWM using hybrid genetic algorithm-based optimisation. *Measurement (Mahwah N J)*, 2016, **85**, 32–42. doi:10.1016/j.measurement.2016.02.012.
32. H. Oktem, T. Erzurumlu and F. Erzincanli: Prediction of minimum surface roughness in end milling mold parts using neural network and genetic algorithm. *Mater. Des.*, 2006, **27**, 735–744. doi:10.1016/j.matdes.2005.01.010.
33. B. Stojanović, A. Vencl, I. Bobić, S. Miladinović and J. Skerlić: Experimental optimisation of the tribological behaviour of Al/SiC/Gr hybrid composites based on Taguchi's method and artificial neural network. *J. Braz. Soc. Mech. Sci. Eng.*, 2018, **40**, 311. doi:10.1007/s40430-018-1237-y.
34. G. Candan and H. R. Yazgan: Genetic algorithm parameter optimisation using Taguchi method for a flexible manufacturing system scheduling problem. *Int. J. Prod. Res.*, 2015, **53**, 897–915. doi:10.1080/00207543.2014.939244.
35. S. Datta, C. Garai and C. Das: Efficient genetic algorithm on linear programming problem for fittest chromosomes. *J. Glob. Res. Comput. Sci.*, 2012, **3**, 8.
36. K. Hanada, H. Yamaguchi and H. Zhou: New spherical magnetic abrasives with carried diamond particles for internal finishing of capillary tubes. *Diam. Relat. Mater.*, 2008, **17**, 1434–1437. doi:10.1016/j.diamond.2008.01.100.
37. H. Kumar, S. Singh and A. Srivastava: Parametric investigations into internal surface modification of brass tubes with alternating magnetic field. *Procedia Manuf.*, 2016, **5**, 1234–1248. doi:10.1016/j.promfg.2016.08.097.
38. D. K. Singh, V. K. Jain and V. Raghuram: Experimental investigations into forces acting during a magnetic abrasive finishing process. *Int. J. Adv. Manuf. Technol.*, 2006, **30**, 652–662. doi:10.1007/s00170-005-0118-6.
39. T. Mori, K. Hirota and Y. Kawashima: Clarification of magnetic abrasive finishing mechanism. *J. Mater. Process. Technol.*, 2003, **143–144**, 682–686. doi:10.1016/S0924-0136(03)00410-2.
40. G. C. Verma, P. Kala and P. M. Pandey: Experimental investigations into internal magnetic abrasive finishing of pipes. *Int. J. Adv. Manuf. Technol.*, 2017, **88**, 1657–1668. doi:10.1007/s00170-016-8881-0.
41. H. J. C. de Souza, M. B. Silva, C. B. Moyses, F. L. Alberto, F. J. Pontes, U. R. Ferreira, R.N. Duarte, C. E. Sanches da Silva: 'Robust design and Taguchi method application' in 'Design of experiments – Applications', (ed. M. Borges Silva), 1–20, 2013, InTech.
42. R. K. Roy: 'A primer on the Taguchi method', 2nd ed, 2010, Dearborn, MI, Society of Manufacturing Engineers.
43. C.-X. Ma, K.-T. Fang and E. Liski: A new approach in constructing orthogonal and nearly orthogonal arrays. *Metrika*, 2000, **50**, 255–268. doi:10.1007/s001840050049.
44. R. Mead: The Non-Orthogonal Design of Experiments. *J. R. Stat. Soc. Ser. A Stat. Soc.*, 1990, **153**, 151–201. doi:10.2307/2982800.
45. C. M. Judd, G. H. McClelland and C. S. Ryan: 'Data analysis: a model comparison approach to regression, ANOVA, and beyond', 3rd ed., 2017, New York, Routledge, Taylor & Francis Group.
46. K. Wang, H. L. Gelgele, Y. Wang, Q. Yuan and M. Fang: A hybrid intelligent method for modelling the EDM process. *Int. J. Mach. Tools Manuf.*, 2003, **43**, 995–999. doi:10.1016/S0890-6955(03)00102-0.
47. V. C. Shukla, P. M. Pandey, U. S. Dixit, A. Roy and V. Silberschmidt: Modeling of normal force and finishing torque considering shearing and ploughing effects in ultrasonic assisted magnetic abrasive finishing process with sintered magnetic abrasive powder. *Wear*, 2017, **390–391**, 11–22. doi:10.1016/j.wear.2017.06.017.
48. V. K. Jain, P. Kumar, P. K. Behera and S. C. Jayswal: Effect of working gap and circumferential speed on the performance of magnetic abrasive finishing process. *Wear*, 2001, **250**, 384–390. doi:10.1016/S0043-1648(01)00642-1.
49. K. Wang, G. L. Kovacs, M. Wozny and M. Fang, Eds.: 'Knowledge enterprise: intelligent strategies in product design, manufacturing, and management', Proceedings of PROLAMAT 2006, IFIP TC5 International Conference, June 15–17, 2006, Shanghai, China, vol. 207. Boston, MA, Springer US, 2006.
50. A. Ghosh and S. Tsutsui: 'Advances in evolutionary computing theory and applications', 2013, Berlin, Springer Berlin.

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Modelling Logistic Growth Model for COVID-19 Pandemic in India

Minni Jain
Delhi Technological University
Department of CSE
New Delhi, India
minnijain@dtu.ac.in

Prajwal Kumar Bhati
Delhi Technological University
Department of CSE
New Delhi, India
prajwalroxman@gmail.com

Pranav Kataria
Delhi Technological University
Department of CSE
New Delhi, India
pranavkataria_bt2k16@dtu.ac.in

Rohit Kumar
Delhi Technological University
Department of CSE
New Delhi, India
rohitmamc2016@gmail.com

Abstract-- **An early analysis of growth dynamics for infectious diseases, like COVID-19, is needed to dissect the crucial driving factors that result in rapid disease transmission, refine the measures taken to control the pandemic and improve disease forecast. The phenomenological models are used to identify the initial climbing growth period of COVID-19 outbreak in India and have modelled 3 major epidemic growth models: Generalized logistic growth, Logistic growth and Generalized growth, to predict the growth in the total number of positive cases, daily increase in the number of positive tested cases and the daily growth rate in confirmed positive cases, dated from Apr 10, 2020, to Apr 20, 2020. The bootstrap resampling method is applied for data prediction to process the sample data, dated from Jan 31, 2020, to Apr 10, 2020, and to calculate the 3 major growth parameters: r (Rate of growth at an early stage), K (Final epidemic size) and C(Number of aggregate cases at time t), which are used to calculate confidence intervals which predict the future direction of the curve and increase in the number of confirmed cases with 95% accuracy for the interval Apr 10, 2020, to Apr 20, 2020. Our models predict exponential and sub-exponential spread rate in the number of positive cases in India from Apr 10, 2020, to Apr 20, 2020. Our findings reveal that significant measures are needed to control the transmission rate of the virus in the community, as the models predict sub-exponential growth in India.**

Keywords – COVID-19; Phenomenological models; Epidemic Growth Models; Logistic Growth; Bootstrap Resampling Method; Growth parameters; Confidence Intervals; Data Prediction.

I. INTRODUCTION

In the early stages of an epidemic, the growth rate of the infection spread follows an exponential path. Extrapolation of available data is the most crucial method used to depict the growth dynamics of the epidemic. The predicted data help the policymakers to take preventive measures in the right direction to curb the spread of disease in the community.[11] In case of an epidemic, epidemic growth models are constructed to extrapolate the rate at which the disease might spread in the community. These epidemic growth models are constructed using the parameters obtained by exploring available data, using bootstrap resampling methods. In the bootstrap resampling method, multiple simulations run on the available dataset. In consecutive simulations, parameters are adjusted to extrapolate the most precise estimates that are further used to plot the respective epidemic growth models.[12] COVID-19 disease is now proclaimed as a pandemic around the globe. India comes among one of the few countries in the world which have implemented complete lockdown as a preventive measure to decrease the growth in positive cases of COVID-19 disease at an early stage of an outbreak. It has now spread across the whole world, badly affecting nations like the USA, Italy and Spain. This paper focuses on extrapolation of available data for India to construct early-stage epidemic growth models to dissect the crucial driving factors that are resulting in the rapid spread of disease, to refine the measures taken by authorities to control the pandemic and improve disease forecast.[4,5].

Logistic growth model construction is favourable and can unravel the growth and fall in the number of cases around the inflection point. If the logistic curve converges at an early stage, it means the pandemic is under control, suggesting that the steps taken to control the pandemic are working. Whereas, if the logistic curve does not converge and shows an exponential growth even after the occurrence of the inflection point, then it indicates that the measures taken in order to manage the pandemic are not working, and the epidemic is still in its early growth stage. Logistic models also involve calculation of important feature parameters like rate of growth at early stage(r), final epidemic size (K) and the number of aggregate cases at time t (C), growth deceleration (p), disease turning point(t)

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:55:03 UTC from IEEE Xplore. Restrictions apply.

which helps in accurate modelling of curve. Hence, the proposed method of logistic modelling is meritorious over conventional methods.

I. LITERATURE SURVEY

Novel coronavirus (COVID-19) belongs to a family of coronaviruses, which can cause various illness like SARS and MERS.[9] According to the World Health Organization, COVID-19 first appeared in the city of Wuhan, Hebei province, in China in December 2019. SARS-CoV-2 is a new mutation Coronavirus, which is identified as the cause of coronavirus disease of 2019.[8]This pandemic has now affected almost every country on the planet, and researchers around the world are racing to find a cure (or vaccine) for this virus.[9] USA, Italy and Spain are the worst hit nations.The situation in USA is very severe, because more than 50,000 people have lost their lives to this virus.[8] Active research in the field of Computers and Mathematics has helped researchers and doctors, to track the rate of growth of this virus in different countries.[1] Research referenced in [1] has shown that the rate of growth of virus in every country follows the same pattern, that is exponential in nature.

II. METHODS AND THEORY

A. DATASET SOURCES

This report is focused on daily data for the number of positive recorded cases from India. The dataset is sourced from the COVID-19 Patient Database[4] which consists of total confirmed cases state wise consisting of patient number, patient gender, age-bracket, travel history,detected city and current status for all cities in India. The data is updated daily. In this report the data from Jan 31 is used when the first case in India was reported to Apr 10 and used the models to forecast outbreak for next 5 to 10 days.

B. EPIDEMIC GROWTH MODELS

India is currently at the initial stage of the spread, that is, the number of positive cases has not reached a saturation peak yet, and the daily increase is still widening.

In the present scenario, the general epidemic growth models like Richards model and complete interval epidemic curves which constitute an extra variable 'α' signifying the imbalance between rise and sink for the growth frequency curve, cannot be constructed as the amount of data is very limited.

Early epidemic growth models usually result in an 'S' shape when they reach saturation stage. This happens when the epidemic is at an early stage and the data available for processing the model is limited.

Hence, this research focuses on 3 major models, which are constructed using few parameters:

1. Logistic Growth Model

$$\frac{dC(t)}{dt} = r(1 - \frac{C(t))}{K})C(t) \quad (1)$$

2. Generalized Logistic Growth Model(GLM):

$$\frac{dC(t)}{dt} = r(1 - \frac{C(t))}{K})C^p(t) \quad (2)$$

3. Generalized Growth Model (GGM):

$$\frac{dC(t)}{dt} = rC^p(t) \quad (3)$$

Here, the following main parameters are calculated to construct accurate growth models:

1. R : Rate of growth at early stage.
2. K : Final epidemic size.
3. C(t) : Number of aggregate cases at time t.
4. p : Signify profile of distinct growth.

Parameter p lies from 0 to 1, where if:

a. p = 0 : Implies constant growth rate.
b. 0 < p < 1 : Implies sub-exponential growth rate.
c. p = 1 : Implies exponential growth rate.

The 2 basic models, which are derived from Richard's model keeping the parameter α equal to 1 (α=1) and p=1, result in logistic development and generalized logistic development models. The development rate for these models deteriorates proportionally to the number of days. These curves tend to saturate the development in the figures for positive cases at an early juncture of epidemic resulting in optimistic results.

The (3) generalized growth model (GGM) allows for a steady increase in outbreak in the early phase (p<1), but it is unable to explain the decomposition of the frequency rate at advanced stages. Therefore, it acts as an upper barrier, or the worst case scenario that is obtained on the presumption that the upsurge will continue to develop following an expanding pattern as before.Logistic development (1) and Generalized Logistic development (2) usually result in an 'S' shaped curve, when the growth in the number of confirmed cases saturates. They predict very optimistic results, which can be seen as the best case scenario in the early stage analysis of the pandemic.

The generalized logistic growth model (2) results in sub-exponential development rate at a preliminary phase, hence it is preferred to study the imbalance between decay and growth dynamics. These three models are constructed using the available data for the total number of actual recorded cases in India.

C. PARAMETER ESTIMATION

For the parameter estimation stage, the Damped least square method, also known as Marquardt-Levenberg algorithm has been used to resolve the problem of least square nonlinear regression for curve fitting. In order to fit the normal function of normal growth in generalized case of Richard's model (α =1, p = 1), commence the starting spike and permit the algorithm to

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:55:03 UTC from IEEE Xplore. Restrictions apply.

calculate one of the 3 parameters optimally. Since, the early phase of growth models does not observe the logistic rate of development, hence to calculate the variability in our model variables, bootstrap resampling method along with a Poisson function is utilized for error estimation $f(\lambda)$, where λ is the expected value of the distribution as in prior inspections [2].

D. *CONFIDENCE INTERVAL*

The confidence interval mean is calculated for parameter estimation developed for Models (1), (2) and (3) by iterating 200 sets of realizations using parametric bootstrap resampling technique as used in prior studies [2].Parameters were then estimated from the realizations to construct 95% accurate Confidence Intervals. A Poisson error structure was used in the parametric bootstrap method for confidence interval estimation. Problems with the appropriate procedure occur when the model cannot be specified separately from the data resulting in unstable variations of the estimated parameters.The models parameters were evaluated for accurate identification using the simulations.

III. RESULTS

The Generalized linear model parameter p, which signifies the distinct profile growth of infected population is estimated to be 0.99 (95% Confidence Interval: [0.98, 1]). This high value of parameter p signifies that the growth rate is heading towards an exponential trend, and major preventive measures need to be enforced at the earliest.

On the contrary, the logistic model depicts a very optimistic best-case scenario , forecasting the total number infected population at 9,547 (95% Confidence Interval: [9349, 9746]) after 5 days (April 15), and 10,723 (95% Confidence Interval: [10514, 10933]) after 10 days.It might be a very optimistic result if the actual infected cases follow the logistic growth model. India announced their confirmed case for the first time on Jan 30. Since 10 April 2020, according to Indian administration, there were 206 deaths in India and 7,062 cases of infection, indicating that the mortality rate stands at 2.91% among the infected population.

While the growth rate of the infected population is slowing as predicted in Figure 5. , future conditions stay undetermined as the total number of infected individuals is close to the tip of inflection. Also, there is a possibility of new major outbursts to occur in the near future.

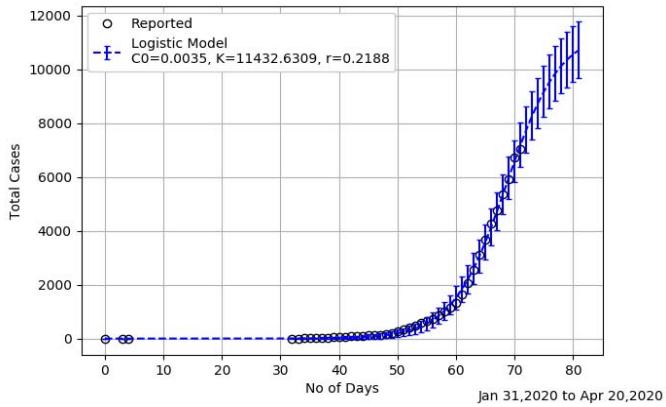


Figure 1. Estimation of Logistic Model to total cumulative confirmed cases in India since Jan31. The error bar is 95% CI extracted from 200 simulations using a poisson error structure.

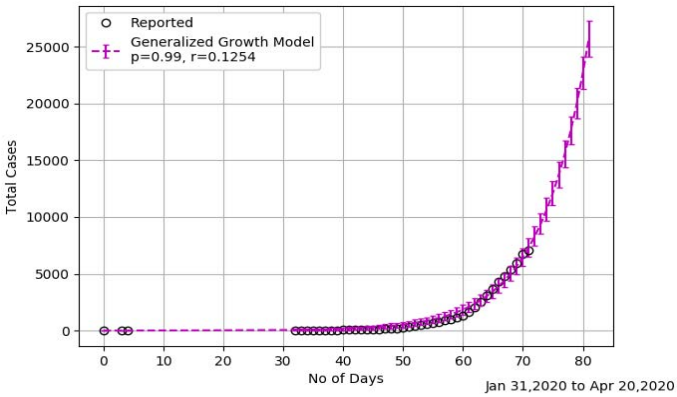


Figure 2. Estimation of Generalized Growth Model to total cumulative confirmed cases in India since Jan31. The error bar is 95% CI extracted from 200 simulations using a poisson error structure

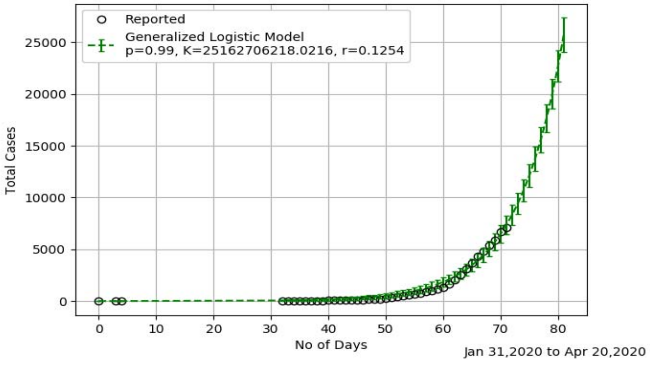


Figure 3. Estimation of Generalized Logistic Model in green to total cumulative confirmed cases in India since Jan31. The error bar is 95% CI extracted from 200 simulations using a poisson error structure.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:55:03 UTC from IEEE Xplore. Restrictions apply.

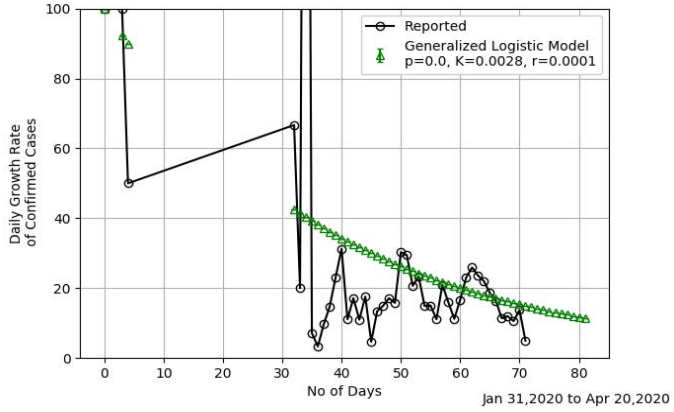


Figure 4. Estimation of models[3](*Generalized Logistic Development Model[1] (in green)*) to rate of daily development of infected population in India since Jan 31. The error bar is 95% CI extracted from 200 simulations using Poisson Error structure[1]

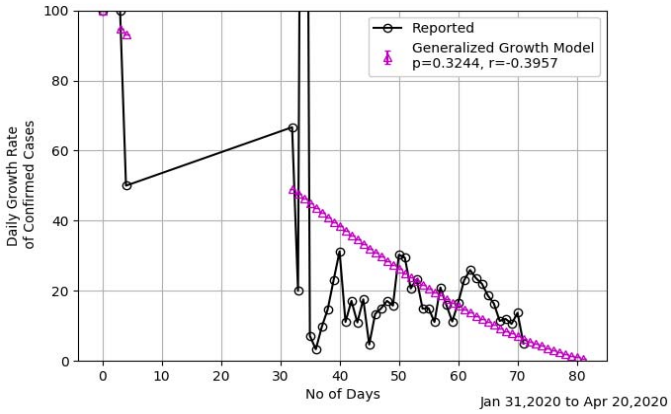


Figure 6. Estimation of models[3](*Logistic Development Model[1] (in blue*),Generalized Growth Model (*in pink*) to rate of daily development of infected population in India since Jan 31. The error bar is 95% CI extracted from 200 simulations using Poisson Error structure[1]

The probability is that the growth rate can keep on fluctuating at current levels for sometime, where the generalized growth model tends to be more favorable for drawing inferences. In this case, the infected population will increase to 13735 (95% CI: [13499, 13970]) in 5 days (April 15th), also to 25706 (95% CI: [25377,26034] within next 10 days (April 20th).

Figure 7. predicts that the point of inflection was passed a few days back, and also the daily number of confirmed infections has started decreasing. This seems to be a positive indication suppose the disease to break out following the logistic depreciation in cases in the following days, presuming the absence of other major public transmission, as was witnessed during phase III of most hotspots in China's research [1].

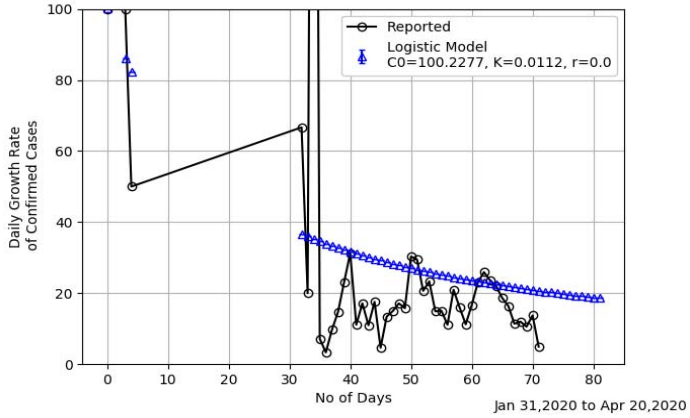


Figure 5. Estimation of models[3](*Logistic Development Model[1] (in blue*),Generalized Growth Model (*in pink*)to rate of daily development of infected population in India since Jan 31. The error bar is 95% CI extracted from 200 simulations using Poisson Error structure[1]

The daily development in the infected population is likely to deteriorate from this junction with an estimated daily surge of 603 (95% Confidence Interval:[550,655]) in 10 days(Apr 20) as predicted by the Logistics Model.

A pessimistic result is obtained from the Generalized Logistic model and Generalized Growth model. Both models estimate the daily increase of infected cases to be : 1355 (95% Confidence Interval: [1279, 1431]) from GLM and 1780 (95% Confidence Interval: [1697,1862]) from GGM, both of which could be predicted to be attained within 10 days

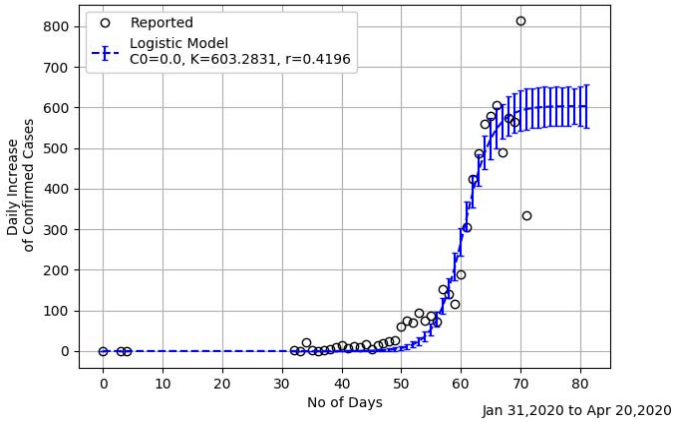


Figure 7. Estimation of models[3] (*Logistic Development Model[1] (in blue)*) to daily increase of infected population in India since Jan 31.The error bar is 95% CI extracted from 200 simulations using Poisson Error structure[1]

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:55:03 UTC from IEEE Xplore. Restrictions apply.

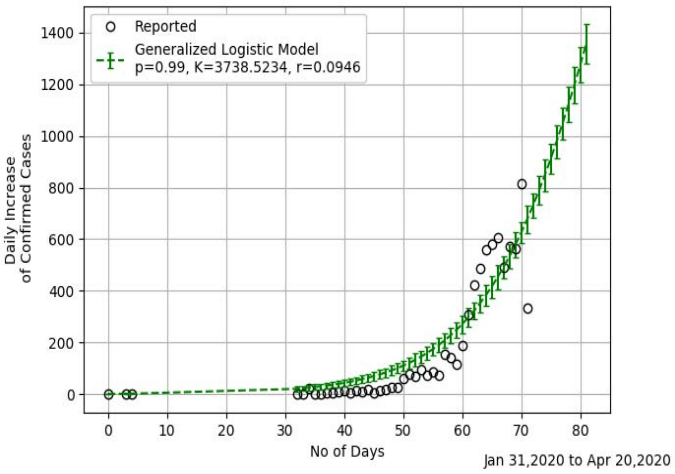


Figure 8. Estimation of models[3](*Generalized Logistic Development Model*[1] *(in green)*) to daily increase of infected population in India since Jan 31. The error bar is 95% CI extracted from 200 simulations using Poisson Error structure[1]

As shown in Figure 4. and Figure 6. the daily growth rate of the confirmed positive cases is definitely not growing exponentially, because the growth rate reduced notably from 40% in February to 1.5% by April 10.The initial period of the Indian outbreak occurred on Jan 30, where most cases are admitted to have a clear travel history to infected places.Although, it is observed from the results that the surge in figures is approaching a constant trend, the uncertainty in future cases might occur.

IV. CONCLUSION AND FUTURE WORK

In this research paper, uniquely introduced variations of logistic models to data of covid-19 pandemic in India and have employed a special 3 parameter Generalized Logistic Model and 2-parameter Generalized Growth Model to model the spread of this outbreak in India. Levenberg–Marquardt algorithm is used for parameter estimation to carry-out least square nonlinear minimization for curve fitting in Python and have calculated confidence intervals with 95% average accuracy, using bootstrap resampling method , where 200 simulations are performed using Poisson error structure. For some days, the number of total cases is predicted with more than 97% of accuracy, which shows the reliability of the models constructed.

With this research paper, tried to model the following growth models: (1)logistic growth development model, the (2)generalized growth development model,(3)generalized logistic growth model to the actual infected population of Covid-19 pandemic from Jan. 30 to Apr 10 of India. Our analysis tries to discover the growth of the covid-19 pandemic in India and the influence of National Lockdown at the average level.

On the grounds of experience in China[1], a high level risk in India is predicted with approximate total actual cases on April 14 to be 12116 (i.e. 95 % CI: [11895, 12336]), 25706 (i.e. 95 % CI: [25337, 26034]) by April 20.The GLM parameter p is around 0.99 (95% Confidence Interval: [0.98, 1]). This indicates that domestic transmission sources in India are out of control as growth is nearing an exponential one.On the contrary, the best case scenario would be the one described in the logistic model, giving the total infected population as 9547 (95% CI: [9349, 9746]) after 5 days (April 15), and 10723 (95% CI: [10514, 10933]) after 10 days. It might be a very good result if the actual infected cases follow the logistic model.Extrapolation of data for India has helped to determine the course of future measures that need to be taken in order to curb the development in COVID-19 disease in India. National lockdown has definitely helped to curb the surge in positive cases in India, but the infection might start spreading rapidly if the lockdown is lifted without prior precautions.

Accuracy of results depends proportionally on the size of the sample data collected. In our research, modelling logistic growth models for COVID-19 pandemic in its most nascent early stages, inferring that the size of sample data available for data extrapolation is very small. Hence, as the sample data size is very limited, saw differences between actual and predicted forecasts using the proposed method.Some external factors also might have resulted in mis-reporting of actual cases.

Table I: Short Term Forecast for daily increase of confirmed cases as predicted by GLM along with actual data and accuracy(%)

Day	Date	Actual	Predicted	Accuracy (%)
72	11 April 2020	854	677	79.27
73	12 April 2020	758	730	96.30
74	13 April 2020	1243	787	63.31
75	14 April 2020	1031	847	82.15
76	15 April 2020	886	911	97.25
77	16 April 2020	1061	978	92.17
78	17 April 2020	922	1188	77.6
79	18 April 2020	1371	1268	92.48
80	19 April 2020	1580	1274	80.63

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:55:03 UTC from IEEE Xplore. Restrictions apply.

Under-testing due to unavailability of testing kits in India might have also resulted in mis-reporting in actual recorded cases, which further resulted in deviation from actual figures in India.

Also, new hotspots might have originated in most affected states like Maharashtra, which might have led to sudden disruptions in the number of reported cases in April 2020.

Regression convergence depends significantly on preliminary values of parameters K, r and p, it might fail to converge accurately for a limited sample dataset.

Therefore, logistic models are highly data-driven in nature and the accuracy in predicted data depends on how well can the data collected mimic the actual dynamics of the pandemic.

The logistic models constructed mimic both the spread of epidemic and also the control in epidemic (S-shaped curve). Curve starts to converge after inflection point, predicting the control in the epidemic. Non-convergence of logistic models implies that the pandemic is not in control and further measures need to be undertaken.

REFERENCES

[1] Ke Wu1,2, Didier Darcet3, Qian Wang4 and Didier Sornette. Generalized logistic growth modeling of the COVID-19 outbreak in 29 provinces in China and in the rest of the world. Preprint: https://www.medrxiv.org/content/10.1101/2020.03.11.20034363v1

[2] Chowell et al., 2007; Chowell et al., 2006a. Transmission dynamics of the great influenza pandemic of 1918 in Geneva, Switzerland: Assessing the effects of hypothetical interventions. Preprint at: https://www.sciencedirect.com/science/article/abs/pii/S0022519305005096

[3] Junling Ma · Jonathan Dushoff · Benjamin M. Bolker · David J.D. Earn. Estimating Initial Epidemic Growth Rates. Preprint at: https://link.springer.com/content/pdf/10.1007/s11538-013-9918-2.pdf

[4] COVID-19 Patient Database for India available at https://docs.google.com/spreadsheets/d/e/2PACX-1vSz8Qs1gE_IYpzlkFkCXGcL_BqR8hZieWVi-rphN1gfrO3H4lDtVZs4kd0C3P8Y9lhsT1rhoB-Q_cP4/pubhtml#

[5] A Literature Review of 2019 Novel Coronavirus During the Early Outbreak Period: Epidemiology, Causes, Clinical Manifestation and Diagnosis, Prevention and Control.

[6] Tian-Mu Chen, Jia Rui, Qiu-Peng Wang,Ze-Yu Zhao, Jing-An Cui, Ling Yin;A mathematical model for simulating the phase-based transmissibility of a novel coronavirus. Preprint at: **https://doi.org/10.1186/s40249-020-00640-3**

[7] Xinguang Chen, Bin Yu; First two months of the 2019 Coronavirus Disease (COVID-19) epidemic in China: real-time surveillance and evaluation with a second derivative model. https://doi.org/10.1186/s41256-020-00137-4

[8] Poudel Adhikari, S.; Meng, S.; Wu, Y.; Mao, Y.; Ye, R.; Wang, Q.; Sun, C.; Sylvia, S.; Rozelle, S.; Raat, H.; Zhou, H. A Literature review of 2019 Novel Coronavirus During the Early Outbreak Period: Epidemiology, Causes, Clinical Manifestation and Diagnosis, Prevention and Control. *Preprints* **2020**, 2020020060 https://www.preprints.org/manuscript/202002.0060/v2

[9] World Health Organization article(WHO); Naming the coronavirus disease (COVID-19) and the virus that causes it https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it

[10] Cécile Viboud, Lone Simonsen, Gerardo Chowell; A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks https://doi.org/10.1016/j.epidem.2016.01.002

[11] Epidemiology article on wikipedia: https://en.wikipedia.org/wiki/Epidemiology

[12] Bootstrap resampling method; Medium Article https://towardsdatascience.com/an-introduction-to-the-bootstrap-method-58bcb51b4d60

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:55:03 UTC from IEEE Xplore. Restrictions apply.

Int. j. inf. tecnol.
https://doi.org/10.1007/s41870-020-00462-4

ORIGINAL RESEARCH

Multi-dimensional aggregation: a viable solution for interval data

Shailender Kumar[1]

Received: 22 November 2019 / Accepted: 21 April 2020
© Bharati Vidyapeeth's Institute of Computer Applications and Management 2020

Abstract Now a day multi-dimensional data modeling and aggregate query processing which are key assets of business intelligence solutions are being frequently realized to the unorthodox data. For interval values which are recorded when the data is on hold, multidimensional aggregation is the only viable solution and the author emphasizes over this aspect in this paper. Actually, such intervals reflect the state of reality of either current data or such data which were part of the present database. Every possible challenge which interval data throws upon is resolved in this paper through introduction of aggregation operator. Although the intervals are unknown at first but they eventually depend on the actual data and it turns out to be quiet handy while associating them with the resulting tuples. Only those result groups are selected for this purpose, which are specified partially. The interval data signifies that data holds either for each interim in the interval or entire interval and in both of these two cases it faces contention with the operators. In this paper, the author presents the empirical analysis of the aggregation operator after its implementation over the huge industrial data sets and claims that it holds an edge over the other temporal aggregation algorithms.

Keywords Aggregates · Databases · Query processing · Static · Stable

1 Introduction

In most of the present reality database employments, e.g., in the fiscal, medicinal, and systematic spheres, handle time-oriented data, which is raw facts with linked time-phases that holds certain time-varying feature of the facts, classically at the point when the certainties were or is right in the displayed veracity or when the realities was or is section of the contemporary condition of the database. In distinction to this, contemporary database administration delivers valuable tiny integral query language support for time-varying facts administration [1]. In phase through the growing dispersion of professional aptitude, collective calculation turn out to be significant. An aggregate operator transfigures a reflection temporal arrangement into a prompt upshot arrangement [2]. Ordinarily, this is accomplished by principal part the contention transient connection into bunches of lines with undistinguishable esteems for more than one components, then appertain an aggregate procedure, e.g., max or min, to every cluster in one go. For time-traverse esteemed time-arranged databases, total is exceptionally testing on the grounds that the time-traverse esteems can likewise be utilized for portraying the bunches of contention columns [3].

In this paper, the author proposes a novel time-oriented aggregation operator which is employed to perform multi-dimensional aggregation [4]. It improves a decent variety of prior proposed conglomeration operator and gives multi-dimensional help to two periods of total: (a) the elucidation of point assemblies for which to report more than one amassed esteems and (b) the elucidation of amassing assemblies, i.e., gathering of transient contention pushes that are associated with the item bunches and over which the time-situated total methods are assessed [5]. The author constructs the elementary structural chunk on

✉ Shailender Kumar
shailenderkumar@dce.ac.in

[1] Department of Computer Science and Engineering, Delhi Technological University, Delhi, India

Published online: 02 May 2020

contemporary developments in multi-dimensional query process and is the central exertion to dynamism these approaches to phase-esteemed data that are proven by time-oriented facts [6]. The paper covenants a well-organized employment of the aggregation operator with an average density of order $n \log z.$, where z signifies the number of product clusters and n signifies the figure of time-oriented aspects rows [7]. In investigational estimations on huge records, the time-oriented aggregation operator display nearly undeviating performance.

Aggregation of time-oriented facts presents complex trials [8]. Most significantly, the stint-phases of product rows can hinge on the authentic records and are not acknowledged in advancement [9]. As a result, the clustering of the artifact tuples can only be quantified moderately. Following, accumulation should provision what is named persistent, flexible, and fissile semantics of the connotation among records and time-intervals [10]. For instance, the quintessence of a retro juncture with a record adjust is static, hugeness that the adjust grasps for every single sub-periods of the traverse [11]. In deviation, contemplate the association of an exacting month with the numeral of periods operated by a demonstrative amid that period, e.g., 150 periods. The stint-relationship is malleable, by way of 150 periods are well-thoroughly deliberated a total of the periods functioned by the wage earner amid every day amid that month. A stint-relationship is fissionable if the realities can't be connected with changed time–values [12]. All techniques so far offer help for static practices. In conclusion, a period arranged accumulation final product may be more noteworthy than the contention time-situated connection [13]. Explicitly, for instant time-oriented aggregates that are gathered to form a static span, the product stint-relation scope can be double that of the time-oriented reflection [14]. To enumerate the product scope, the paper outlines the concept of a time-oriented aggregation aspect; and to regulate the time-oriented aggregation, the capability to state immobile time-span for the product rows is comprised [15].

The paper is organized in the following manner: Sect. 2 crisply explicates the literatures available in the domain of temporal databases and temporal aggregate operators. Section 3 includes key concepts and basic parameters. In Sect. 4, the author explains and introduce aggregation operator. Section 5 gives a brief overview about the employment of the operator. Section 6 reports on an empirical evaluation and Sect. 7 sum up the things by providing the concision the beneficiation of this article.

2 Background

The notion of hasty as well as amassed time-oriented aggregate operators has been, as of now, expressed by a few authors [16]. The importance of a rapid time-oriented amassed at stint 'st' is assumed by capturing the cluster of rows that confiscation at 'st' [17]. The implication of an amassed time-oriented accumulative at stint 'st' is assumed by capturing the cluster of rows that seizure in the progressive span [st − s, st], s ≥ 0. Misty amassed upshots with uninterrupted stints are amalgamated into static time spans. Most analysis has been improved the situation moment time-arranged totals, e.g. [4, 6], and accumulative aggregates [5, 18] and time-oriented aggregates with auxiliary succession predicates [7, 14] have acknowledged slender contemplation.

An introductory plan by Tuma [8, 19] for presuming time-oriented aggregates entails two examinations of the response stint-oriented schema—one for the assessment of the stint-interims of the item lines and the assessment of the time-arranged totals. A scheme by Kline and Snodgrass [4, 20] look at the info time-arranged connection just a single time, development of an accumulation sequencing tree in principle memory. In the interim, the sequencing tree isn't steady, the most pessimistic scenario time intricacy is $O(n^2)$ where cardinality of a relation is n. A rise, giving under the indistinguishable most pessimistic scenario unpredictability, is the k-ordered conglomeration sequencing tree [21]. A methodology deeds fractional congregation of columns for junk get together of timeworn protuberances.

A steady tree, for collection in principle memory, is grounded on the requested entire sign of time [4, 22]. This elucidation performs for the total capacities like summation, average, and count; for minimum and maximum, a merge-sort like procedures are planned. The two calculations have a most pessimistic scenario. For minor stockpiling, a procedure of bucket calculation is advocated over the information columns to mic-buckets giving to a dividing of the time reflections and furthermore figures out how to pay for enduring lines particular administration. Conglomeration is then accomplished on every single mic-bucket in partition. The calculation includes right to utilize the complete databank thrice.

The disk-grounded calculation, preservation of prompt and cumulative time-oriented conglomeration [6, 23] has already been carried out. A SB-tree encompasses a step of traverses related with modestly figured totals. With the SB-tree, total questions are practical to a complete base assembly—it isn't conceivable to grasp gathering establishes. The multilevel-styled SB-tree [23] emphases on arrangement total explorations amalgamated with go

predicates. The two strategies are focused on to spread grounds, the stint-interim of the info row is erased; thus the outcome isn't a stint situated connection.

The overall strategies portions three aspects. To begin with, the time-arranged gathering process combines the subdividing of the course of events with the array of the info lines [21]. The course of events is isolated into ranges, and an info push fit into an express board if its entire sign of time associate that panel. Second, the item pushes are very much characterized for stint-focuses and not after some stint-interims. Third, they permit the exploitation of at utmost one non-time-situated trait for time-arranged congregation.

Our opportunity arranged collection administrator, which spreads the multi-dimensional crease administrator to engage time-situated total, overpowers these limits and rearranges the conglomeration administrators pondered [3, 16]. It de-couples the isolating of the timetable from the collection of the info columns, consequently allowing to require item pushes over likely covering ranges and to control the degree of the item connection. Furthermore, it bolsters various characteristic highlights [17]. For a viable commercial the creators deed the classification of the info connection.

3 Preliminaries

3.1 Temporal aggregates

The paper commences a divergent stint sphere, where the modules are entitled stints (or stint-facts), well-found with a whole request $<$ T, and the author employs these as our stint-sphere. A whole mark of stint (or stint-facts) is a curled cluster over the stint-sphere and is signified by two stints, $[T_{start}, T_{end}]$, signifying its universal initial and absolute end facts, consistently. The author employs T as shorthand for [T_start, T_end]. For whole mark of time, the author presents numerous schemas: st $\in$ sT represents the stint 'st' is incorporated over the whole mark of stint sT. For two whole mark of stint sT and sT′, sT′ $\subseteq$ sT iff every stints in sT′ are also included in sT, and sT $\cap$ sT′ yields the group of stints over the whole mark of time. If sT $\cap$ sT′ $\neq \varnothing$, the paper states that the binary segments transect.

A time-oriented schema is a multi-tuple Schema = (ψ, ς, dom), where ψ is a determinate regular group of time-oriented attributes, ς is a determinate group of spheres, and sphere: $\rightarrow \varsigma$ is a procedure that links a sphere with every time-oriented attribute. A time-oriented schema is a schema with single whole mark of stint valued attribute (the sphere of whole mark of time corresponds to ς). For the determination of this paper, the author expresses time-oriented stint-relation schemas $V = (w_1, w_2, w_3, \ldots, w_n, \text{Timestamp})$ and $Z = (u_1, u_2, u_3, \ldots, u_n)$. Note that the statement that time-oriented schemas have a whole mark of time attribute. Timestamp is only for appropriateness. There is no concealed time-oriented property, and all clarifications are parameterized with a settled characteristic. Generally, the retitle administrator μ can be utilized to direct schemas as reasonable.

3.2 Attribute characteristics

The paper differentiates amongst three semantics of the stint-relationship of a standard attribute with a whole mark of time attribute. For a time-oriented stint-relation with schema $(w_1, w_2, w_3, \ldots, w_n, Timestamp)$ the attribute features. Timestamp are assumed as $vt = (v_1, v_2, v_3, \ldots, v_n)$, where $v_i \in \{v, m, a\}$. The values v, a, and m represent persistent, flexible, and nuclear features, correspondingly. For example, $vt = (v, m)$ for the schema (X, Y, Timestamp) represents that X has a persistent features and Y has a flexible features. If numerous time-oriented properties are employed, e.g., substantial stint and exchange stint, a standard characteristics enhanced the highlights over more than one stint. The rest of the paper utilizes just a single stint property T, and V denotes to the features associated with sT.

Definition 1 (*Modification of aspect values*). Let $t = (v_1, v_2, v_3, \ldots, v_n|T)$. be a super tuple over time-oriented schema $(w_1, w_2, w_3, \ldots, w_n, \text{Timestamp})$., U be a whole mark of time, and let $Z = (u_1, u_2, u_3, \ldots, u_n)$. be the aspect value features. The alteration in the aspect values is demarcated as streams:

$$\begin{aligned} &\text{adjct}\ (t, U, Z) = (\text{adjct}\ (t.w_1, t.T, U, u_1), \\ &\ldots, \text{adjct}(t.w_n, t.T, U, u_n), U). \end{aligned} \tag{1}$$

Considering the features Z, the adjct procedure alters every standard aspect value of t to the stint-sphere U and produces the tuple which associates T. For example, for the tuple (Apr, 1000, [2004/03, 2004/12]), the features (v, m), and the time-sphere [2004/05, 2004/06] the adjct procedure yields (Apr, 1000, [2004/05, 2004/06]).

4 Multi-dimensional aggregate operator

4.1 Aggregation clusters

A common time-oriented conglomeration operator should handle the dimension of two orthogonal individualities of conglomeration. Categorization of the artifact assemblages

is done on the basis of fallouts in time-oriented aggregate. The clusters of rows which are named as aggregation cluster significantly represents the stint-relationship with product cluster. Every product cluster can be characterized as a row in a time-oriented stint-relation R with schema $(k_1, k_2, k_3, \ldots, k_n|T)$.. Each k_i. is an element of the time-oriented stint-relation which reflects the reflection of the conglomeration operator; tuples undertake standards from k_i. that ensue in the reflection schema. For the whole mark of time, there are two scenarios: static spans and stable spans. With static spans, the whole mark of time attribute undertakes as values the utmost, non-intersecting spans over which the group of reflection rows is persistent. Intervals to aggregate are represented in Fig. 1.

Definition 2 (*Static spans*). Let v be a time-oriented stint-relation with whole mark of time attribute T. The paper describes the static span of v as:

$$\mathrm{SS(v)} = \{\mathrm{sT}|\forall \mathrm{v} \in \mathrm{v}(\mathrm{v.sT} \supseteq \mathrm{sT''} \vee \mathrm{r.sT} \cap \mathrm{sT'} = \emptyset) \wedge \forall \mathrm{sT'} \supset \mathrm{sT}\left(\exists \mathrm{v} \in \mathrm{v}\left(\mathrm{v.sT} \not\supseteq \mathrm{sT''} \wedge \mathrm{r.sT} \cap \mathrm{sT'} \neq \emptyset\right)\right)\Big\}. \tag{2}$$

The leading portion of the equation confirms that product spans do not cross margins of reflection spans and the subsequent portion of the equation confirms that the product spans are uttermost. The static spans for any time-oriented stint-relation are clustered by its sub-group.

Theorem 1 (Cardinality of static spans). *For a time-oriented stint-relation v with m rows, $m > 0$, the cardinality of static spans is partial by the following equation*:

$$|SS(v)| \leq 2m - 1 \tag{3}$$

Proof The end-points of the rows with whole mark of time can be undeviatingly wl-ordered giving at most 2 m time-points. 'm' time-points yields at most m − 1 successive time spans.

Definition 3 (*Stable spans*). Let v be a time-oriented stint-relation with whole mark of time attribute T. The Stable-span(v), has been described as a user-specified group of time-stamps, that fulfills thetandard temporal

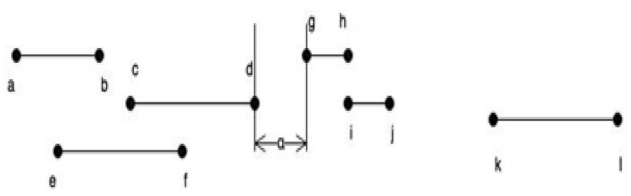


Fig. 1 Intervals to aggregate

statements. These statements describe that stable spans must overlaps with other spans in the time-oriented reflection stint-relation. The categorical ordering of product clusters with stable spans permits control of the cardinality of the artifact schema.

Definition 4 (*Partially quantified product clusters*). A product cluster with schema $(k_1, k_2, k_3, \ldots, k_n)$ is moderately quantified iff the standard of the whole mark of time attribute is not quantified properly. The paper characterizes a partially quantified product row as $v = (k_1, k_2, k_3, \ldots, k_n, [*, *])$.

With partially quantified product clusters in place, the author pushes the conclusion of the product clusters with static spans into the algorithm for the estimation of the time-oriented multi-dimensional conglomeration operator. The incompletely evaluated item pushes are simulated to every single static span for the predictable conglomeration clusters. The cross connection from proclamation that assigns the huge information columns to the static spans is utilized certainly by the estimation of a calculation.

5 Employment of the aggregation operator

5.1 Notion and outline

The work of the collection administrator for static spans is grounded on the various articulations: on the off chance that we analyze the time-situated contention connection, which is very much requested by the span begin time estimations of the lines, whenever point 'pt' assess the item pushes in advance to the point 'pt' (supposing, no columns that begin next to 'pt' added to the item group). Consequently, time-arranged contention connection is being scrutinized, item bunch are framed, and obsolete-columns are also isolated from primary memory. Just the columns those are legitimate at point-stint 'pt'. For the utilization with stable spans, the entire characteristic of time of the item group is measured in the gathering connection. Thus, it isn't compulsory to protect the information pushes in primary memory, however can assess them and refresh the total esteems as we filter the information connection. Data flow of the employment is clearly represented in Fig. 2.

5.2 Algorithm for static span

The algorithm employed for aggregation has five input elements: the cluster stint-relation c, the time-oriented reflection stint-relation t, a list of aggregate procedures $P = \{ p_1, p_2, \ldots, p_n. \}$, a assortment base α, and trait features F. The yield is a period situated connection that incorporates c delayed by the estimations of the total

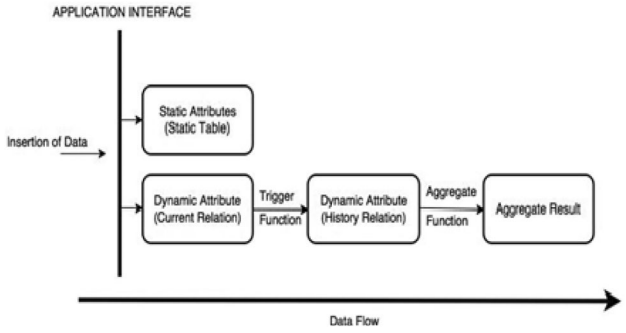


Fig. 2 Data flow

techniques in P. The calculation utilizes two sorts of information assemblies.

The initial step of the calculation is to instate the cluster reflection ct. In the event that c is a proj-reflection over relation, the gathering reflection is at first vacant and will be occupied while scrutinizing the time-situated contention rows. Else, ct is introduced with c, with the begin stint of the records settled to $-\infty$. In conclusion, a record over the non-fleeting properties is created. T succeeding advance is to assess the time-situated contention connection r sequentially on the premise of begin time of the lines. On the off chance that the gathering connection is a social variable based math articulation, we may need to spread the assemblage table with another section before as far as possible all product groups to which information tuple r contributes.

5.3 Algorithm for stable span

The algorithm for the assessment of administrator CT with stable traverse semantics. The key information configuration is the assemblage relation ct that holds the cluster connection c and has a supplementary section named Pi for each time-situated total methods Pi $\in$ P. The item groups, involving their entire sign of time, are entirely measured. The information rows cannot be kept in an end-point assemblage tree.

5.4 Complexity

For the complexity examination of time-oriented aggregate operator, the treatment of the information connection r is proper, which is isolated into four stages: query in the record, manufacture of result lines, refresh of the list and inclusion of the tuple at last point assemblage tree. A refresh list and the query related to the clter reflection have multifaceted nature and order of complexity $\log n_c$, the number of tuples present in the cluster relation is reflected by n_c. The development of a performance outcome push is undeviating as related to the uncovered columns. The support for different trait highlights infers at a cost. For every outcome, it needs a sweep over the entire E-point assemblage tree as well as a change in the characteristic esteems. On the off chance that exclusive static qualities were utilized, the total esteems could be assessed incrementally.

6 Experimental results and analysis

The author executed a number of trials with the static span and stable span algorithms, investigative their presentation for numerous backgrounds. All trials were track on an Intel core i3 with a 1.8 GHz processor and 2 GB memory. The author uses facts stint-schema that comprise from 20,000 to 1, 00,000 entries. The lifecycle of the facts stint-relations is about [0, 2^{20}]. The bunch connection holds one of a kind record. This is a vilest incident since all entire characteristic of time wind up on the comparable E-point assemblage tree. A cluster reflection with extra columns constructed over minor E-point assemblage trees and perform superior. Group partitioning scheme is represented in Fig. 3.

The primary trial examines the versatility of the static spans. The time intricacy depends on the quantity of unclog-tuples and the accumulation encouragement. Connection r-most noticeably awful with the major comprehensible number of unclog-tuples and the most extreme accumulation encouragement for static span has a handling stint that is quadratic in the extent of the information connection. For every other datum gatherings, the preparing stint divulges an undeviating execution. Connection r-measure up to have a collection perspective near 0 despite the fact that the quantity of open tuples is most extreme.

The routine of stable spans isn't misrepresented by crossing columns since the total result is computed augmentation. The indispensable limitation as far as execution is the quantity of upshot assemblages and the efficacy of the current query hone (for every single tuple in r, we should discover all bunches that gratify the standard state

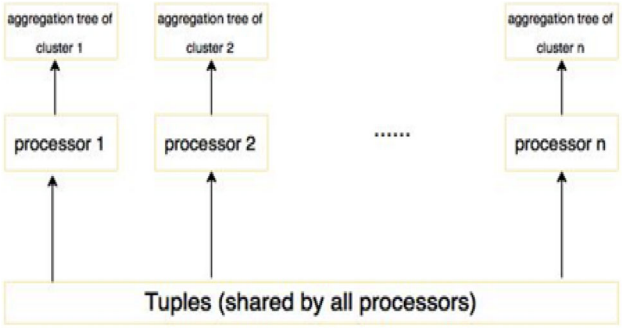


Fig. 3 Group partitioning scheme

and along these lines must be refreshed). Since we utilize AVL-trees for ordering item bunches, the execution increments steeply alongside the quantity of groups. In the event that we utilize hashing, the query time would be stable. In any case, the hashing just backs the fairness conditions.

The execution impact of variable key components on the algorithms utilized by the information connection r-rand is explored. The preparing time of dynamic number of open-tuples is high when contrasted with static one. The execution decays bit by bit since with adaptable properties, all open-tuples must be kept at last point tree. When a static span has been build up, the end-point tree is navigated, the eminence esteems are attuned, and the preceding total strategy is assessed. We have likewise fused the execution of a digressed static spans that backings static characteristics as it were. Static spans are incrementally figures out the totals, and its accomplishment is self-representing the quantity of unclog-tuples as in Fig. 4.

6.1 Static span versus stable span

The result of evaluating aggregates with static span semantics is represented in two ways: (1) with moderately quantified outcome clusters, and (2) a priori calculation of static spans by consuming SQL. The results confirm that static span with moderately quantified outcome clusters is certainly an effective way of evaluating aggregates with static span semantics. The performance of stable span declines as the number of clusters rises as clearly shown in Table 1. Nonetheless, up to an aggregation part of very nearly 46% Stable span outflanks static spans. In this manner, stable spans are effective for sensible aggregation elements, and it licenses to unequivocally control the aggregation factor.

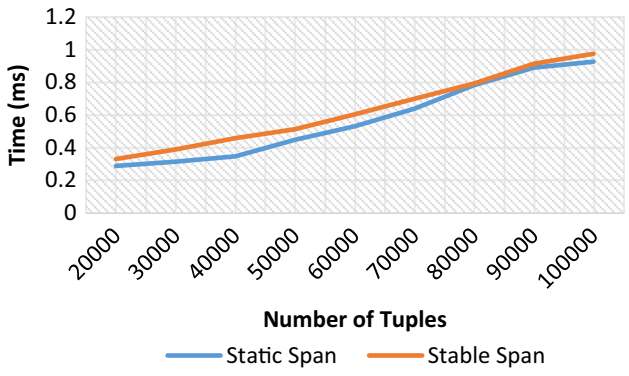


Fig. 4 Performance comparison of static span versus stable span

Table 1 Perfromance comparison of static span vs stable span at each tuple value

Static span		Stable span	
No. of tuples	Time (ms)	No. of tuples	Time (ms)
20,000	0.286	20,000	0.329
30,000	0.312	30,000	0.388
40,000	0.344	40,000	0.457
50,000	0.447	50,000	0.511
60,000	0.529	60,000	0.603
70,000	0.638	70,000	0.698
80,000	0.782	80,000	0.793
90,000	0.889	90,000	0.914
100,000	0.926	100,000	0.975

6.2 Static span versus balanced tree

The very last trial equates static span over the weighted balanced assemblage tree quantified in [3, 5]. The utmost operational procedure come into play, but prompt that it provisions summation, count, or average—it provisions stretchy aspects. Static span supports only static aspects and consents incremental calculation of aggregate standards. While stable span has the similar presentation as the balanced tree algorithm, the trials demonstrate that the provision for numerous stint-aspects features in static span is expensive and the DBMS should recognize circumstances which do not include flexible aspects and stable spans are employed. The recollection investigation of static span hinges on the numeral of unclog-rows.

7 Conclusion and future scope

This paper exhibits a novel time-oriented total administrator that controls most recent headway in multiple dimension query processing to relate with span-esteemed information. The conglomeration administrator streamlines an assorted variety of formerly proposed time-arranged collection administrators. It splits the portrayal of artifact stint-clusters from the explanation of conglomeration assemblages, i.e., the congregations of stint-oriented valued tuples that are linked with the artifact clusters and over which the time-oriented conglomeration procedures are evaluated. This prompts a definite informative system that licenses additionally to direct the extent of the result connection. Consequently, the conglomeration operator provision multiple time-oriented aspect features, embracing stretchy values which embrace aspect standards have to be attuned if the rows time-value altered. Lastly, the author delivers two dissimilar algorithms for the assessment of the

aggregation operator with static span and stable span, correspondingly. Comprehensive trial evaluated over the procedures which are climbable with reverence to facts group extent and associate well with other time-oriented aggregation algorithms. The appraisal additionally exposes that the help for several property determines at a cost. Future work includes various enhancement phases in static and stable span calculations, containing the following ones: improvement procedures for stint-relational algebra expressions that interrelate with the conglomeration administrator, the introduction of the cluster relation on the fly, indexes on the cluster relation, and a variety of the E-point assemblage tree that does not entail a completely arranged information connection.

References

1. Aljawarneh Shadi A, Radhakrishna V, Kumar V, Vinjamuri J (2017) G-SPAMINE: an approach to discover temporal association patterns andtrends in internet of things. Future Gener Comput Syst 74:430–443
2. Carafoli L, Mandreoli F, Martoglia R, Penzo W (2017) Streaming tables: native support to streaming data in DBMSs. IEEE Trans Syst Man Cybern: Syst 47(10):2768–2782
3. Cheng K (2016) Approximate temporal aggregation with nearby coalescing. In: International conference on database and expert systems applications. Springer, Berlin, pp 426–433
4. Ding W, Zhang S, Zhao Z (2017) A collaborative calculation on real-time stream in smart cities. Simul Model Pract Theory 72:23–33
5. Kumar S, Sharma N, Duggal R (2017) Veiled endpoints: a novel dimension for aggregation. Adv Comput Sci Technol 10(5):945–963
6. Dyreson CE (2003) Temporal coalescing with now granularity, and incomplete information. In: Proceedings of the 2003 ACM SIGMOD international conference on Management of data. ACM, pp 169–180
7. Freytag JC, Goodman N (1986) Translating aggregate queries into iterative programs. In: VLDB, vol 86, pp 25–28
8. Gordevičius J, Gamper J, Böhlen M (2012) Parsimonious temporal aggregation. VLDB J Int J Very Large Databases 21(3):309–332
9. Guzzo A, Pugliese A, Rullo N, Sacca D, Piccolo A (2017) Malevolent activity detection with hypergraph-based models. IEEE Trans Knowl Data Eng 29(5):1115–1128
10. Halawani SM, AlBidewi I, Ahmad ABR, Al-Romema NA (2012) Retrieval optimization technique for tuple timestamp historical stint-relation temporal data model. J Comput Sci 8(2):243
11. Kline N, Snodgrass RT (1995) Computing temporal aggregates. In: Data engineering, 1995. Proceedings of the eleventh international conference on, pp 222–231. IEEE
12. Kumar S, Rishi R (2016) Retrieval of meteorological data using temporal data modeling. Ind J Sci Technol 9(37):1–10
13. Kvet M, Matiako K, Kvet M (2016) Managing and storing function results in temporal approach. In: Open and big data (OBD), international conference on, pp 87–94. IEEE
14. Lopez IFV, Snodgrass RT, Moon B (2005) Spatiotemporal aggregate computation: a survey. IEEE Trans Knowl Data Eng 17(2):271–286
15. Marcellino M (1999) Some consequences of temporal aggregation in empirical analysis. J Bus Econ Stat 17(1):129–136
16. Moon B, Lopez IFV, Immanuel V (2003) Efficient algorithms for large-scale temporal aggregation. IEEE Trans Knowl Data Eng 15(3):744–759
17. Navathe SB, Ahmed R (1989) A temporal stint-relational model and a query language. Inf Sci 49(1–3):147–175
18. Snodgrass RT, Gomez S, McKenzie LE (1993) Aggregates in the temporal query language TQuel. IEEE Trans Knowl Data Eng 5(5):826–842
19. Terenziani Paolo (2016) Irregular indeterminate repeated facts in temporal stint-relational databases. IEEE Trans Knowl Data Eng 28(4):1075–1079
20. Yang J, Widom J (2001) Incremental computation and maintenance of temporal aggregates. In: Data engineering, 2001. Proceedings 17th international conference on. IEEE, pp 51–60
21. Zhang D, Markowetz A, Tsotras VJ, Gunopulos D, Seeger B (2008) On computing temporal aggregates with range predicates. ACM Trans Database Syst 33(2):12
22. Zhou X, Wang F, Zaniolo C (2006) Efficient temporal coalescing query support in stint-relational database systems. In: International conference on database and expert systems applications. Springer, Berlin, pp 676-686
23. Kumar S, Rishi R (2016) A relative analysis of modern temporal data models. In: Computing for sustainable global development (INDIACom), 2016 3rd international conference on. IEEE, pp 2851–2855

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Node Classification in Complex Networks using Network Embedding Techniques

Sanjay Kumar
Computer Science and Engineering
Delhi Technological University
Delhi,India
Sanjay.kumar@dtu.ac.in

Sanidhya Chaudhary
Computer Science and Engineering
Delhi Technological University
Delhi,India
sanidhya.chaudhary97@gmail.com

Saksham Kumar
Computer Science and Engineering
Delhi Technological University
Delhi,India
saksham.kmr.21@gmail.com

Raj Kumar Yadav
Computer Science and Engineering
Delhi Technological University
Delhi,India
rajkumar.raj.yadav.2015@gmail.com

***Abstract*—Node classification is a technique to assign node labels using the network topology and various node features. Many techniques like random walk-based method, label propagation graph regularizations are proposed in this research. Network embedding is a popular method to capture graph structure in low dimensional vector space and can capture the features of the nodes properly. In this paper, the popular embedding techniques are visualized on various data-sets and introduce the idea of network embedding techniques to perform the node classification task.**

***Index Terms*—Label propagation, Network embedding, Node classification, Online Social Networks, Random walk.**

I. INTRODUCTION

Node classification [1,2,3] is one of the most popular techniques, which is aimed to determine the labels of the vertices (nodes) of the network according to the other labelled vertices (nodes) and the structure of network graph. Many times, in a network, only a part of total nodes is labelled. There can be many reasons behind the unknown labels for a larger number of nodes so it becomes extremely essential to classify these nodes to acquire knowledge about their characteristics and behaviour. The labelled nodes and the links can be used to infer the missing labels.

The Node Classification problem can be stated as: providing a social network (or any other network structure) where some of the nodes are labeled and rest unlabeled, in what way estimate a good quality labeling for each node of the network. This problem is called the "node classification problem", fundamental problem being providing labeling for each node in a network.

Problem Statement:

Given a graph **G (V, E, W)**

V is the set of all nodes in the graph (—V— = n)

E is the set of edges in the graph

W is the weight matrix of the graph

Identify applicable funding agency here. If none, delete this.

Let **Vl** represent the labelled subset of V

Vu = (V - Vl) represents the set of unlabelled nodes

Let **Y** represent the set of all possible labels (be m)

Let **Yl** be the set of initial labels given on nodes of set Vl. To solve the problem, our method uses the information already available in the partly labeled network to predict labels for the rest of the nodes. Our goal is to train the classifier using the nodes that are pre labeled. Then the trained classifier is used to label the unlabeled nodes and also to nodes that already have some labels to augment their existing labels.

II. NETWORK OR GRAPH EMBEDDING

Transforming graph properties into vector space is Graph Embedding. Embedding must capture the geometric structure of graph, its vertex-to-vertex connection, and all other appropriate information about the graph and its vertices.

A. Methods used for Graph Embedding

1) DeepWalk[9]

DeepWalk is a random-walk based network embedding procedure. Random walks begin from a chosen node then moving to random neighbors at each step for a defined number of times. DeepWalk consists of 3 basic steps:

a) Sampling (Sampling of graph with random walks)

b) Training the skip-gram network (Skip-Gram network accepts node as one-hot-vector and maximizes the chances of predicting its neighbor nodes)

c) Computing the Embeddings (the hidden layer of the trained network is used as the embedding)

2) LINE [10]

Large-scale Information Network Embedding (LINE) defines two basic functions for first and second order proximity. These are explained below:

a) First Order Proximity refers to similarity between two nodes if they are connected by an edge.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:51:46 UTC from IEEE Xplore. Restrictions apply.

b) Second Order Proximity refers to similarity of two nodes' neighborhood's structure.

3) Node2Vec [11]

Node2Vec is similar to DeepWalk with the addition of biasing in random walks over the discovery of larger neighborhood or discovery of microscopic view around the node. Two parameters P and Q were introduced to control the probability of returning to the previous node and discovering the undiscovered part of the network respectively. All other steps (Training, computing Embeddings) is same as DeepWalk.

4) SDNE [12]

Structural Deep Network Embedding (SDNE) does not use random walks and is designed to preserve the first and second-order proximity on nodes in the embeddings.

SDNE uses an autoencoder neural net consisting of two parts Two (left and right) Vanilla auto encoders which accepts node adjacency vectors and learn the second order proximity.

5) Struc2Vec [13]

Struc2Vec embeds a network using node's structural identity. The idea is to construct a meta-graph which encodes the structural information of the graph of interest, and then use random walks from every node in this meta-graph to determine the contexts.

Struc2Vec has four basic steps:

a) Determine the structural similarity between every pair of nodes of the network, over different neighborhood sizes.

b) Construct a weighted multi layered graph where each layer of this graph relates to a level in hierarchy determining the structural similarity.

c) Use this multi layered graph to generate context for every node in the network based on biased random walks.

d) Use standard techniques to learn a latent representation from this context.

B. Graph Embedding Visualization

1) DeepWalk Visualization

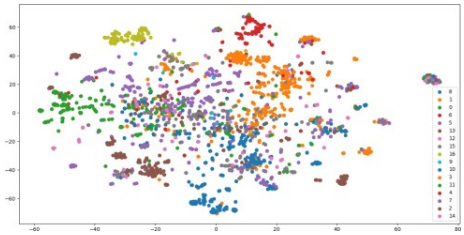


Fig 1. DeepWalk on wiki Dataset.

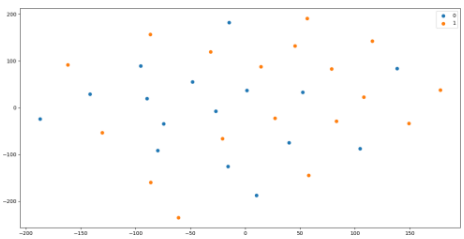


Fig 2. DeepWalk on karate Dataset.

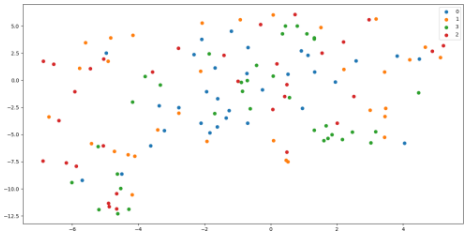


Fig 3. DeepWalk on Flight(Brazil) Dataset.

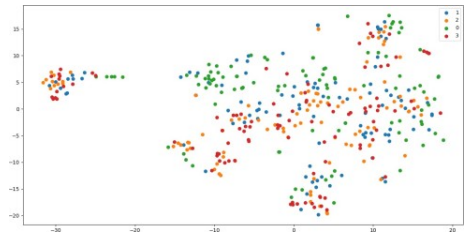


Fig 4. DeepWalk on Flight(Europe) Dataset.

2) LINE Visualization

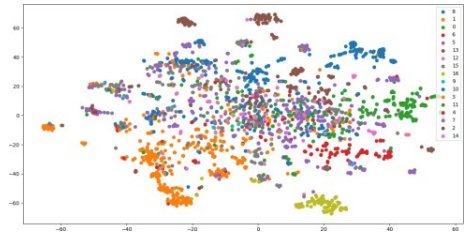


Fig 5. LINE on wiki Dataset.

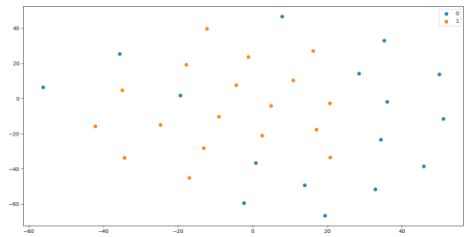
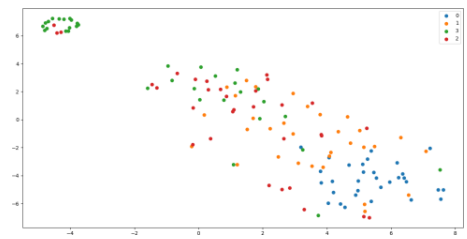


Fig 6. LINE on karate Dataset.



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:51:46 UTC from IEEE Xplore. Restrictions apply.

Fig 7. LINE on Flight(Brazil) Dataset.

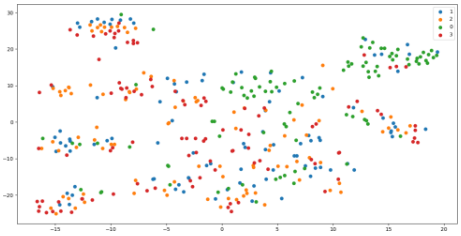


Fig 8. LINE on Flight(Europe) Dataset.

3) Node2Vec Visualization

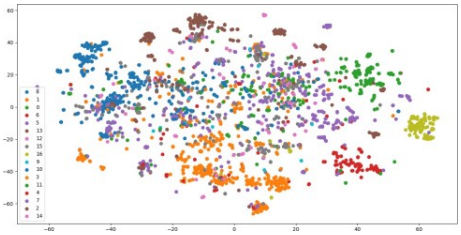


Fig 9. Node2Vec on wiki Dataset.

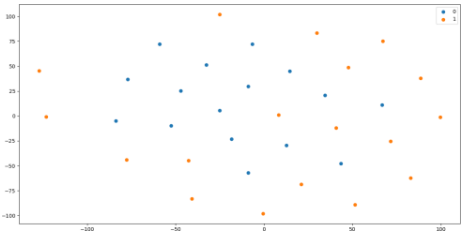


Fig 10. Node2Vec on karate Dataset.

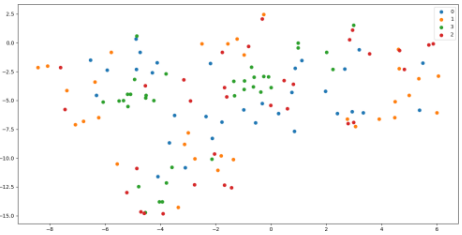


Fig 11. Node2Vec on Flight(Brazil) Dataset.

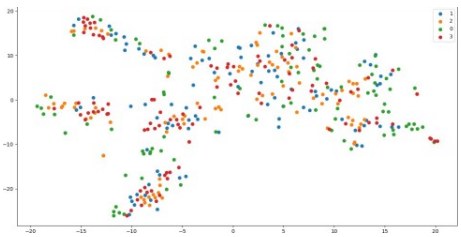


Fig 12. Node2Vec on Flight(Europe) Dataset.

4) SDNE Visualization

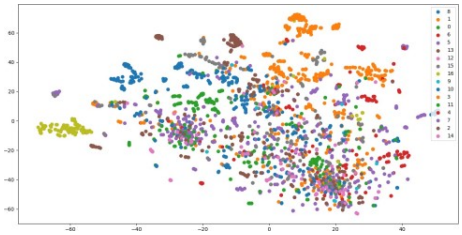


Fig 13. SDNE on wiki Dataset.

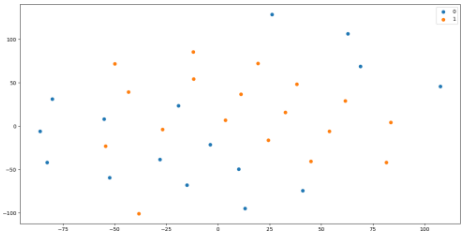


Fig 14. SDNE on karate Dataset.

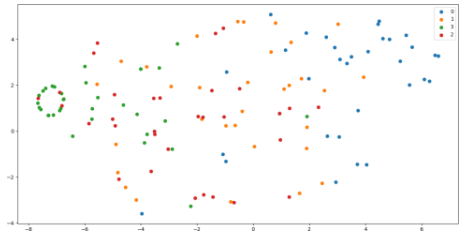


Fig 15. SDNE on Flight(Brazil) Dataset.

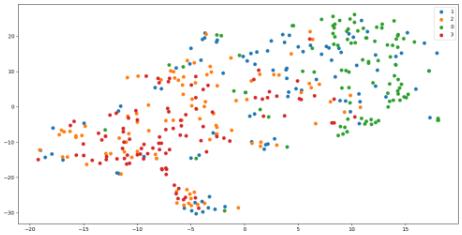


Fig 16. SDNE on Flight(Europe) Dataset.

5) Struc2Vec Visualization

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:51:46 UTC from IEEE Xplore. Restrictions apply.

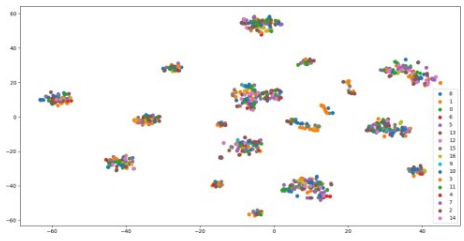


Fig 1. Struc2Vec on wiki Dataset.

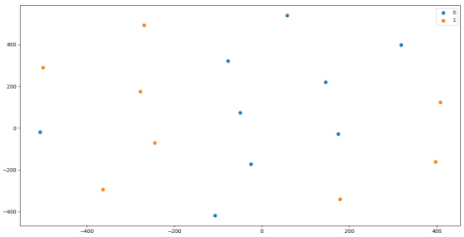


Fig 1. Struc2Vec on karate Dataset.

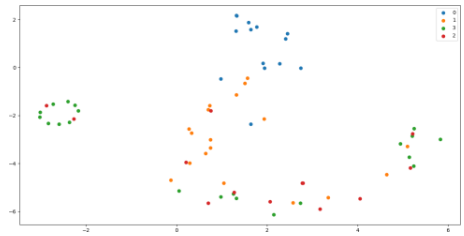


Fig 1. Struc2Vec on Flight(Brazil) Dataset.

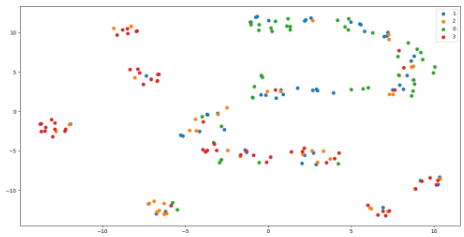


Fig 1. Struc2Vec on Flight(Europe) Dataset.

III. METHODOLOGY

A. Algorithm

Using the idea of network embedding techniques followed by label propagation and regression for classification and a proper mixture of test and training data, an algorithm for node classification is proposed which works as follows:

Input: Graph (G)
Output: Accuracy Score, F1_score

NodeClassification(G)
G ← DiGraph(edge_list)
model ← EmbeddingTechnique(G)
model.train(window_size, iter)
embeddings ← model.get_embeddings()
for embedding in embeddings
evaluate_embedding(embedding)
plot_embedding(embedding)
evaluate_embeddings(embedding):
X, Y ← read_node_labels(label _list)
tr_frac ← 0.8
clf ← classifier(embedding, clf = LogisticRegression())
clf.evaluate(Y_test, Y_predicted, trac_frac)
label_propagation(G)
label_prop_model←labelPropagation()
label_prop_model.fit(X_train,y_train)
prediction ← label_prop_model.predict(X_test)
evaluate(Y_test, prediction)

The graph is given as an input to the algorithm in step 2, which is converted to lower dimensions using different graph embedding techniques in step 3 and 4. Step 4 trains the model to evaluate embeddings. Plot_embedding is used to plot the reduced vector form in pictorial form. Evaluate_embeddings function in step 9 is used to classify nodes by logistic regression in which tr_fact is the fraction of pre-labelled dataset that is the number of nodes which are already labelled. Step 13 evaluates the predicted labels and gives accuracy and f1_score as output. Similarly, in function 'label_propagation' (step 14), 80% nodes are used to predict the labels of remaining nodes and again the predicted labels are evaluated in step 18.

A directed graph is constructed using the edge list for various datasets like karate, wiki etc and have used 80% of the nodes for training our model and then used it to classify the remaining 20% nodes using various methods. Initially, classified using label propagation in which each node checks for the labels of its neighbors and then adopts the label with highest frequency.

Then tried to classify nodes using feature- based methods and then applying logistic regression to classify them. In non-embedded graph the neighbor list is considered as the features of the node, however in embedded graph the dimensionality is used a feature list and then after the again using 80% of the classified nodes have tried to classify the rest 20% of the nodes. In order to classify using logistic regression, used OneVsAll classifier in which one classifier is fitted per class. Interpretability is the main advantage of this classifier as the inspection of classifier will give the knowledge of the class as each class is represented by one classifier.

A comprehensive comparison of various classification techniques is shown and suitability of different embedding techniques for different datasets and have also proved the scalability problem of old classifying methods as the size of dataset increases the gap between the evaluation score of embedded and non-embedded graph also increases substantially, however the evaluation score was quite compa-rable in smaller datasets.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:51:46 UTC from IEEE Xplore. Restrictions apply.

B. Evaluation Metrices

- **Accuracy score:** In classification, this function computes subset accuracy: the set of labels predicted for a sample must exactly match the corresponding set of labels in y true i.e. true labels.
- **F1 score** F-score or F-measure are the equivalent terms of F1 score. The weighted average of the recall and precision is interpreted as F1 score. The best value of F1_ score is 1 whereas worse is 0. Equal weightage is given precision and recall. Mathematically, it can be stated as:

F1 = (recall * precision) * 2 / (recall + precision)

a) Micro Metrics are calculated over the global domain by counting the total number of false negatives, false positives and true positives.
b) Macro Metrics are calculated for every label and their unweighted mean is found. The label imbalance is not taken into account

IV. Result

Evaluation and comparisons of node classification using both non-embedded and embedded graphs. The graph was embedded using different techniques mentioned above in Section 2.1. The metrics mentioned in Section 3.2 are used to evaluate and compare the results obtained.

Application of Label Propagation on various datasets without the use of any embedding techniques produced results shown in Table 1. The values obtained are one of the lowest showing limitations of Label Propagation on big datasets. Using Logistic Regression on non-embedded graphs shows slight improvement over previous Label Propagation, shown in Table 2.Various graph embedding techniques are used in the proposed algorithm and results are obtained showing improvements over previous approaches. Results with graph embeddings are shown from Table 3 through Table 8.

Traditional methods of node classification may not scale well with the increasing size of the network. This is also evident by observing the value of F1_score and accuracy below as they are comparable in smaller datasets such as karate and flight but as the size of the dataset increases the F1_score and accuracy of embedded graphs is much better than that of non-embedded graphs like in wiki data set, classification using logistic regression gave 32.6% better results when applied on embedded graph(LINE embedding technique) than non-embedded graph by considering neighbors as features. Classification of wiki data set using logistic regression on embedded graph gave 86.4% better results than label propagation on non-embedded graph. Similarly, on comparing the struc2vec embedding of flight (Brazil) dataset and found that the accuracy score is 61.36% better than that of non-embedded one.

TABLE I
Using Label Propagation on non-Embedded Graphs

DATASET	*MICRO*	*MACRO*	*ACCURACY*
Wiki	0.3775	0.13189	0.3776
Karate	0.3775	0.13189	0.3776
Flight(Brazil)	0.3775	0.13189	0.3776
Flight(Europe)	0.3775	0.13189	0.3776

TABLE II
Using Logistic Regression on non-Embedded Graph where neighbours are considered as features

DATASET	*MICRO*	*MACRO*	*ACCURACY*
Wiki	0.5218	0.3560	0.5218
Karate	0.5714	0.3636	0.5714
Flight(Brazil)	0.3529	0.1666	0.4444
Flight(Europe)	0.2702	0.2941	0.4875

TABLE III
Results obtained using DeepWalk for graph embedding

DATASET	*MICRO*	*MACRO*	*ACCURACY*
Wiki	0.6902	0.6070	0.6902
Karate	0.4285	0.3000	0.4285
Flight(Brazil)	0.1481	0.1491	0.1481
Flight(Europe)	0.4750	0.4706	0.4750

TABLE IV
Results obtained using LINE for graph embedding

DATASET	*MICRO*	*MACRO*	*ACCURACY*
Wiki	0.6299	0.5023	0.6299
Karate	0.4285	0.3000	0.4285
Flight(Brazil)	0.5185	0.4495	0.5185
Flight(Europe)	0.4375	0.4351	0.4375

TABLE V
Results obtained using Node2Vec for graph embedding

DATASET	*MICRO*	*MACRO*	*ACCURACY*
Wiki	0.6881	0.5945	0.6881
Karate	0.4285	0.3000	0.4286
Flight(Brazil)	0.3704	0.3216	0.3703
Flight(Europe)	0.4000	0.3983	0.4000

TABLE VI
Results obtained using SDNE for graph embedding

DATASET	*MICRO*	*MACRO*	*ACCURACY*
Wiki	0.6216	0.4859	0.6216
Karate	0.8571	0.8571	0.8571
Flight(Brazil)	0.5925	0.5527	0.5925
Flight(Europe)	0.5750	0.5660	0.5750

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:51:46 UTC from IEEE Xplore. Restrictions apply.

TABLE VII
RESULTS OBTAINED USING STRUC2VEC FOR GRAPH EMBEDDING

DATASET	*MICRO*	*MACRO*	*ACCURACY*
Wiki	0.1950	0.1100	0.1950
Karate	0.7500	0.7333	0.7500
Flight(Brazil)	0.7143	0.7222	0.7143
Flight(Europe)	0.4500	0.4428	0.4500

CONCLUSION AND FUTURE WORK

In this paper, various approaches to node classification techniques are analyzed using label propagation on non-embedded graphs and logistic regression on both embedded and non-embedded graphs. Node classification using logistic regression on embedded graphs gives better results than both logistic regression and label propagation on non-embedded graphs. Various embedding techniques work best for different data sets because of their methods of feature generation i.e., dimensionality. For example, struc2vec works best on flight networks which are identified by the network structure however LINE, SDNE and node2vec gave better results on large data sets such as wiki dataset. The results obtained after embedding are quite significant and give commendable results. The embedding techniques can be further used for link prediction i.e., predicting the missing links which can be used for building a recommendation engine, Graph reconstruction. The node classification efficiency can be further enhanced using a neural network, which can greatly capture the characteristic dynamics of the network, by training the network data-sets.

REFERENCES

[1] Goyal, P., Ferrara, E.: Graph embedding techniques, applications, and performance: A survey. Knowledge-Based Systems, 151, 78-94.(2018).
[2] Bhagat, S., Cormode, G., Muthukrishnan, S. : Node classification in social networks. In Social network data analytics (pp. 115-148). Springer, Boston, MA. (2011).
[3] Tang, J., Aggarwal, C., Liu, H.: Node classification in signed social networks. In Proceedings of the 2016 SIAM international conference on data mining (pp. 54-62). Society for Industrial and Applied Mathematics. (2016, June).
[4] Grover, A., Zweig, A., Ermon, S.: Graphite: Iterative generative modeling of graphs. arXiv preprint arXiv:1803.10459. (2018)
[5] Zhang, X. K., Ren, J., Song, C., Jia, J., Zhang, Q. : Label propagation algorithm for community detection based on node importance and label influence. Physics Letters A, 381(33), 2691-2698. (2017).
[6] Zhu, X., Ghahramani, Z.: Learning from labeled and unlabeled data with label propagation.(2002).
[7] Jokar, E., Mosleh, M.: Community detection in social networks based on improved Label Propagation Algorithm and balanced link density. Physics Letters A, 383(8), 718-727. (2019).
[8] Zhang, T., Popescul, A., Dom, B.: Linear prediction models with graph regularization for web-page categorization. In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 821-826). (2006, August).
[9] Perozzi, B., Al-Rfou, R., Skiena, S.: Deepwalk: Online learning of social representations. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 701-710).(2014).
[10] Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., Mei, Q.: Line: Large-scale information network embedding. In Proceedings of the 24th international conference on world wide web (pp. 1067-1077).(2015, May).
[11] Grover, A., Leskovec, J.node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 855-864).(2016, August).
[12] Ribeiro, L. F., Saverese, P. H., Figueiredo, D. R.: struc2vec: Learning node representations from structural identity. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 385-394). (2017, August).
[13] Ribeiro, L. F., Saverese, P. H., Figueiredo, D. R.: struc2vec: Learning node representations from structural identity. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 385-394). (2017, August).
[14] Kipf, T. N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907. (2016).
[15] Wang, D., Cui, P., Zhu, W: Structural deep network embedding. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1225-1234).(2016, August).
[16] Sun, G., Zhang, X.: Graph Embedding with Rich Information through Bipartite Heterogeneous Network. CoRR. (2017).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:51:46 UTC from IEEE Xplore. Restrictions apply.

Nonlinear Dyn
https://doi.org/10.1007/s11071-020-05762-9

ORIGINAL PAPER

Nonlinear dynamics of a time-delayed epidemic model with two explicit aware classes, saturated incidences, and treatment

Kanica Goel · Abhishek Kumar · Nilam

Received: 28 April 2020 / Accepted: 10 June 2020
© Springer Nature B.V. 2020

Abstract Whenever a disease emerges, awareness in susceptibles prompts them to take preventive measures, which influence individuals' behaviors. Therefore, we present and analyze a time-delayed epidemic model in which class of susceptible individuals is divided into three subclasses: unaware susceptibles, fully aware susceptibles, and partially aware susceptibles to the disease, respectively, which emphasizes to consider three explicit incidences. The saturated type of incidence rates and treatment rate of infectives are deliberated herein. The mathematical analysis shows that the model has two equilibria: disease-free and endemic. We derive the basic reproduction number R_0 of the model and study the stability behavior of the model at both disease-free and endemic equilibria. Through analysis, it is demonstrated that the disease-free equilibrium is locally asymptotically stable when $R_0 < 1$, unstable when $R_0 > 1$, and linearly neutrally stable when $R_0 = 1$ for the time delay $\varrho > 0$. Further, an undelayed epidemic model is studied when $R_0 = 1$, which reveals that the model exhibits forward and backward bifurcations under specific conditions, which also has important implications in the study of disease transmission dynamics. Moreover, we investigate the stability behavior of the endemic equilibrium and show that Hopf bifurcation occurs near endemic equilibrium when we choose time delay as a bifurcation parameter. Lastly, numerical simulations are performed in support of our analytical results.

Keywords Full and partial awareness · Time delay · Nonlinear incidences and treatment rates · Bifurcations · Stability · Numerical simulations

Mathematics Subject Classification 34D20 · 92B05 · 37M05

K. Goel · Nilam (✉)
Department of Applied Mathematics, Delhi Technological University, Delhi 110042, India
e-mail: nilam.research18@gmail.com; rathi.nilam@gmail.com

K. Goel
e-mail: kanica.dtu2016@gmail.com

A. Kumar
Department of Mathematics, School of Basic Sciences and Research, Sharda University, Greater Noida 201310, India
e-mail: abhishek.dtu14@gmail.com; abhishek.kumar6@sharda.ac.in

1 Introduction

The last two decades have seen several large-scale epidemics outbreaks such as Ebola, SARS, Zika virus, and swine flu, which leads to low socioeconomic status and inadequate access to health care. People get information about these outbreaks quite quickly due to significant advances in social media, which can have an insightful effect on the actual epidemic dynamics [17,31]. Therefore, at the beginning of an epidemic outbreak, the initial step is to make the individuals aware of the disease and its preventive methods. Awareness programs can alert the susceptible population toward the

Published online: 04 July 2020

contagious disease [23]. If susceptible individuals have adequate knowledge and proper information about the infection, then they take precautionary measures such as regular hand sanitization, the use of face masks, wear hand gloves, vaccination, and even quarantine, which can reduce the impact of illness, for instance, ranging from the plague outbreak in the English village of Eyam in 1665–1666 [9], where the town completely sealed itself off to prevent further transmission of plague, to more recent outbreaks of swine influenza [17] and Ebola [31]. In the situation of rural health services, the challenge of government health-care scheme is that there are numerous gaps in primary health services, and the health-care facilities are mainly urban centric. The rural residents may have low health knowledge awareness, and their receiving way of health knowledge can be traditional and straightforward. Knowledge level increases with higher education levels; therefore, one of the main factors is their education level. Therefore, a lot of people are not fully but partially aware of the spread and control of infectious diseases. Due to partial awareness, some individuals often medicate themselves adopting antibiotics, even when advised against doing so which weakens their immune system makes them at a high risk of catching the infection [18]. Therefore, complete awareness about the cycle of disease, along with the utilization of appropriate precautions and adequate decontamination procedures, is vital. Full awareness of the disease in humans develops a habit of taking precautionary measures against it, and they follow the instructions given by health workers and government and lower their risk of becoming infected.

In the study of the transmission of infectious diseases, the incidence rate has a vital role as it determines the number of infectives per unit of time. In 1927, Kermack Mckendrick considered the bilinear incidence rate, which follows the law of mass action [1]. In the bilinear incidence rate, the number of infectives linearly increases, which might be real for a small population of infected individuals, but it is impractical for a large number of infectives. Therefore, several studies are devoted to consider nonlinear incidence rate for disease transmission dynamics (e.g., [2,3,15,16,26,34–40]. In the present study, we have incorporated the nonlinear saturated incidence rate, where the interaction term is of the form $Sg(I)$, $g(I) = \beta I/(1 + \varepsilon I)$. Here, βI measures the force of infection, and $1/(1 + \varepsilon I)$ measures the inhibition effect of the behavioral change of the susceptibles when their number increases and crowding effect of the infective individuals [2,22]. In this incidence rate, the number of infectives depends on the nonlinear bounded map $g(I)$, which tends to saturation level β/ε when I gets large. It prevents the unboundedness of the contact rate as it includes the psychological effects and, thus, more realistic than the bilinear incidence rate.

Treatment is vital to cure the infection and prevent the development of resistant bacteria. Therefore, consideration of the treatment rate in the epidemic model is of great importance. In 2004, Wang and Ruan [10] studied the SIR epidemic model with the constant treatment rate and showed various bifurcations. The consideration of constant treatment rate might be real when there are small infected populations because there are limited medical resources available in society. Therefore, in 2012, Zhou and Fan [24] improved the treatment rate by considering Holling type II functional response as given below:

$$h(I) = \frac{\beta I}{1 + \alpha I}, \quad I \geq 0, \quad \beta \geq 0, \quad \alpha \geq 0.$$

and explore the SIR epidemic model to understand the effect of the limited medical resources and their supply efficiency on the transmission of infectious diseases. Motivated by the work mentioned above, we deliberate the Holling type II treatment rate (also called saturated treatment rate) and study the effect of treatment rate with limited medical resources in the current epidemic model.

The inclusion of time delay in the study of epidemiology is an important aspect. Persons with asymptomatic infections play an essential role in the spread of infectious disease, especially as they are unaware of their infection and therefore take no special hygiene precautions. Thus, the study of epidemiology involves time delay, which needs to be considered for practical purposes. The inclusion of time delay may arise due to delays caused by the latency in a vector and delay caused by a latent period in the host [4]. Recently, many mathematical models studied the impact of time delay in their epidemic model [7,13,35–37,40].

The impact of information and awareness on the spread of epidemics has been studied by many authors [14,19–21,27–29,33,34]. Funk et al. [19] considered the aware susceptible and aware infected populations in their epidemic model, on the assumption that the

aware susceptibles will be at a lower rate of catching the infection than the unaware susceptibles, and studied that disease dynamics in a well-mixed population. Kiss et al. [20] studied the effect of information transmission on the dynamics of sexually transmitted diseases. They assumed that the entire population is aware of the risk of infection; however, only a specific portion decides to react by constraining their contact with contaminated people. Misra et al. and Dubey et al. [21,34] investigated the impact of awareness programs on the transmission dynamics of infectious diseases in their nonlinear epidemic models. Some researchers studied the impact of awareness on the disease transmission dynamics along with the influence of time delay. Zuo et al. [29] introduced a time delay in the media variable to emphasize the delay in reporting cases of infections. Zhao et al. [27] studied the SIRS epidemic model by incorporating time delay in media coverage. Zuo and Liu [28] studied the effect of awareness programs driven by media and the delay on the prevalence of the infectious disease. In all these models, the study reveals that the disease-free equilibrium (DFE) is stable when a basic reproduction number R_0 is less than one, unstable when $R_0 > 1$, irrespective of the value of the time delay, and has a stable endemic equilibrium for time delay equals to zero. In the study of Zuo et al. [29], Zhao et al. [27], and Misra et al. [21], the occurrence of Hopf bifurcation is shown for the particular value of the time delay. Greenhalgh et al. [33] included two delays, one in reporting of infected cases and another delay representing the loss of disease awareness after a fixed period. Their study reveals the reduction in infected individuals with an increment in the duration of awareness. It is also shown that both the time delays can produce oscillations and destabilize the endemic equilibrium.

The public is a coalition of many subgroups of individuals with vastly different social, educational, and economic backgrounds. During an outbreak, people adopt full or partial awareness, depending on the understanding of how they perceive risks, and communicate about the effectiveness of protective measures. Therefore, instead of going directly from susceptible to infected class, we introduce fully aware and partial aware susceptible compartments into the SIR epidemic model due to heterogeneous protection level and extend the epidemic model to include the behavioral change of susceptibles, which can change the transmission patterns and reduce the prevalence of disease to the more extent. The precaution level of susceptible individuals is heterogeneous as they may take different levels of precautions to protect themselves from being infected based on the severity of epidemics or their characteristics. We consider three specific nonlinear incidence rates of unaware susceptibles, fully aware susceptibles, and partially aware susceptible, respectively, with the inclusion of time delay as a latent phase having a fixed duration. Also, with awareness, treatment to infectives is essential to mitigate the infection. Therefore, we consider the nonlinear saturated treatment rate, which includes the fact of limitation in the availability of resources. We formulate the nonlinear time-delayed mathematical epidemic model and perform the stability analysis to demonstrate the eradication or persistence of the disease with the help of the basic reproduction number R_0 for the time delay $\varrho \geq 0$. The bifurcation theory approach using the center manifold theory is performed, revealing the existence of backward and forward bifurcations, which shows that reducing R_0 below unity is not sufficient to eradicate the disease due to the presence of backward bifurcation, which has an important implication in the study of disease dynamics. Further, choosing time delay as a bifurcation parameter, the periodic and oscillatory solutions appear via Hopf bifurcation. Moreover, the numerical experiments show the importance of considering full and aware susceptible individuals and nonlinear terms such as time delay, incidences, and treatment.

The rest of the manuscript is organized as follows: The model and its basic properties are presented in Sect. 2. In Sect. 3, the existence of disease-free and endemic equilibria is presented, and stability analysis is performed using the basic reproduction number. The conditions for the backward and forward bifurcations are derived. Also, the existence of stability switches is shown via the presence of Hopf bifurcation by choosing the time delay as a bifurcation parameter. In Sect. 4, numerical simulations are presented to signify the theoretical results. Lastly, Sect. 5 is devoted to a brief discussion.

2 Mathematical model and basic properties

We assume that N denotes the total constant population, and it is divided into five compartments according to the disease status: unaware susceptible class $S(t)$, fully aware class $A_F(t)$, partially aware class $A_P(t)$, infected individuals class $I(t)$, and removed individu-

als class $R(t)$. The unaware susceptible class consists of those individuals who are vulnerable to the disease and taking no precautions against it. A fully aware susceptible class involves those individuals who have adequate knowledge and proper information about the spread of the disease and taking precautionary measures against it. Partially aware susceptibles consist of those individuals who have an incomplete understanding of the spread and prevention of the disease and have low resources available to escape from the infection. Including a behavioral response among unaware susceptibles would unavoidably call for response among fully aware susceptibles and partially aware susceptibles. Therefore, we have considered three explicit same functional types of saturated incidence rates described below:

(i) $F_1(S(t-\varrho), I(t-\varrho)) = (\beta_1 S(t-\varrho) I(t-\varrho))/(1+\varepsilon I(t-\varrho))$, representing the saturated incidence rate among unaware susceptibles,

(ii) $F_2(A_F(t-\varrho), I(t-\varrho)) = (\beta_2 A_F(t-\varrho) I(t-\varrho))/(1+\varepsilon I(t-\varrho))$, representing the saturated incidence rate among fully aware susceptibles,

(iii) $F_3(A_P(t-\varrho), I(t-\varrho)) = (\beta_2 A_F(t-\varrho) I(t-\varrho))/(1+\varepsilon I(t-\varrho))$, representing the saturated incidence rate among partially aware susceptibles,

where ϱ denotes the time delay, representing the latent period. The latent period is defined as the period between exposure and infection, since the pathogen is present in a latent stage, without clinical symptoms or signs of infection in the host. The delay term in $S(t)$, $A_F(t)$, and $A_P(t)$ is introduced because people, who are unaware, fully aware, or partially aware of the disease may consider themselves in their respective classes after becoming infected. We assume that such individuals are in the latent period.

Let κ denote the constant rate of inflow of new unaware susceptibles due to the recruitment of new members by the current members or immigration. Let δ_1 denote the rate at which unaware susceptibles $S(t)$ become fully aware. δ_2 denotes the rate at which $S(t)$ becomes partially aware of the disease. The parameters β_1, β_2, and β_3 are the transmission rates of infection of unaware, fully aware, and partially aware susceptible populations, respectively. We suppose that all the susceptible classes can become infected by contact with infected individuals, but the fully aware class has less chance to be infected as compared to the unaware susceptible and partially aware susceptible individuals [33]. Therefore, it is assumed that β_1 and β_3 are greater β_2. The parameter ε represents the behavioral changes or measures of inhibition adopted by infectives. ϑ denotes the natural mortality rate of all individuals, whereas d and θ denote the disease-induced mortality rate and recovery rate of the infectives, respectively. The function $h(I(t)) = aI/(1+bI)$ represents the treatment rate of infectives, where a and b denote the cure rate and limitation rate in the availability of resources, respectively. The symbols and description of the parameters and state variables are given in Table 1 briefly.

The flow diagram of the proposed epidemic model is given in Fig. 1, and model is represented mathematically under the following system of delay differential equations:

$$\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= \kappa - \delta_1 S - \delta_2 S - \frac{\beta_1 S(t-\varrho) I(t-\varrho)}{1+\varepsilon I(t-\varrho)} - \vartheta S,\\
\frac{\mathrm{d}A_F}{\mathrm{d}t} &= \delta_1 S - \frac{\beta_2 I(t-\varrho) A_F(t-\varrho)}{1+\varepsilon I(t-\varrho)} - \vartheta A_F,\\
\frac{\mathrm{d}A_P}{\mathrm{d}t} &= \delta_2 S - \frac{\beta_3 I(t-\varrho) A_P(t-\varrho)}{1+\varepsilon I(t-\varrho)} - \vartheta A_P,\\
\frac{\mathrm{d}I}{\mathrm{d}t} &= \frac{\beta_1 S(t-\varrho) I(t-\varrho)}{1+\varepsilon I(t-\varrho)} + \frac{\beta_2 I(t-\varrho) A_F(t-\varrho)}{1+\varepsilon I(t-\varrho)}\\
&\quad + \frac{\beta_3 I A_P(t-\varrho)}{1+\varepsilon I(t-\varrho)} - (d+\theta+\vartheta) I - \frac{aI}{1+bI},\\
\frac{\mathrm{d}R}{\mathrm{d}t} &= \theta I + \frac{aI}{1+bI} - \vartheta R.
\end{aligned} \tag{1}$$

Suppose that the initial conditions of the model (1) take the form

$$\begin{aligned}
&S(\nu) = \chi_1(\nu),\ A_F(\nu) = \chi_2(\nu),\ A_P(\nu) = \chi_3(\nu),\\
&I(\nu) = \chi_4(\nu),\ R(\nu) = \chi_5(\nu),\\
&\chi_i(\nu).\ \nu \in [-\varrho,\ 0],\ \chi_i(0) > 0\ (i = 1, 2, 3, 4, 5),
\end{aligned} \tag{2}$$

where $(\chi_1(\nu), \chi_2(\nu), \chi_3(\nu), \chi_4(\nu), \chi_5(\nu)) \in C([-\varrho, 0], \mathbb{R}^5_+)$, which is the Banach space of continuous functions mapping the interval $[-\varrho, 0]$ into $\mathbb{R}^5_+$, where

$$\begin{aligned}
\mathbb{R}^5_+ = \{&(S, A_F, A_P, I, R)\,/S \geq 0, A_F \geq 0,\\
&A_P \geq 0, I \geq 0, R \geq 0\}.
\end{aligned}$$

Using Proposition 2.3 in [32] and Proposition 2.1 in [25], it follows that all state variables of the model (1) are nonnegative, i.e., $(S, A_F, A_P, I, R) \in \mathbb{R}^5_+$. For

Table 1 Notations of model variables and parameters

Symbol	Description
N	Total constant population
$S(t)$	Unaware susceptibles
$A_F(t)$	Fully aware susceptibles
$A_P(t)$	Partially aware susceptibles
$I(t)$	Infected population
$R(t)$	Removed population
ϱ	Latent period (time delay)
κ	Constant recruitment rate of unaware susceptibles
β_1	Transmission rate of susceptibles to infected individuals
β_2	Transmission rate of fully aware to infected individuals
β_3	Transmission rate of partially aware to infected individuals
ε	Inhibition measures by infectives
δ_1	Rate of full awareness in unaware susceptibles
δ_2	Rate of partial awareness in unaware susceptibles
ϑ	Natural death rate
d	Disease-induced death rate
θ	Recovery rate
a	Cure rate
b	Rate of limitations in treatment availability

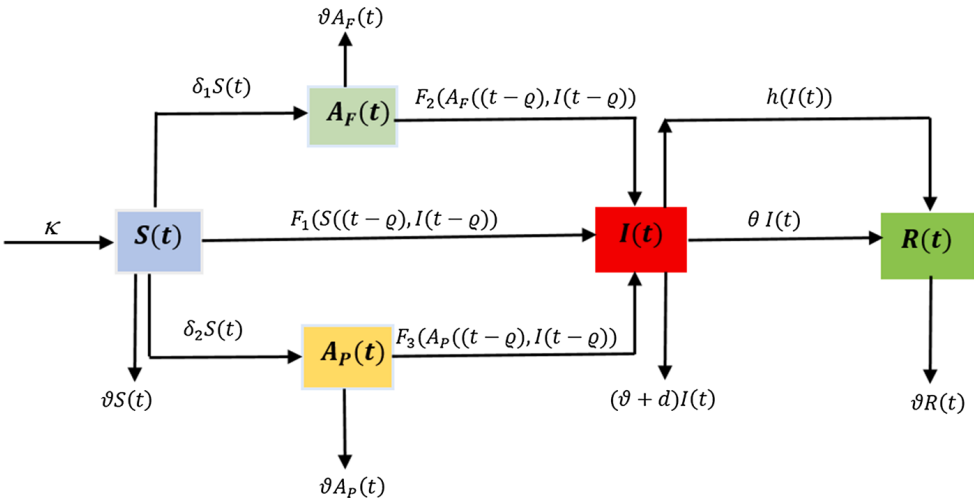


Fig. 1 Flow diagram of the model (1)

biological reasons, we assume that the model parameters κ, δ_1, δ_2 β_1, β_2, β_3, ε, ϑ, θ, d, a, b are positive. By the fundamental theory of functional differential equations [6], the model (1) has a unique solution $(S(t)\, A_F(t), A_P(t), I(t), R(t))$ satisfying the initial conditions (2).

Lemma 1 *The compact set*

$$\Omega = \left\{(S(t), A_F(t), A_P(t), I(t), R(t)) \in \mathbb{R}^5_+ : S(t) + A_F(t) + A_P(t) + I(t) + R(t) \leq \frac{\kappa}{\vartheta}\right\}$$

is invariant for the solutions of the model (1).

Proof Since the right-hand side of the model (1) and its derivatives are continuous, it assures the well-posedness of the model (1).

On adding the equations of the model (1), we get

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\left(S(t)+A_F(t)+A_P(t)+I(t)+R(t)\right)\\&=\kappa-\vartheta\left(S(t)+A_F(t)+A_P(t)\right.\\&\quad\left.+I(t)+R(t)\right)-dI\\&\leq\kappa-\vartheta\left(S(t)+A_F(t)+A_P(t)+I(t)+R(t)\right).\end{aligned}\tag{3}$$

Thus, the invariant region for the existence of the solutions is given as

$$0<\lim_{t\to\infty}\left(S(t)+A_F(t)+A_P(t)+I(t)+R(t)\right)\leq\frac{\kappa}{\vartheta}.\tag{4}$$

Hence, the solution of the model (1) is compact. □

In the next section, we obtain the model's equilibria and perform stability analysis to investigate the behavior of the equilibrium points.

3 Mathematical Analysis

Since $R(t)$ does not appear in the first four equations of the model (1), without loss of generality, it is sufficient to consider first four equations of the model for the analysis purpose:

$$\begin{aligned}\frac{\mathrm{d}S}{\mathrm{d}t}&=\kappa-\delta_1S-\delta_2S-\frac{\beta_1S(t-\varrho)I(t-\varrho)}{1+\varepsilon I(t-\varrho)}-\vartheta S,\\\frac{\mathrm{d}A_F}{\mathrm{d}t}&=\delta_1S-\frac{\beta_2I(t-\varrho)A_F(t-\varrho)}{1+\varepsilon I(t-\varrho)}-\vartheta A_F,\\\frac{\mathrm{d}A_P}{\mathrm{d}t}&=\delta_2S-\frac{\beta_3I(t-\varrho)A_P(t-\varrho)}{1+\varepsilon I(t-\varrho)}-\vartheta A_P,\\\frac{\mathrm{d}I}{\mathrm{d}t}&=\frac{\beta_1S(t-\varrho)I(t-\varrho)}{1+\varepsilon I(t-\varrho)}+\frac{\beta_2I(t-\varrho)A_F(t-\varrho)}{1+\varepsilon I(t-\varrho)}\\&\quad+\frac{\beta_3IA_P(t-\varrho)}{1+\varepsilon I(t-\varrho)}-(d+\theta+\vartheta)I-\frac{aI}{1+bI}.\end{aligned}\tag{5}$$

For the existence of model equilibrium, we set the right-hand side terms of the (5) to zero and obtain that the model (5) has two equilibria given below:

1. Disease-free equilibrium $E_0(S_0, A_{F_0}, A_{P_0}, 0)$ (represents the eradication of infected individuals, i.e., $I\equiv 0$ for all $t>0$), discussed in Sect. 3.1.
2. Endemic equilibrium $E_e(S^*, A_F^*, A_P^*, I^*)$ (represents the persistence of infected individuals above a certain positive level, i.e., $I>0$ for all $t>0$), discussed in Sect. 3.2.

3.1 Disease-free equilibrium

The system (5) has a disease-free equilibrium of the form $E_0(S_0, A_{F_0}, A_{P_0}, 0)$, where

$$S_0=\frac{\kappa}{\delta_1+\delta_2+\vartheta},\ A_{F_0}=\frac{\delta_1\kappa}{\vartheta(\delta_1+\delta_2+\vartheta)},$$
$$A_{P_0}=\frac{\delta_2\kappa}{\vartheta(\delta_1+\delta_2+\vartheta)}.$$

The characteristic equation of the system (5) at disease-free equilibrium is obtained as

$$\begin{aligned}&(-\vartheta-\lambda)^2\left(-\delta_1-\delta_2-\vartheta-\lambda\right)\left(\lambda+(\vartheta+d+\theta+a)\right.\\&\left.-\left(\beta_1S_0+\beta_2A_{F_0}+\beta_3A_{P_0}\right)\mathrm{e}^{-\lambda\varrho}\right)=0.\end{aligned}\tag{6}$$

Equation (6) is fourth-degree transcendental equation in λ with roots $\lambda_1=-\vartheta$, $\lambda_2=-\vartheta$, $\lambda_3=-(\delta_1+\delta_2+\vartheta)$, and the other roots are the solution of

$$\begin{aligned}&\lambda+(\vartheta+d+\theta+a)\\&-\left(\beta_1S_0+\beta_2A_{F_0}+\beta_3A_{P_0}\right)\mathrm{e}^{-\lambda\varrho}=0.\end{aligned}\tag{7}$$

On simplification, Eq. (7) can be written as

$$\begin{aligned}X(\lambda):=\lambda+(\vartheta+d+\theta+a)\\-\frac{\kappa\left(\beta_1\vartheta+\beta_2\delta_1+\beta_3\delta_2\right)}{\vartheta\left(\delta_1+\delta_2+\vartheta\right)}\mathrm{e}^{-\lambda\varrho}=0.\end{aligned}\tag{8}$$

The term $\frac{\kappa(\beta_1\vartheta+\beta_2\delta_1+\beta_3\delta_2)}{\vartheta(\vartheta+d+\theta+a)(\delta_1+\delta_2+\vartheta)}\mathrm{e}^{-\lambda\varrho}$ at $\varrho=0$ is defined as the basic reproduction number of the system (5). Thus, the basic reproduction number of the system (5) is

$$R_0=\frac{\kappa\left(\beta_1\vartheta+\beta_2\delta_1+\beta_3\delta_2\right)}{\vartheta\left(\vartheta+d+\theta+a\right)\left(\delta_1+\delta_2+\vartheta\right)}.$$

The basic reproduction number R_0 is the average number of secondary infections caused by a single infected agent, during his/her entire infectious period, in an entirely susceptible population [8].

3.1.1 Analysis for $\mathbf{R_0\neq 1}$

As mentioned above, Eq. (6) has three negative real roots $\lambda_1=-\vartheta$, $\lambda_2=-\vartheta$, $\lambda_3=-(\delta_1+\delta_2+\vartheta)$, and other roots are the solutions of Eq. (8).

Note that

$$X(0)=(\vartheta+d+\theta+a)\,(1-R_0).\tag{9}$$

If $R_0 > 1$, then $X(0) < 0$. Also,

$$X'(\lambda) = 1 + \frac{\kappa\,(\beta_1\vartheta + \beta_2\delta_1 + \beta_3\delta_2)}{\vartheta\,(\delta_1 + \delta_2 + \vartheta)}\mathrm{e}^{-\lambda\varrho}\varrho > 0. \quad (10)$$

Thus, $\lim_{\lambda\to\infty} X(\lambda) = +\infty$.

Hence, $X(\lambda) = 0$, $X(0) < 0$, $X'(\lambda) > 0$, and $\lim_{\lambda\to\infty} X(\lambda) = +\infty$ indicate that there always exists a unique and positive real root of $X(\lambda) = 0$ if $R_0 > 1$.

Now, if $R_0 < 1$, then

$$\begin{aligned}\mathrm{Re}\ \lambda &= \frac{\kappa\,(\beta_1\vartheta + \beta_2\delta_1 + \beta_3\delta_2)\,\mathrm{e}^{-(\mathrm{Re}\ \lambda)\varrho}\cos(Im\lambda)\varrho}{\vartheta\,(\delta_1 + \delta_2 + \vartheta)} \\ &\quad - (\vartheta + d + \theta + a) \\ &< \frac{\kappa\,(\beta_1\vartheta + \beta_2\delta_1 + \beta_3\delta_2)}{\vartheta\,(\delta_1 + \delta_2 + \vartheta)} - (\vartheta + d + \theta + a) \\ &= (\vartheta + d + \theta + a)\,(R_0 - 1) \\ &< 0. \end{aligned} \quad (11)$$

Therefore, $R_0 < 1$ implies that λ is a root of Eq. (6) with negative real part.

3.1.2 Analysis at $\mathbf{R_0 = 1}$

In this subsection, the system (5) is analyzed at $E_0(S_0, A_{F_0}, A_{P_0}, 0)$ and $R_0 = 1$ for the time delay $\varrho \geq 0$.

Case (i) $\varrho > 0$

If $R_0 = 1$, then $\lambda = 0$ is a simple characteristic root of Eq. (8). Note that $R_0 = 1$ implies that $\kappa\,(\beta_1\vartheta + \beta_2\delta_1 + \beta_3\delta_2) = \vartheta\,(\vartheta + d + \theta + a)\,(\delta_1 + \delta_2 + \vartheta)$.

Let $\lambda = r + is$ be any of the other solutions, then (8) becomes:

$$\begin{aligned} &r + \mathrm{i}s + (\vartheta + d + \theta + a) \\ &\quad - (\vartheta + d + \theta + a)\,\mathrm{e}^{-r\varrho}\,(\cos s\varrho - \mathrm{i}\sin s\varrho) = 0. \end{aligned} \quad (12)$$

By applying Euler's formula and splitting real and imaginary parts, Eq. (12) can be written as

$$\begin{aligned} r + \vartheta + d + \theta + a &= (\vartheta + d + \theta + a)\cos s\varrho\ \mathrm{e}^{-r\varrho}, \\ s &= -(\vartheta + d + \theta + a)\sin s\varrho\ \mathrm{e}^{-r\varrho}. \end{aligned} \quad (13)$$

On squaring and adding both the equations of Eq. (13), we obtain

$$\begin{aligned} &(r + \vartheta + d + \theta + a)^2 + s^2 \\ &\quad = (r + \vartheta + d + \theta + a)^2\,\mathrm{e}^{-2r\varrho}. \end{aligned} \quad (14)$$

If there exists a root satisfying both the equations of (13), then this root also satisfies Eq. (14) obtained by squaring and adding these equations. For Eq. (14) to be verified, we must have $r \leq 0$. Therefore, E_0 is linearly neutrally stable.

Case (ii) $\varrho = 0$

We now analyze the qualitative behavior of the system (5) without time delay, i.e., we take $\varrho = 0$. This analysis has an interest in itself and will also allow getting some information on the stability of the coexistence equilibrium in the case with delay [30]. The bifurcation theory approach obtained in [11], which is based on the center manifold theory [5], has been used to decide the local stability of a nonhyperbolic equilibrium (i.e., linearization matrix has at least one eigenvalue with zero real part) near the criticality E_0 and $R_0 = 1$. It allows us to clarify the direction of the bifurcation and describes the local behavior of disease-free equilibrium near $R_0 = 1$. For simplicity, we redefine the state variables as $S = x_1$, $A_F = x_2$, $A_P = x_3$, and $I = x_4$. So, the system (5) takes the form:

$$\begin{aligned} \frac{\mathrm{d}x_1}{\mathrm{d}t} &= \kappa - \delta_1 x_1 - \delta_2 x_1 - \frac{\beta_1 x_1 x_4}{\varepsilon x_4 + 1} - x_1\vartheta \equiv g_1, \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} &= -\frac{\beta_2 x_2 x_4}{\varepsilon x_4 + 1} - \vartheta x_2 + \delta_1 x_1 \equiv g_2, \\ \frac{\mathrm{d}x_3}{\mathrm{d}t} &= -\frac{\beta_3 x_3 x_4}{\varepsilon x_4 + 1} - \vartheta x_3 + \delta_2 x_1 \equiv g_3, \\ \frac{\mathrm{d}x_4}{\mathrm{d}t} &= \frac{\beta_1 x_1 x_4}{\varepsilon x_4 + 1} + \frac{\beta_2 x_2 x_4}{\varepsilon x_4 + 1} + \frac{\beta_3 x_3 x_4}{\varepsilon x_4 + 1} \\ &\quad - (d + \theta + \vartheta)x_4 - \frac{a x_4}{b x_4 + 1} \equiv g_4. \end{aligned} \quad (15)$$

We observe that when $R_0 = 1$, the chosen bifurcation parameter β_2 takes the form

$$\begin{aligned} \beta_2 &= \beta_2^* \\ &= \frac{\delta_2\,(\vartheta(a + d + \theta + \vartheta) - \beta_3\kappa) + \vartheta\,(\delta_1 + \vartheta)\,(a + d + \theta + \vartheta) + \beta_1\kappa(-\vartheta)}{\delta_1\kappa}. \end{aligned}$$

The Jacobian matrix $J(E_0,\ \beta_2^*)$ of the system (15) obtained at criticality (that is, at E_0 and β_2^*) is

$$J(E_0, \beta_2^*) = \begin{pmatrix} -\vartheta - \delta_1 - \delta_2 & 0 & 0 & -\frac{\kappa\beta_1}{\vartheta + \delta_1 + \delta_2} \\ \delta_1 & -\vartheta & 0 & -\frac{\kappa\beta_2^*\delta_1}{\vartheta\,(\vartheta + \delta_1 + \delta_2)} \\ \delta_2 & 0 & -\vartheta & -\frac{\kappa\beta_3\delta_2}{\vartheta\,(\vartheta + \delta_1 + \delta_2)} \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The eigenvalues of the Jacobian matrix $J(E_0, \beta_2^*)$ are $\lambda_1 = -\vartheta,\ \lambda_2 = -\vartheta,\ \lambda_3 = -\vartheta - \delta_1 - \delta_2,\ \lambda_4 = 0$. We see that λ_1, λ_2, and λ_3 are real and negative eigenvalues and $\lambda_4 = 0$ is a simple zero eigenvalue (since the algebraic multiplicity of $\lambda_4 = 0$ is 1) of $J(E_0, \beta_2^*)$. Thus, when $R_0 = 1$, the DFE E_0 is a non-hyperbolic equilibrium.

The right eigenvector $v = (v_1, v_2, v_3, v_4)$ of $J(E_0, \beta_2^*)$ corresponding to $\lambda_4 = 0$ is obtained as

$$
\begin{aligned}
v_1 &= -\frac{\beta_1 \kappa}{(\delta_1 + \delta_2 + \vartheta)^2},\\
v_2 &= -\frac{\delta_1 \kappa\,(\beta_2\,(\delta_1 + \delta_2 + \vartheta) + \beta_1 \vartheta)}{\vartheta^2\,(\delta_1 + \delta_2 + \vartheta)^2},\\
v_3 &= -\frac{\delta_2 \kappa\,(\beta_3\,(\delta_1 + \delta_2 + \vartheta) + \beta_1 \vartheta)}{\vartheta^2\,(\delta_1 + \delta_2 + \vartheta)^2},\\
v_4 &= 1.
\end{aligned}
$$

The left eigenvector $w = (w_1, w_2, w_3, w_4)$ of $J(E_0,\ \beta_2^*)$ corresponding to $\lambda_4 = 0$ is obtained as

$$
\begin{aligned}
w_1 &= 0,\\
w_2 &= 0,\\
w_3 &= 0,\\
w_4 &= 1.
\end{aligned}
$$

Let g_k's, $k = 1, 2, 3, 4$, denote the right-hand side of the system (15). The bifurcation coefficients a_1 and b_1 defined in Theorem 4.1 of [33] are given by:

$$
\begin{aligned}
a_1 &= \sum_{k,i,j=1}^{4} w_k v_i v_j \left(\frac{\partial^2 g_k}{\partial x_i \partial x_j}\right)_{E_0},\\
b_1 &= \sum_{k,i=1}^{4} w_k v_i \left(\frac{\partial^2 g_k}{\partial x_i \partial \beta_2^*}\right)_{E_0}.
\end{aligned}
$$

The nonzero partial derivatives of the functions g_k's at E_0 are evaluated as

$$
\begin{aligned}
\left(\frac{\partial^2 g_4}{\partial x_4 \partial x_1}\right)_{E_0} &= \beta_1, \left(\frac{\partial^2 g_4}{\partial x_4 \partial x_2}\right)_{E_0} = \beta_2^*,\\
\left(\frac{\partial^2 g_4}{\partial x_4 \partial x_3}\right)_{E_0} &= \beta_3, \left(\frac{\partial^2 g_4}{\partial x_1 \partial x_4}\right)_{E_0} = \beta_1,\\
\left(\frac{\partial^2 g_4}{\partial x_2 \partial x_4}\right)_{E_0} &= \beta_2^*,\\
\left(\frac{\partial^2 g_4}{\partial x_3 \partial x_4}\right)_{E_0} &= \beta_3, \left(\frac{\partial^2 g_4}{\partial x_4^2}\right)_{E_0}\\
&= -\frac{2\varepsilon\beta_1\kappa}{\delta_1 + \delta_2 + \vartheta} - \frac{2\varepsilon\beta_2^*\delta_1\kappa}{\vartheta\,(\delta_1 + \delta_2 + \vartheta)}\\
&\quad -\frac{2\varepsilon\beta_3\delta_2\kappa}{\vartheta\,(\delta_1 + \delta_2 + \vartheta)} + 2ab, \textit{and}\\
\left(\frac{\partial^2 g_4}{\partial x_4 \partial \beta_2^*}\right)_{E_0} &= \frac{\delta_1\kappa}{\vartheta\,(\delta_1 + \delta_2 + \vartheta)}.
\end{aligned}
$$

The bifurcation coefficients a_1 and b_1 are calculated at the bifurcation parameter β_2^* as follows:

$$
\begin{aligned}
a_1 &= -\frac{2}{\delta_1\kappa\vartheta^2\,(\delta_1 + \delta_2 + \vartheta)}\Big(\delta_1(\delta_2(\vartheta^2\,(\kappa(a(\varepsilon - b)\\
&\quad +\varepsilon(d + \theta + \vartheta)) + 2(a + d + \theta + \vartheta)^2\Big)\\
&\quad + \beta_3\kappa\,(\beta_3\kappa - 2\vartheta(a + d + \theta + \vartheta)))\\
&\quad + \vartheta^3\,(\kappa(a(\varepsilon - b) + \varepsilon(d + \theta + \vartheta))\\
&\quad +2(a + d + \theta + \vartheta)^2\Big))\\
&\quad + \delta_1^2\vartheta^2\,(\kappa(a(\varepsilon - b) + \varepsilon(d + \theta + \vartheta))\\
&\quad +(a + d + \theta + \vartheta)^2\Big)\\
&\quad + (\delta_2\,(\vartheta(a + d + \theta + \vartheta) - \beta_3\kappa)\\
&\quad + \vartheta^2(a + d + \theta + \vartheta))^2\\
&\quad - \beta_1\kappa\vartheta\,(2\delta_2\,(\vartheta(a + d + \theta + \vartheta) - \beta_3\kappa)\\
&\quad +\vartheta\,(\delta_1 + 2\vartheta)\,(a + d + \theta + \vartheta))\\
&\quad + \beta_1^2\kappa^2\vartheta^2\Big)\\
&= -\frac{2}{\delta_1\kappa\vartheta^2\,(\delta_1 + \delta_2 + \vartheta)} \times G(\beta_3),\\
b_1 &= \frac{\kappa}{\delta_1 + \delta_2 + \vartheta}.
\end{aligned}
$$

where

$$
G(\beta_3) = \left(\zeta_2\,\beta_3^2 + \zeta_1\,\beta_3 + \zeta_0\right), \tag{16}
$$

where the coefficients $\zeta_0,\ \zeta_1$ and ζ_2 are

$$
\begin{aligned}
\zeta_2 &= \delta_2\,(\delta_1 + \delta_2)\,\kappa^2,\\
\zeta_1 &= 2\delta_2\kappa\vartheta\,(\beta_1\kappa - (\delta_1 + \delta_2 + \vartheta)\,(a + d + \theta + \vartheta)),\\
\zeta_0 &= \vartheta^2\Big(\big(\delta_1 + \delta_2 + \vartheta\big)\big(\delta_1\big(\kappa(a(\varepsilon - b)\\
&\quad + \varepsilon(d + \theta + \vartheta))\\
&\quad + (a + d + \theta + \vartheta)^2\big)
\end{aligned}
$$

$$+\big(\delta_2+\vartheta\big)(a+d+\theta+\vartheta)^2\big)$$
$$+\beta_1\delta_1\kappa(a+d+\theta+\vartheta)$$
$$+\beta_1\kappa\big(\beta_1\kappa-2\big(\delta_1+\delta_2+\vartheta\big)(a+d+\theta+\vartheta)\big)\big). \tag{17}$$

It can be seen that the bifurcation coefficient b_1 is always positive and the sign of a_1 depends the sign of $G(\beta_3)$, given in Eq. (16). If $G(\beta_3) < 0$, then $a_1 > 0$, whereas if $G(\beta_3) > 0$, then $a_1 < 0$.

The discriminant of quadratic polynomial $G(\beta_3)$ is obtained as

$$D = a\big(\delta_1+\delta_2\big)\kappa(b-\varepsilon)\big(\delta_1+\delta_2+\vartheta\big)$$
$$+\big(-\delta_1-\delta_2-\vartheta\big)\big(\big(\delta_1+\delta_2+\vartheta\big)(a+d+\theta+\vartheta)^2$$
$$+\varepsilon\big(\delta_1+\delta_2\big)\kappa(d+\theta+\vartheta)\big)$$
$$+\beta_1\kappa\big(\delta_1+\delta_2+2\vartheta\big)(a+d+\theta+\vartheta)$$
$$+\beta_1^2\big(-\kappa^2\big). \tag{18}$$

Let β_3^* and β_3^{**} be two roots of Eq. (16), then we get

$$\beta_3^* = \frac{-\zeta_1+D}{2\zeta_2}, \text{ and } \beta_3^{**} = \frac{-\zeta_1-D}{2\zeta_2}. \tag{19}$$

Using Theorem 4.1(iv) in [11], the type of bifurcation is governed by the sign of a_1 and hence by the sign of $G(\beta_3)$. If $G(\beta_3)$ is of positive sign, then forward bifurcation occurs, whereas if the sign of $G(\beta_3)$ is negative, then the system (15) reveals a backward bifurcation. The study of forward and backward bifurcation is important in the study of epidemiology as they play a relevant role in disease control and eradication. In the case of forward bifurcation, the condition $R_0 < 1$ is a necessary and sufficient condition for disease eradication, whereas reducing R_0 below unity is no longer sufficient for disease eradication when a backward bifurcation occurs. These behavior differences are essential in planning how to control a disease; a backward bifurcation at $R_0 = 1$ makes control more difficult. These two cases are discussed below separately.

(I) Forward bifurcation: When there is a forward bifurcation at $R_0 = 1$, it is not possible for a disease to invade a population if $R_0 < 1$ because the system will return to the disease-free equilibrium $I = 0$ if some infectives are introduced into the population. For values of R_0 slightly greater than 1, E_0 changes its stability from stable to unstable and the model admits a unique endemic equilibrium, which is locally asymptotically stable [11]. Therefore, it is imperative to find the range for which forward bifurcation occurs. The range of forward bifurcation is governed by the positivity of $G(\beta_3)$. Thus, there are two cases in which $G(\beta_3)$ is found to be positive. These are given as follows:

$$\begin{cases} \zeta_1 > 0, \\ \zeta_0 > 0. \end{cases} \tag{20}$$

or

$$\begin{cases} D > 0, \\ \text{either of } (\zeta_1 > 0,\ \zeta_0 < 0),\ (\zeta_1 < 0.\ \zeta_0 > 0),\ \text{or } (\zeta_1 < 0,\ \zeta_0 < 0) \text{ holds}, \\ \beta_3 < \beta_3^* \text{ or } \beta_3 > \beta_3^{**}. \end{cases} \tag{21}$$

(II) Backward bifurcation: The backward bifurcation is characterized as when $R_0 < 1$; a small unstable endemic equilibrium appears while the disease-free equilibrium and a larger endemic equilibrium are locally asymptotically stable. When $R_0 > 1$, then an unstable disease-free equilibrium and a stable endemic equilibrium exist [11]. It is illustrated in Fig. (4). The range of existence of backward bifurcation (i.e., when $G(\beta_3) < 0$) as follows:

$$\begin{cases} D > 0, \\ \text{either of } (\zeta_1 > 0,\ \zeta_0 < 0),\ (\zeta_1 < 0.\ \zeta_0 > 0),\ \text{or } (\zeta_1 < 0,\ \zeta_0 < 0) \text{ holds}, \\ \beta_3^{**} < \beta_3 < \beta_3^*. \end{cases} \tag{22}$$

Based on the analysis above, we state the following theorems:

Theorem 1 *The disease-free equilibrium* $E_0(S_0, A_{F_0}, A_{P_0}, 0)$ *of the delayed system (5) is asymptotically stable if* $R_0 < 1$ *and unstable if* $R_0 > 1$ *for* $\varrho > 0$.

Theorem 2 *The disease-free equilibrium* $E_0(S_0, A_{F_0}, A_{P_0}, 0)$ *of the delayed system (5) at* $R_0 = 1$ *is linearly neutrally stable for* $\varrho > 0$.

Theorem 3 *When* $R_0 = 1$, *then the undelayed system (15) reveals a backward (forward) bifurcation at disease-free equilibrium* $E_0(S_0, A_{F_0}, A_{P_0}, 0)$ *if and only if* $G(\beta_3) < 0\ (> 0)$.

The graphical presentations of the forward and backward bifurcations are shown in Figs. 2, 3, and 4 for the experimental data mentioned below: $\kappa = 2$, $\varepsilon = 0.1$, $\vartheta = 0.01$, $d = 0.001$, $b = 0.8$, $\theta = 0.01$, $a = 0.9$, $\beta_1 = 0.004$, $\delta_1 = 0.01$, $\delta_2 = 0.009$. At these values of parameters, we evaluate that the range of β_3 is $\beta_3^{**} = 0.00098164 \leq$

$\beta_3 \leq 0.0088652 = \beta_3^*$. The cases of occurrence of forward and backward bifurcations, given in the inequalities (20), (21), and (22), are illustrated from (1)–(3) as below:

(1) If we take $\beta_3 = 0.00008$, then we obtain that $\zeta_2 = 0.000522$, $\zeta_1 = 1.242 \times 10^{-6} > 0$, and $\zeta_0 = 7.75344 \times 10^{-9} > 0$. This case illustrates the inequality (20), and the graph is shown in Fig. 2.
(2) On considering $\beta_3 = 0.01$, we obtain that $\zeta_2 = 0.000522$, $\zeta_1 = 1.242 \times 10^{-}6 > 0$, $\zeta_0 = -5.44699 \times 10^{-8} < 0$, and the discriminant $D = 1.15276 \times 10^{-10} > 0$. It illustrates the inequality (21), and the graph is shown in Fig. 3.
(3) If we take $\beta_3 = 0.006$, which lies between β_3^{**} and β_3^*, then we obtain that the coefficients $\zeta_2 = 0.000522$, $\zeta_1 = 1.242 \times 10^{-6} > 0$ and $\zeta_0 = -4.05047 \times 10^{-8} < 0$, and the discriminant $D = 8.61164 \times 10^{-11} > 0$. This case illustrates the inequality (22), and the graph is shown in Fig. 4.

In Figs. 2, 3, and 4, the solid lines show stability and the dashed lines show instability.

3.2 Endemic equilibrium and stability

Assuming that $S,\ A_F,\ A_P,\ I \neq 0$. To determine the conditions for the existence of endemic (positive) equilibrium $E_e(S^*, A_F^*, A_P^*, I^*)$, we put the right-hand side of the system (5) to zero. We get:

$$\kappa - \delta_1 S - \delta_2 S - \frac{\beta_1 S I}{1+\varepsilon I} - S\vartheta = 0, \tag{23}$$

$$-\frac{\beta_2 A_F I}{1+\varepsilon I} - \vartheta A_F + \delta_1 S = 0, \tag{24}$$

$$-\frac{\beta_3 A_P I}{1+\varepsilon I} - \vartheta A_P + \delta_2 S = 0, \tag{25}$$

$$\frac{\beta_1 S I}{1+\varepsilon I} + \frac{\beta_2 A_F I}{1+\varepsilon I} + \frac{\beta_3 A_P I}{1+\varepsilon I} - (d+\theta+\vartheta)I - \frac{aI}{1+bI} = 0. \tag{26}$$

Solving for S, A_F and A_P in terms of I, from Eqs. (23), (24), and (25), respectively, and substituting the resulting expression in Eq. (26), after simplification, we obtain the following equation in I:

$$P(I) := A_0 + A_1 I + A_2 I^2 + A_3 I^3 + A_4 I^4 = 0, \tag{27}$$

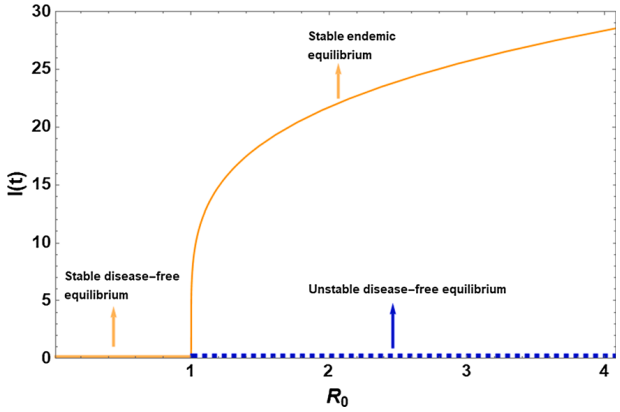


Fig. 2 Plot of R_0 versus $I(t)$, showing the occurrence of forward bifurcation

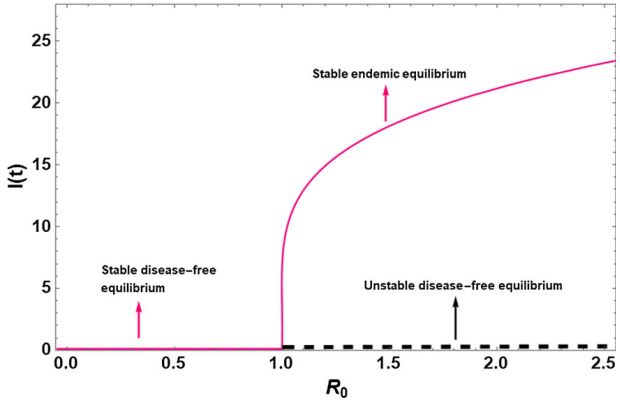


Fig. 3 Plot of R_0 versus $I(t)$, showing the occurrence of forward bifurcation

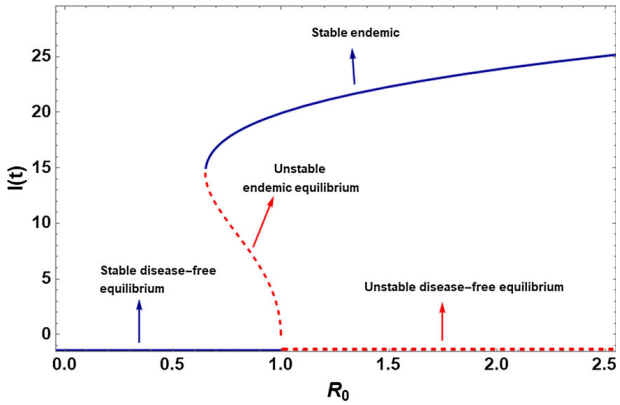


Fig. 4 Plot of R_0 versus $I(t)$, showing the presence of backward bifurcation

where

$$
\begin{aligned}
A_0 &= \vartheta\,(\delta_1+\delta_2+\vartheta)\,(a+d+\theta+\vartheta)(1-R_0),\\
A_1 &= \vartheta\,((\delta_1+\delta_2+\vartheta)\,(3\varepsilon\vartheta+\beta_2+\beta_3)+\beta_1\vartheta)\\
&\quad (a+d+\theta+\vartheta)-\beta_2\delta_1\kappa\vartheta\,(2\varepsilon+b)\\
&\quad -\beta_3\delta_2\kappa\vartheta\,(2\varepsilon+b)\\
&\quad -\beta_1\kappa\vartheta^2(2\varepsilon+b)+b\vartheta^2\,(\delta_1+\delta_2+\vartheta)\\
&\quad (d+\theta+\vartheta)-\beta_2\beta_3\delta_1\kappa-\beta_2\beta_3\delta_2\kappa\\
&\quad -\beta_1\,(\beta_2+\beta_3)\,\kappa\vartheta,\\
A_2 &= \vartheta\Big(\beta_3\,(\delta_2(2a\varepsilon-\varepsilon\kappa(\varepsilon+2b)+(2\varepsilon+b)(d+\theta+\vartheta))\\
&\quad +(\delta_1+\vartheta)\,(2a\varepsilon+(2\varepsilon+b)(d+\theta+\vartheta)))\\
&\quad +3\varepsilon\vartheta\,(\delta_1+\delta_2+\vartheta)\,(a\varepsilon+(\varepsilon+b)(d+\theta+\vartheta))\Big)\\
&\quad +\beta_2\Big(\beta_3\Big(\,(\delta_1+\delta_2)\,(a-\kappa(\varepsilon+b)+d+\theta+\vartheta)\\
&\quad +\vartheta(a+d+\theta+\vartheta)\Big)+\vartheta\big(\delta_1(2a\varepsilon-\varepsilon\kappa(\varepsilon+2b)\\
&\quad +(2\varepsilon+b)(d+\theta+\vartheta))\\
&\quad +(\delta_2+\vartheta)\,(2a\varepsilon+(2\varepsilon+b)(d+\theta+\vartheta))\big)\Big)\\
&\quad -\beta_1\Big(\beta_2\,(-\vartheta(a+d+\theta+\vartheta)+\kappa\vartheta(\varepsilon+b)+\beta_3\kappa)\\
&\quad +\vartheta\,(\vartheta(-2a\varepsilon+\varepsilon\kappa(\varepsilon+2b)\\
&\quad -(2\varepsilon+b)(d+\theta+\vartheta))\\
&\quad -\beta_3(a-\kappa(\varepsilon+b)+d+\theta+\vartheta))\Big),\\
A_3 &= -\,(\varepsilon\vartheta+\beta_3)\,((\varepsilon\vartheta+\beta_2)\,(-a\varepsilon\vartheta-\beta_1(a-b\kappa))\\
&\quad -\varepsilon\delta_1\,(a\varepsilon\vartheta+\beta_2(a-b\kappa)))+\varepsilon\delta_2\,(\varepsilon\vartheta+\beta_2)\,\big(a\varepsilon\vartheta\\
&\quad +\beta_3(a-b\kappa)\big)-(d+\theta+\vartheta)\big(\beta_1\,(\vartheta\,(-\beta_3(\varepsilon+b)\\
&\quad -\varepsilon\vartheta(\varepsilon+2b))-\beta_2\,(\vartheta(\varepsilon+b)+\beta_3))\\
&\quad -(\delta_1+\delta_2+\vartheta)\,(\beta_2\,(\beta_3(\varepsilon+b)+\varepsilon\vartheta(\varepsilon+2b))\\
&\quad +\varepsilon\vartheta\,(\beta_3(\varepsilon+2b)+\varepsilon\vartheta(\varepsilon+3b)))\big),\\
A_4 &= b\,(\varepsilon\vartheta+\beta_2)\,(\varepsilon\vartheta+\beta_3)\,(d+\theta+\vartheta)\\
&\quad (\varepsilon\,(\delta_1+\delta_2+\vartheta)+\beta_1)\,.
\end{aligned}
\tag{28}
$$

For a positive root of I^* of the polynomial $P(I)$, we have

$$S^* = \frac{\kappa+\varepsilon\kappa I^*}{(\delta_1+\delta_2+\vartheta)\,(\varepsilon I^*+1)+\beta_1 I^*}, \tag{29}$$

$$A_F^* = \frac{\delta_1 S^*(\varepsilon I^*+1)}{\varepsilon I^*\vartheta+\beta_2 I^*+\vartheta}, \tag{30}$$

$$A_P^* = \frac{\delta_2 S^*(\varepsilon I^*+1)}{\varepsilon I^*\vartheta+\beta_3 I^*+\vartheta}. \tag{31}$$

So, $E_e(S^*, A_F^*, A_P^*, I^*)$ is a positive equilibrium of the system (5).

Theorem 4 *When $R_0 > 1$, then there is either a unique or three positive endemic equilibria if all equilibria are simple roots.*

Proof Let $R_0 > 1$. We see that the coefficient A_4 is always positive. On the other hand, $A_0 < 0$ when $R_0 > 1$. From Eq. (27), we have a fourth-degree polynomial in I, given below:

$$P(I) := A_0 + A_1 I + A_2 I^2 + A_3 I^3 + A_4 I^4 = 0.$$

The following possibilities for the signs of A_1, A_2, and A_3 exist:

$$
\begin{aligned}
&\mathbf{V}_1 : A_1 > 0,\ A_2 > 0,\ \text{and } A_3 > 0,\\
&\mathbf{V}_2 : A_1 < 0,\ A_2 < 0,\ \text{and } A_3 > 0,\\
&\mathbf{V}_3 : A_1 < 0,\ A_2 > 0,\ \text{and } A_3 > 0,\\
&\mathbf{V}_4 : A_1 < 0,\ A_2 < 0,\ \text{and } A_3 < 0,\\
&\mathbf{V}_5 : A_1 > 0,\ A_2 > 0,\ \text{and } A_3 < 0,\\
&\mathbf{V}_6 : A_1 > 0,\ A_2 < 0,\ \text{and } A_3 > 0,\\
&\mathbf{V}_7 : A_1 > 0,\ A_2 < 0,\ \text{and } A_3 < 0,\\
&\mathbf{V}_8 : A_1 < 0,\ A_2 > 0,\ \text{and } A_3 < 0.
\end{aligned}
$$

Using Descartes' rule of signs [12], $P(I)$ can have either a unique or three positive roots. If any of the conditions $\mathbf{V}_1$–$\mathbf{V}_4$ holds, then there is unique endemic equilibrium, whereas for the existence of three endemic equilibria, any one of the conditions $\mathbf{V}_5$–$\mathbf{V}_8$ must satisfy. □

For the present study, we consider the case of unique endemic equilibrium only. **H1:** Suppose that any of the conditions ($\mathbf{V}_1$–$\mathbf{V}_4$, and $R_0 > 1$) holds, then the system (5) admits a unique endemic equilibrium.

Now, we study the local stability behavior of endemic equilibrium $E_e(S^*, A_F^*, A_P^*, I^*)$. For this, we obtain the characteristic equation of the system (5) at $E_e(S^*, A_F^*, A_P^*, I^*)$ as given below:

$$P_0(\lambda)+\mathrm{e}^{-\lambda\varrho}P_1(\lambda)+\mathrm{e}^{-2\lambda\varrho}P_2(\lambda)+\mathrm{e}^{-3\lambda\varrho}P_3(\lambda)=0, \tag{32}$$

where,

$$
\begin{aligned}
P_0(\lambda) &= \lambda^4+K_1\lambda^3+K_2\lambda^2+K_3\lambda+K_4,\\
P_1(\lambda) &= K_5\lambda^3+K_6\lambda^2+K_7\lambda+K_8,\\
P_2(\lambda) &= K_9\lambda^2+K_{10}\lambda+K_{11},\\
P_3(\lambda) &= K_{12}\lambda+K_{13}.
\end{aligned}
\tag{33}
$$

where the coefficients K_i, $i = 1$ to 13 are given in Appendix. For $\varrho = 0$, the characteristic equation becomes:

$$\lambda^4 + Q_3\lambda^3 + Q_2\lambda^2 + Q_1\lambda + Q_0 = 0, \tag{34}$$

where $Q_i's$, $i = 0, 1, 2, 3$, are given in Appendix.

Based on Routh–Hurwitz criterion, it can be concluded that all the roots of Eq. (34) have negative real parts if the following inequalities hold:

H2: $Q_i > 0$ $(i = 0, 1, 2, 3)$, $Q_3Q_2 - Q_1 > 0$, and
$$Q_3Q_2Q_1 - Q_1^2 - Q_3^2Q_0 > 0. \tag{35}$$

Thus, we state the following theorem:

Theorem 5 *At $\varrho = 0$, the endemic equilibrium $E_2(S^*, A_F^*, A_P^*, I^*)$ is locally asymptotically stable if **H2** holds.*

Now multiplying by $e^{-\lambda\varrho}$ on both sides of the characteristic equation (32), we get:

$$P_0(\lambda)e^{\lambda\varrho} + P_1(\lambda) + e^{-\lambda\varrho}P_2(\lambda) + e^{-2\lambda\varrho}P_3(\lambda) = 0. \tag{36}$$

Change of stability and hence the existence of oscillatory solution may appear if the roots of the characteristic equation are purely imaginary. Therefore, to study the stability of E_e, assuming that $\lambda = i\omega$, $(\omega > 0)$ is a root of Eq. (36). Then, Eq. (36) becomes:

$$P_0(i\omega)e^{i\omega\varrho} + P_1(i\omega) + e^{-i\omega\varrho}P_2(i\omega) + e^{-2i\omega\varrho}P_3(i\omega) = 0. \tag{37}$$

Equation (37) can be rewritten as

$$\begin{aligned}&(M_0 + iN_0)(\cos\omega\varrho + i\sin\omega\varrho) + (M_1 + iN_1)\\ &+ (M_2 + iN_2)(\cos\omega\varrho - i\sin\omega\varrho)\\ &+ (M_3 + iN_3)(\cos 2\omega\varrho - i\sin 2\omega\varrho) = 0,\end{aligned} \tag{38}$$

where M_0, M_1, M_2, M_3, N_0, N_1, N_2 and N_3 denote the real and imaginary parts of $P_0(i\omega)$, $P_1(i\omega)$, $P_2(i\omega)$ and $P_3(i\omega)$, respectively, given as:

$$\begin{aligned}&M_0 = \omega^4 - K_2\omega^2 + K_4,\ N_0 = -K_1\omega^3 + K_3\omega,\\ &M_1 = -K_6\omega^2 + K_8,\ N_1 = -K_5\omega^3 + K_7\omega,\\ &M_2 = -K_9\omega^2 + K_{11},\ N_2 = K_10\omega,\\ &M_3 = K_{13},\ N_3 = K_{12}\omega.\end{aligned}$$

When Eq. (38) splits into its real and imaginary parts, we get

$$M_1 + M_4\cos\omega\varrho + V_4\sin\omega\varrho = -N_3\sin 2\omega\varrho - M_3\cos 2\omega\varrho, \tag{39}$$

$$N_1 + N_5\cos\omega\varrho + M_5\sin\omega\varrho = -N_3\cos 2\omega\varrho + M_3\sin 2\omega\varrho, \tag{40}$$

where $M_4 = M_0 + M_2$, $N_4 = N_2 - N_0$, $M_5 = M_0 - M_2$ and $N_5 = N_2 + N_0$.

On squaring Eqs. (39) and (40) and then adding, we obtain

$$(M_1 + M_4\cos\omega\varrho + N_4\sin\omega\varrho)^2 + (N_1 + N_5\cos\omega\varrho) + M_5(\sin\omega\varrho)^2 - N_3^2 - M_3^2 = 0. \tag{41}$$

On substituting $\sin\omega\varrho = \sqrt{1 - \cos^2\omega\varrho}$ in Eq. (41), we obtain

$$L_1\cos^4\omega\varrho + L_2\cos^3\omega\varrho + L_3\cos^2\omega\varrho + L_4\cos\omega\varrho + L_5 = 0, \tag{42}$$

where

$$\begin{aligned}&L_1 = p_1^2 + 4p_5^2,\ L_2 = 4p_1p_2 + 8p_4p_5,\\ &L_3 = 4p_2^2 + 2p_1p_3 + 4p_4^2 - 4p_5^2,\\ &L_4 = 4p_2p_3 - 8p_4p_5,\\ &L_5 = p_3^2 - 4p_4^2,\\ &p_1 = M_4^2 + N_5^2 - N_4^2 - M_5^2,\ p_2 = M_1M_4 + N_1N_4,\\ &p_3 = M_1^2 + N_4^2 + N_1^2 + M_5^2 - N_3^2 - M_3^2,\\ &p_4 = M_1N_4 + N_1M_5,\\ &p_5 = M_4N_4 + M_5N_5.\end{aligned}$$

Let $\cos\omega\varrho = x$, then Eq. (42) can be written as

$$f(x) := L_1x^4 + L_2x^3 + L_3x^2 + L_4x + L_5 = 0. \tag{43}$$

From Eq. (43), we obtain

$$f'(x) = 4L_1x^3 + 3L_2x^2 + 2L_3x + L_4 = 0. \tag{44}$$

For convenience, we assume that $x = y - L_2/4L_1$. Then, Eq. (44) becomes:

$$y^3 + Ay + B = 0, \tag{45}$$

where

$$A = \frac{L_3}{2L_1} - \frac{3}{16}\left(\frac{T_2}{T_1}\right)^2,$$

$$B = \frac{1}{32}\left(\frac{L_2}{L_1}\right)^3 - \frac{L_2L_3}{8L_1^2} + \frac{L_4}{4L_1}.$$

Now, roots of Eq. (45) are given as

$$y_1 = \sqrt[3]{-\frac{B}{2} + \sqrt{O}} + \sqrt[3]{-\frac{B}{2} - \sqrt{O}},$$

$$y_2 = \eta_1 \sqrt[3]{-\frac{B}{2} + \sqrt{O}} + \eta_2 \sqrt[3]{-\frac{B}{2} - \sqrt{O}},$$
$$y_3 = \eta_2 \sqrt[3]{-\frac{B}{2} + \sqrt{O}} + \eta_1 \sqrt[3]{-\frac{B}{2} - \sqrt{O}},$$

where $\eta_1 = \frac{-1+\sqrt{3}\mathrm{i}}{2}$, $\eta_2 = \frac{-1-\sqrt{3}\mathrm{i}}{2}$ and $O = \frac{B^2}{4} + \frac{A_3}{27}$.

It follows from $\cos \omega\varrho = x$ and $x = y - L_2/4L_1 = F_1(\omega)$ that

$$\cos \omega\varrho = F_1(\omega). \tag{46}$$

On substituting Eq. (46) in Eq. (41) and solving for $\sin \omega\varrho$, we obtain

$$\sin \omega\varrho = F_2(\omega). \tag{47}$$

From Eqs. (46) and (47), we get

$$F_1^2(\omega) + F_2^2(\omega) = 1. \tag{48}$$

Assume that **H3:**, Eq. (48) has at least one positive root ω_0, such that the characteristic equation (32) has a pair of purely imaginary roots $\mathrm{i}\omega_0$.

For ω_0, the corresponding critical value of time delay ϱ_k can be obtained as:

$$\varrho_k = \frac{1}{\omega_0}\left(\arccos F_1(\omega_0) + 2k\pi\right), \ k = 0, 1, 2, \ldots. \tag{49}$$

Assume that ϱ is a bifurcation parameter and ϱ_0=min ϱ_k, $k = 0, 1, 2, \ldots$ is the critical value.

To establish the Hopf bifurcation, we show that $Re\left[\frac{\mathrm{d}\lambda}{\mathrm{d}\varrho}\right]^{-1}\Big|_{\lambda=\mathrm{i}\omega_0} \neq 0$.

Differentiating Eq. (32) with respect to ϱ, we obtain:

$$\begin{aligned}
&\frac{\mathrm{d}\lambda}{\mathrm{d}\varrho}\left(P_0'(\lambda) + \mathrm{e}^{-\lambda\varrho}P_1'(\lambda) - \varrho\mathrm{e}^{-\lambda\varrho}P_1(\lambda) + \mathrm{e}^{-2\lambda\varrho}P_2'(\lambda) - 2\varrho P_2(\lambda)\mathrm{e}^{-2\lambda\varrho} + \mathrm{e}^{-3\lambda\varrho}P_3'(\lambda) - 3\varrho P_3(\lambda)\mathrm{e}^{-3\lambda\varrho}\right)\\
&\quad - \lambda\mathrm{e}^{-\lambda\varrho}P1(\lambda) - 2\lambda P_2(\lambda)\mathrm{e}^{-2\lambda\varrho} - 3\lambda P_3(\lambda)\mathrm{e}^{-3\lambda\varrho} = 0.\\
&\Longrightarrow \frac{\mathrm{d}\lambda}{\mathrm{d}\varrho}\left(P_0'(\lambda) + \mathrm{e}^{-\lambda\varrho}P_1'(\lambda) - \varrho\mathrm{e}^{-\lambda\varrho}P_1(\lambda) + \mathrm{e}^{-2\lambda\varrho}P_2'(\lambda) - 2\varrho P_2(\lambda)\mathrm{e}^{-2\lambda\varrho} + \mathrm{e}^{-3\lambda\varrho}P_3'(\lambda) - 3\varrho P_3(\lambda)\mathrm{e}^{-3\lambda\varrho}\right)\\
&\quad = \lambda\mathrm{e}^{-\lambda\varrho}P1(\lambda) + 2\lambda P_2(\lambda)\mathrm{e}^{-2\lambda\varrho} + 3\lambda P_3(\lambda)\mathrm{e}^{-3\lambda\varrho}.\\
&\Longrightarrow \frac{\mathrm{d}\lambda}{\mathrm{d}\varrho} = \frac{\lambda\left(P_1(\lambda) + 2P_2(\lambda)\mathrm{e}^{-\lambda\varrho} + 3P_3(\lambda)\mathrm{e}^{-2\lambda\varrho}\right)}{\left(P_0'(\lambda)\mathrm{e}^{\lambda\varrho} + P_1'(\lambda) - \varrho P_1(\lambda) + \mathrm{e}^{-\lambda\varrho}\left(P_2'(\lambda) - 2\varrho P_2(\lambda)\right) + \mathrm{e}^{-2\lambda\varrho}\left(P_3'(\lambda) - 3\varrho P_3(\lambda)\right)\right)}.
\end{aligned}$$

Thus, we get

$$\left[\frac{\mathrm{d}\lambda}{\mathrm{d}\varrho}\right]^{-1} = \frac{\left(P_0'(\lambda)\mathrm{e}^{\lambda\varrho} + P_1'(\lambda) + \mathrm{e}^{-\lambda\varrho}P_2'(\lambda) + \mathrm{e}^{-2\lambda\varrho}P_3'(\lambda)\right) - \left(\varrho P_1(\lambda) + 2\varrho P_2(\lambda)\mathrm{e}^{-\lambda\varrho} + 3\varrho P_3(\lambda)\mathrm{e}^{-2\lambda\varrho}\right)}{\lambda\left(P_1(\lambda) + 2P_2(\lambda)\mathrm{e}^{-\lambda\varrho} + 3P_3(\lambda)\mathrm{e}^{-2\lambda\varrho}\right)}.$$

$$\text{i.e.,} \left[\frac{\mathrm{d}\lambda}{\mathrm{d}\varrho}\right]^{-1} = \frac{\left(P_0'(\lambda)\mathrm{e}^{\lambda\varrho} + P_1'(\lambda) + \mathrm{e}^{-\lambda\varrho}P_2'(\lambda) + \mathrm{e}^{-2\lambda\varrho}P_3'(\lambda)\right)}{\lambda\left(P_1(\lambda) + 2P_2(\lambda)\mathrm{e}^{-\lambda\varrho} + 3P_3(\lambda)\mathrm{e}^{-2\lambda\varrho}\right)} - \frac{\varrho}{\lambda}.$$

$$Re\left[\frac{d\lambda}{d\varrho}\right]^{-1}\Big|_{\lambda=\mathrm{i}\omega_0} = \frac{X_1X_2 + Y_1Y_2}{X_2^2 + Y_2^2},$$

where

$$
\begin{aligned}
X_1 &= -\sin\omega_0\varrho_0\left(-4\omega_0^3+2K_2\omega_0\right)\\
&\quad+\cos\omega_0\varrho_0\left(-3K_1\omega_0^2+K_3\right)+\left(-3K_5\omega_0^2+K_7\right)\\
&\quad+\left(K_{10}\cos\omega_0\varrho_0+2\omega_0K_9\sin\omega_0\varrho_0\right)\\
&\quad+K_{16}\cos 2\omega_0\varrho_0,\\
X_2 &= K_5\omega_0^4-K_7\omega_0^2-2K_{10}\omega_0^2\cos\omega_0\varrho_0\\
&\quad+2\left(-K_9\omega_0^2+\omega_0K_{11}\right)\sin\omega_0\varrho_0\\
&\quad-3\omega_0^2K_{12}\cos 2\omega_0\varrho_0\\
&\quad+3\omega_0K_{13}\sin 2\omega_0\varrho_0,\\
Y_1 &= \cos\omega_0\varrho_0\left(-4\omega_0^3+2K_2\omega_0\right)\\
&\quad+\sin\omega_0\varrho_0\left(-3K_1\omega_0^2+K_3\right)+2\omega_0K_6\\
&\quad+2\omega_0K_9\cos\omega_0\varrho_0-K_{10}\sin\omega_0\varrho_0\\
&\quad-K_{16}\sin 2\omega_0\varrho_0,\\
Y_2 &= -K_6\omega_0^3+\omega_0K_8+2\left(-K_9\omega_0^3+\omega_0K_{11}\right)\cos\omega_0\varrho_0\\
&\quad+3\left(\omega_0^2K_{12}\sin 2\omega_0\varrho_0+\omega_0K_{13}\cos 2\omega_0\varrho_0\right)\\
&\quad+2\omega_0^2K_{10}\sin\omega_0\varrho_0.
\end{aligned}
$$

Obviously, if **H4:** $X_1X_2+Y_1Y_2\neq 0$, then $Re\left[\frac{d\lambda}{d\varrho}\right]^{-1}\Big|_{\lambda=i\omega_0}\neq 0$.

Thus, we state the following theorem:

Theorem 6 *For the system (5), if the conditions (**H1–H4**) hold, then the endemic equilibrium* $E_e=(S^*, A_F^*, A_P^*, I^*)$ *is locally asymptotically stable when* $\varrho\in[0,\varrho_0)$*; the system (5) undergoes a Hopf bifurcation at* $E_e=(S^*, A_F^*, A_P^*, I^*)$ *when* $\varrho=\varrho_0$*, and a family of periodic solutions bifurcate from* $E_e=(S^*, A_F^*, A_P^*, I^*)$.

4 Numerical simulation

In this section, numerical experiments are presented to show the analytical results using Mathematica 11.

We have considered the following set of experimental data: $\kappa=2$, $\vartheta=0.01$, $\beta_1=0.02$, $\beta_2=0.006$, $\beta_3=0.008$, $\delta_1=0.1$, $\delta_2=0.02$, $a=0.04$, $d=0.009$, $\theta=0.004$, $\varepsilon=0.5$, $b=0.09$.

At these parameters values, the endemic equilibrium is $E_e(S^*, A_F^*, A_P^*, I^*)=(11.903, 55.602, 9.443, 38.504)$ with $R_0>1$.

Figure 5 shows the behavior of different subpopulations for the time delay $\varrho=1$. Evidently, as time increases, unaware susceptibles decrease, and the fully aware and partially aware, infected, and removed individuals population increase and then start decaying and settle down to steady state $E_e(11.903, 55.602, 9.443, 38.504)$.

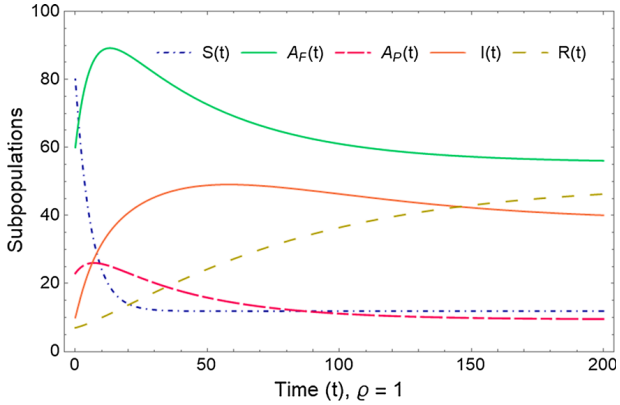


Fig. 5 Subpopulations $S-A_P-A_F-I-R$ at $\varrho=1$

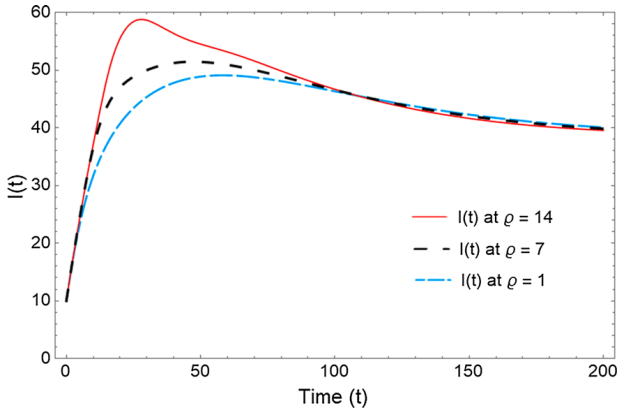
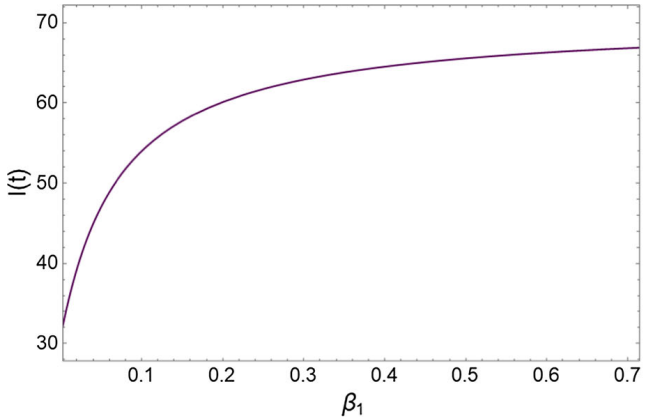


Fig. 6 Infected population $I(t)$ at different values of time delay ϱ

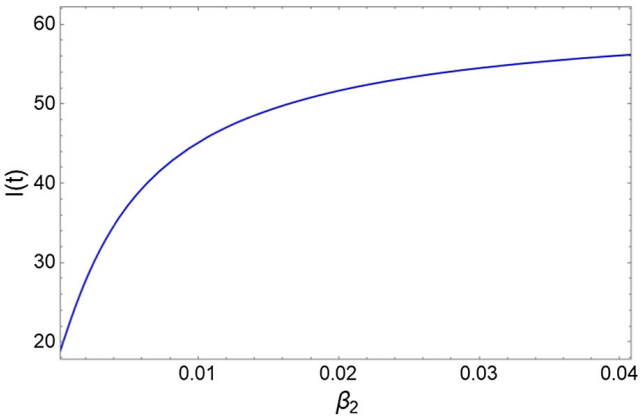
Figure 6 depicts the effect of time delay on the infected population. We plot the infected population for different values of time delay $\varrho=1, 7$, and 14, respectively. It reveals that the large value of time delay causes an increment in the infected population.

Figure 7a–c shows the influence of different transmission rates on infected population $I(t)$. It validates the increment in the number of infected population as the transmission rates increase, which is biologically true.

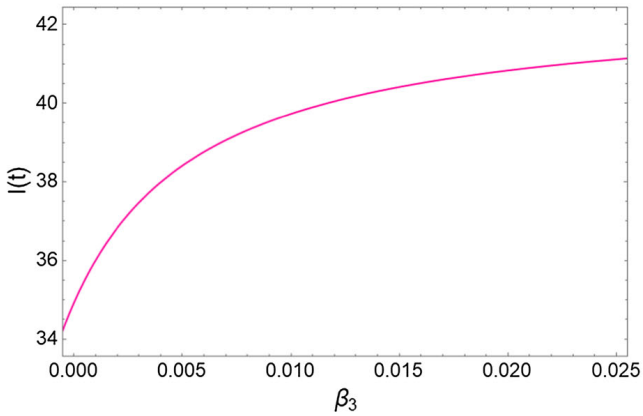
Figure 8a, b shows the impact of full and partial awareness rates (δ_1 and δ_2) on the infected population for the time delay $\varrho=1$. It is evident that if the full awareness rate is high, then infection diminishes at a high level. Partially awareness in humans also helps



(a) Infected population $I(t)$ when β_1 varies.

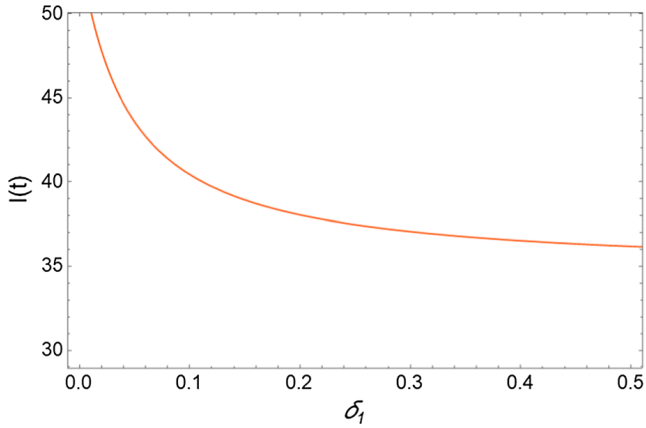


(b) Infected population $I(t)$ when β_2 varies.

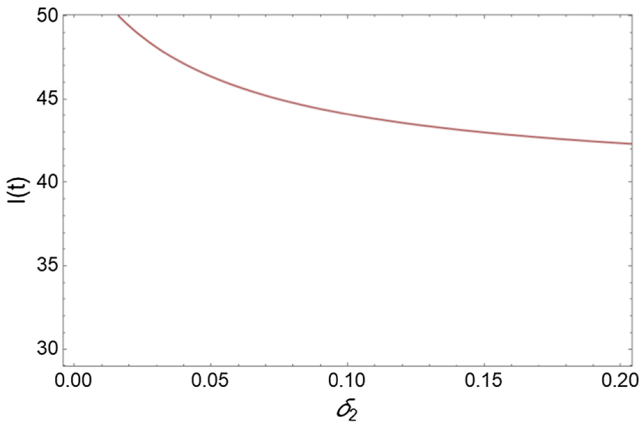


(c) Infected population $I(t)$ when β_3 varies.

Fig. 7 Infected population for the transmission rates of unaware, fully aware, and partially aware susceptibles for the time delay $\varrho = 1$



(a) Infected population $I(t)$ for δ_1.



(b) Infected population $I(t)$ for δ_2.

Fig. 8 Impact of full and partial awareness rates on Infected population for the time delay $\varrho = 1$

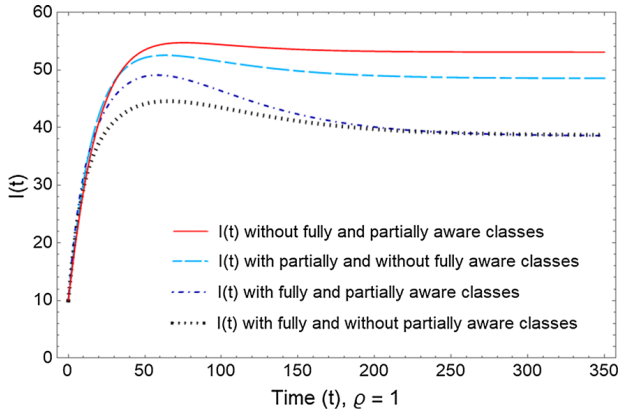
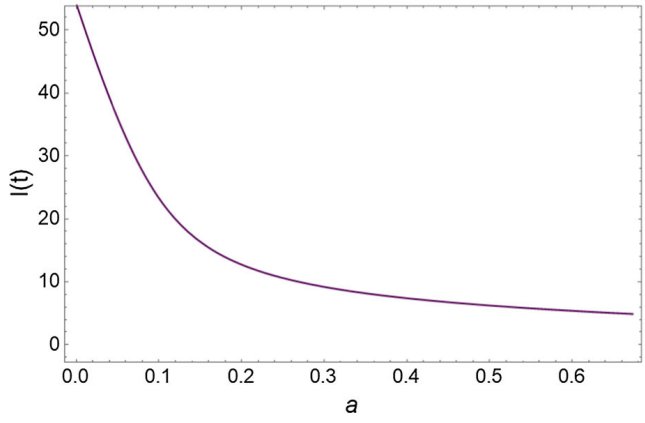


Fig. 9 Dynamics of infectious diseases showing the impact of aware classes on infected individuals $I(t)$ for the time delay $\varrho = 1$

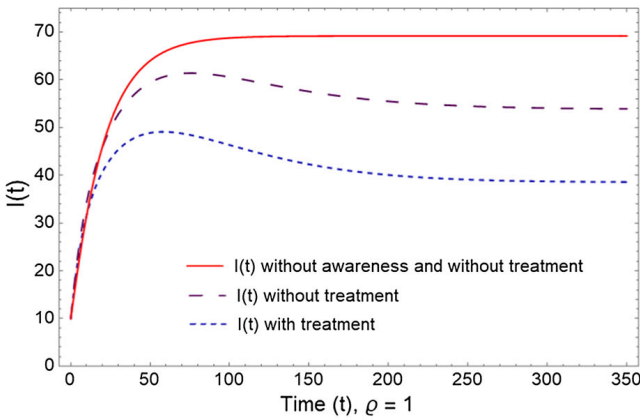
them to escape from the infection, as depicted in Fig. 8b. Thus, more efforts should be put to spread full awareness among people.

Figure 9 examines the potential of fully and partially aware classes in minimizing the impact of an epidemic. From the graph, it is evident that when people are not aware of the spread of disease at all (shown by a solid red line), then the infection occurs at a higher rate. Awareness in humans motivates them to escape from the infection. By the full and right information about a disease, individuals change their attitudes and actions to reduce their chances of becoming infected, spreading the disease further, or experiencing prolonged periods of medical treatment. Attempts to raise awareness of an infectious disease may also build a sense of threat in individuals who are inadequately informed. When there is only partial information available, then infection reduces, but a low level (as shown by the dashed green line). Because of weak, inaccessible, and absence of education, uneducated individuals are rarely formally trained in the process of handling diseases, its prognosis, and diagnosis, preventions, and cures. They take preventive measures by word of mouth or by social media, which leads to a reduction in the infection at a low level. Also, if people are partially mindful, then there are chances that they can take unnecessary medications due to fear of catching the infection, which can weaken their immune system. So those people are at more risk of getting infected (shown by the dot-dashed blue line). Therefore, the absence of partial awareness and presence of full awareness about the spread and prevention of disease are leading to a reduction in the range of illness at the high rate (shown by the dotted black line).

Figure 10a shows that the increase in the cure rate can increase the reduction in infected individuals. Figure 10b shows the influence of saturated treatment rate on infected population $I(t)$. We have plotted the infected population when there is neither treatment nor awareness available. Clearly, in this case, infection is spreading at a very high rate. Purple dashed line shows the infected population when people are aware, but treatment is not available, and the solid red line shows the infected population in the absence of awareness and treatment. The major difference can be seen between these two lines. Blue dashed line shows



(a) Effect of cure rate a on Infected population $I(t)$ for the time delay $\varrho = 1$.



(b) Infected individuals $I(t)$ with and without saturated treatment rate and aware classes for the time delay $\varrho = 1$.

Fig. 10 Impact of cure rate, awareness, and saturated treatment on the infected population for $\varrho = 1$

the infected population when both awareness and treatment are present, which is stabilizing at the lowest level among three lines. Thus, the treatment rate with awareness helps in reducing infection at a faster rate.

To show the presence of Hopf bifurcation, we take the following experimental data: $\kappa = 10$, $\varepsilon = 0.92$, $\beta_1 = 0.2$, $\beta_2 = 0.02$, $\beta_3 = 0.04$, $\delta_1 = 0.05$, $\delta_2 = 0.03$, $\vartheta = 0.1$, $a = 0.2$, $b = 0.5$, $\theta = 0.01$, $d = 0.01$.

We obtain that $E_e(S^*, A_F^*, A_P^*, I^*) = (25.5152, 10.5267, 5.37598, 42.1257)$, at these values of parameters.

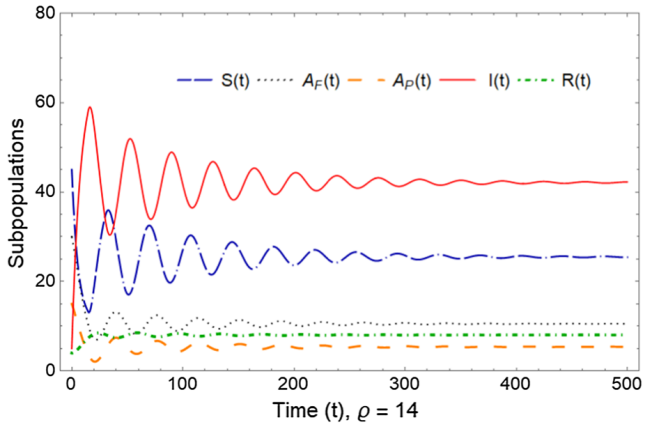
Figure 11 depicts the time series solutions of the model (1) with their respective SIR phase plot for distinct values of time delay ϱ. It is observed that the spread of infectious disease can be controlled for $\varrho = 14$, and $\varrho = 21$. That is, Fig. 11a–d shows that initially periodic solutions appear but after a time; the endemic equilibrium E_e reached to its steady state. Figure 11e–h shows the unstable limit cycle around the endemic equilibrium when time delay $\varrho > \varrho_0 = 24$.

5 Discussion

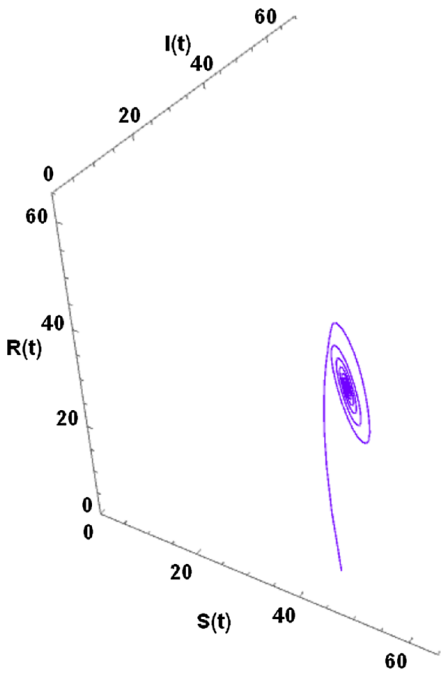
In the present article, we divide the total population into five compartments: unaware susceptibles, fully aware susceptibles, partially aware susceptibles, infected, and removed individuals and study a time-delayed epidemic model with the inclusion of saturated incidences and treatment rates. We analyze the model mathematically and study the dynamic behaviors of the epidemic model with and without time delay ϱ. The mathematical analysis of the model shows that it exhibits two equilibria: disease-free and endemic. By deriving the basic reproduction number R_0, we prove that the disease-free equilibrium is locally asymptotically stable when $R_0 < 1$, unstable when $R_0 > 1$, and linearly neutrally stable when $R_0 = 1$ for the time delay $\varrho > 0$. When we don't consider the time delay, then using the center manifold theory, it is obtained that the forward or backward bifurcation occurs when $R_0 = 1$. We obtain the bifurcation range of forward and backward bifurcations, given in the inequalities (20), (21), and (22), respectively. In the presence of forward bifurcation, as R_0 increases through unity, the disease-free equilibrium loses its stability, and a stable endemic equilibrium appears. In contrast, the presence of backward bifurcation shows that a stable endemic equilibrium coexists with a stable DFE when $R_0 < 1$. It has an important implication as it fails the ideal condition of reducing R_0 below unity to eradicate the diseases from society. Thus, the control programs must reduce R_0 further than below unity to eliminate the disease. Schematic diagrams of forward and backward bifurcation are depicted in Figs. (2), (3), and (4). Further, the local stability of the endemic equilibrium has been studied, which demonstrates that by choosing time delay as a bifurcation parameter, the Hopf bifurcation occurs near the endemic equilibrium, revealing the presence of oscillatory and periodic solutions.

The numerical simulations show the graphical representation of the effectiveness of theoretical results. We see that the consideration of time delay has a significant role as it has an impact on the number of infectives. It is seen that when the time delay is high, then infection spreads at a higher rate. The periodic and oscillatory solutions have been plotted near endemic equilibrium, which shows the presence of Hopf bifurcation at different values of time delay ϱ. As the delay ϱ passes through the critical value ϱ_0, the endemic equilibrium loses its stability, and an unstable limit cycle appears. When there is no treatment available, then knowledge about the spread of disease is the main focus. The role of full and partial awareness in susceptibles, with and without saturated treatment, has been shown numerically. When treatment is not given to infected individuals, then only susceptibles' full or partial awareness shows the significant difference in the number of infectives (Figs. 9, 10b), whereas if we consider saturated treatment rate along with awareness in the susceptibles, then the transmission pattern of infectious diseases changes more effectively, and the reduction in the prevalence of the disease can be seen to the more extent.

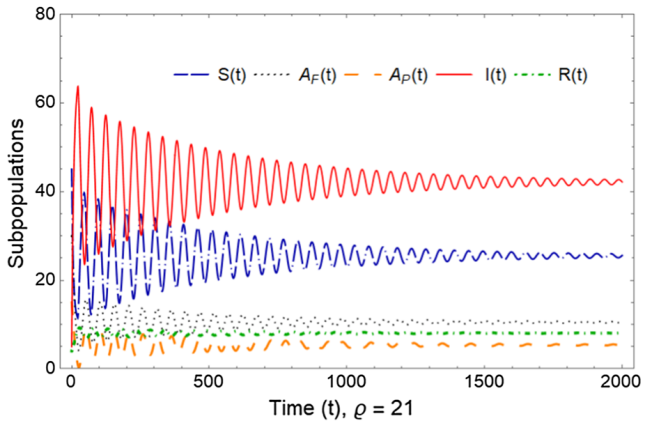
The findings of the model, consisting of explicit saturated incidences with latent period and saturated treatment rate, are capable of demonstrating the significant role of the latent period, the behavior of susceptibles through different subclasses, and limitation in available facilities of treatment. The results are capable of understanding the role of varying protection levels of susceptibles in the transmission pattern of the infectious diseases and hence further suggesting the control strategies to prevent the spread of infections at a massive scale. The full awareness about the spread of infectious disease increases public perception to avert infection and willingness to adopt prevention methods, which mitigates the transmission of disease. Due to the



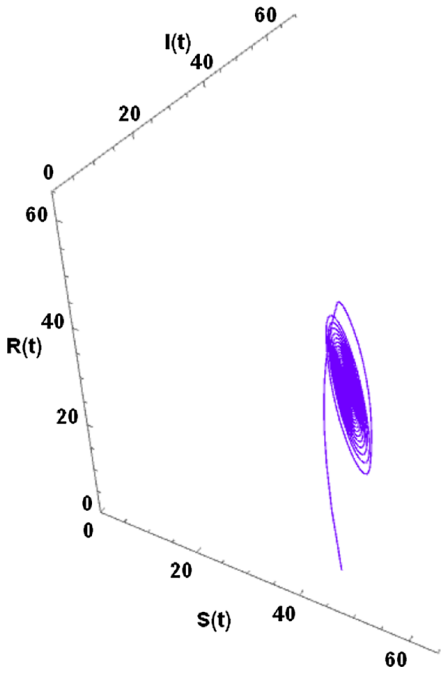
(a) Time series solution of the model (1) for $\varrho = 14$.



(b) Phase plot of susceptible, infected and removed population when $\varrho = 14$.

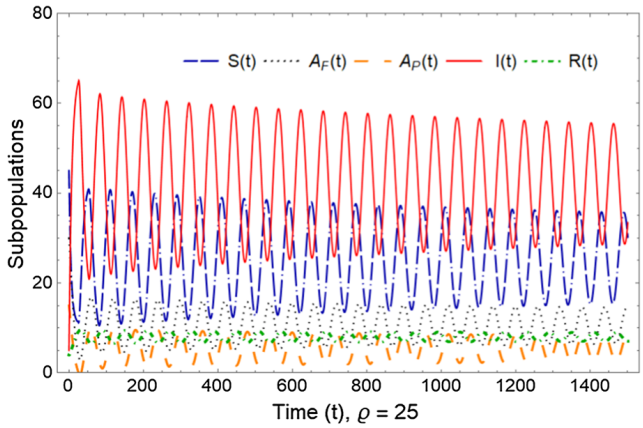


(c) Time series solution of the model (1) for $\varrho = 21$.

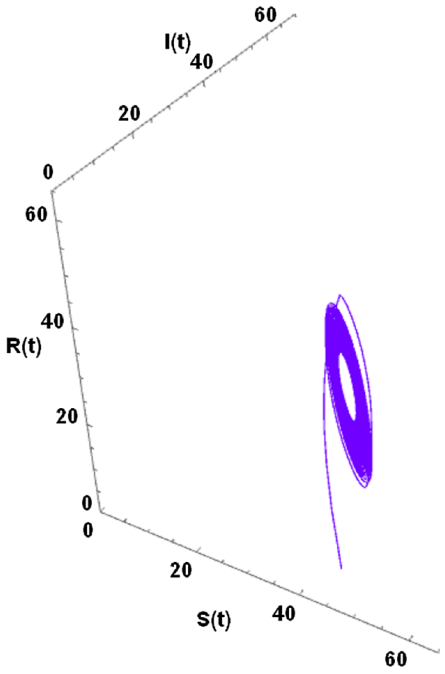


(d) Phase plot of susceptible, infected and removed population when $\varrho = 21$.

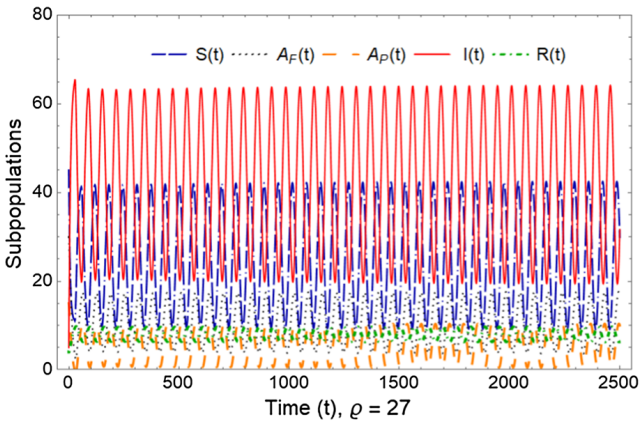
Fig. 11 Graphs depicting the presence of Hopf bifurcation for different values of time delay ϱ



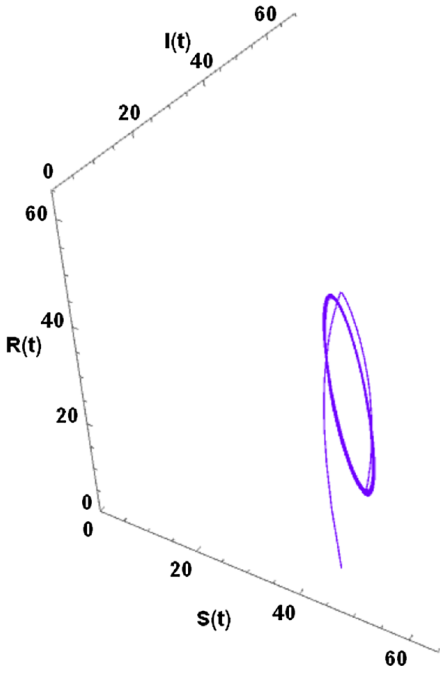
(e) Time series solution of the model (1) for $\varrho = 25$.



(f) Phase plot of susceptible, infected and removed population when $\varrho = 25$.



(g) Time series solution of the model (1) for $\varrho = 27$.



(h) Phase plot of susceptible, infected and removed population when $\varrho = 27$.

Fig. 11 continued

different social, educational, and limited information resources, some people may have incomplete information. Therefore, these partially aware individuals adopt insufficient preventive methods, through which infection reduces but at a low level. Public health initiatives can put extra effort into making individuals fully aware by enhancing health literacy and providing sufficient information resources. It can contribute to early case detection and help in reducing the transmission of infection. Thus, for the eradication of the disease, programs related to regular knowledge, full information, education, and communication, concerning the spread of infectious diseases and its importance to the public and health-care workers, will help them in improving their general attitude toward it. Timely disease pieces of information updates are highly needed. Together with awareness, appropriate treatment to infectives and the availability of health resources will help in diminishing the infection from society, effectively.

Acknowledgements The authors are grateful to the Delhi Technological University, Delhi, India, for providing financial support to carry out this research work. They are also indebted to the editor and the anonymous reviewers for their constructive comments and suggestions, which led to an improvement in the original manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interest regarding the publication of this article.

Appendix

Using Mathematica 11, we obtain the coefficients of the characteristic equation (32) as follows:

$$K_1 = \frac{a + (1+bI^*)^2(d+\theta+4\vartheta) + \delta_1(1+bI^*)^2 + \delta_2(1+bI^*)^2}{(1+bI^*)^2},$$

$$K_2 = \frac{(\delta_1+\delta_2)\left(a+(1+bI^*)^2(\theta+d+3\vartheta)\right) + 3\vartheta\left(a+(d+\theta+2\vartheta)(bI^*+1)^2\right)}{(1+bI^*)^2},$$

$$K_3 = \vartheta\left((\delta_1+\delta_2)\left(2a+(1+bI^*)^2(2(d+\theta)+3\vartheta)\right) + \vartheta\left(3a+(bI^*+1)^2(3(d+\theta)+4\vartheta)\right)\right),$$

$$K_4 = \frac{\vartheta^2(\delta_1+\delta_2+\vartheta)\left(a+(d+\theta+\vartheta)(1+bI^*)^2\right)}{(1+bI^*)^2},$$

$$K_5 = \frac{\beta_3\left(-A_P^* + \varepsilon I^{*2} + I^*\right) + \beta_2\left(-A_F^* + \varepsilon I^{*2} + I^*\right) + \beta_1\left(-S^* + \varepsilon I^{*2} + I\right)}{(\varepsilon I^*+1)^2},$$

$$\begin{aligned}
K_6 = {} & \frac{1}{(1+bI^*)^2(1+\varepsilon I^*)^2}(\beta_3(I^*(\varepsilon I^*+1)\\
& \left(a+(d+\theta+3\vartheta)(1+bI^*)^2\right)\\
& +(1+bI^*)^2((\delta_1+\delta_2)\,I^*(1+\varepsilon I^*)\\
& -A_P^*(\delta_1+\delta_2+3\vartheta))) + \beta_2(I^*(\varepsilon I^*+1)\\
& \left(a+(1+bI^*)^2(\theta+d+3\vartheta)\right)\\
& +(1+bI^*)^2((\delta_1+\delta_2)\\
& \times I^*(\varepsilon I^*+1) - A_F^*(\delta_1+\delta_2+3\vartheta)))\\
& +\beta_1(aI^*(\varepsilon I^*+1) + (1+bI^*)^2\\
& (I^*(1+\varepsilon I^*)(\theta+d+3\vartheta)\\
& -3S^*\vartheta) - (\delta_1+\delta_2)\,S^*(bI^*+1)^2)),
\end{aligned}$$

$$\begin{aligned}
K_7 = {} & \frac{1}{(1+bI^*)^2(1+\varepsilon I^*)^2}(\beta_3(I^*(1+\varepsilon I^*)((\delta_1+\delta_2)\\
& \left(a+(1+bI^*)^2(\theta+d+2\vartheta)\right)\\
& +\vartheta(2a+(1+bI^*)^2\\
& \times(2(d+\theta)+3\vartheta))) - \vartheta A_P^*(1+bI^*)^2\\
& (2\delta_1+2\delta_2+3\vartheta)) + \beta_2(I^*(\varepsilon I^*+1)((\delta_1+\delta_2)\\
& \times\left(a+(1+bI^*)^2(\theta+d+2\vartheta)\right)\\
& +\vartheta\left(2a+(1+bI^*)^2(2(d+\theta)+3\vartheta)\right))\\
& -\vartheta A_F^*(bI^*+1)^2\\
& (2\delta_1+2\delta_2+3\vartheta)) + \beta_1\vartheta(2aI^*(\varepsilon I^*+1)\\
& +(bI^*+1)^2(I^*(\varepsilon I^*+1)(2(d+\theta)+3\vartheta)\\
& -3S^*\vartheta) - 2\,(\delta_1+\delta_2)\,S^*(bI^*+1)^2)),
\end{aligned}$$

$$\begin{aligned}
K_8 = {} & \frac{\vartheta}{(bI^*+1)^2(\varepsilon I^*+1)^2}((\delta_1+\vartheta+\delta_2)\\
& (\beta_3(I^*(\varepsilon I^*+1)\\
& \left(a+(1+bI^*)^2(\theta+d+\vartheta)\right)\\
& -\vartheta A_P^*(1+bI^*)^2)
\end{aligned}$$

$$
\begin{aligned}
&+\beta_2\left(I^*\left(a+(bI^*+1)^2(\theta+d+\vartheta)\right)(1+\varepsilon I^*)\right.\\
&\left.-\vartheta A_F^*(1+bI^*)^2\right)+\beta_1\vartheta(aI^*(1+\varepsilon I^*)\\
&+(1+bI^*)^2(I^*(\theta+d+\vartheta)(\varepsilon I^*+1)-S^*\vartheta)\\
&-(\delta_1+\delta_2)S^*(1+bI^*)^2)),
\end{aligned}
$$

$$
\begin{aligned}
K_9=&\frac{1}{(1+\varepsilon I^*)^3}(I^*(\beta_1\left(\beta_3\left(-A_P^*-S^*+\varepsilon I^{*2}+I^*\right)\right.\\
&\left.+\beta_2\left(-A_F^*-S^*+\varepsilon I^{*2}+I^*\right)\right)\\
&+\beta_2\beta_3\left(-A_P^*-A_F^*+\varepsilon I^{*2}+I^*\right))^3),
\end{aligned}
$$

$$
\begin{aligned}
K_{10}=&\frac{I^*}{(1+\varepsilon I^*)^3(1+bI^*)^2}(\beta_1(\beta_3((1+\varepsilon I^*)aI^*\\
&-2\vartheta A_P^*(1+bI^*)^2+(1+bI^*)^2\\
&(I^*(\theta+d+2\vartheta)(1+\varepsilon I^*)\\
&-2S^*\vartheta)-\delta_1S^*(1+bI^*)^2)+\beta_2((1+\varepsilon I^*)aI^*\\
&-2\vartheta A_F^*(1+bI^*)^2+(1+bI^*)^2(I^*(1+\varepsilon I^*)\\
&\times(\theta+d+2\vartheta)-2S^*\vartheta)-\delta_2S^*(1+bI^*)^2))\\
&+\beta_2\beta_3((1+\varepsilon I^*)I^*\left(a+(1+bI^*)^2(\theta+d+2\vartheta)\right)\\
&+(1+bI^*)^2((\delta_1+\delta_2)\times I^*(\varepsilon I^*+1)\\
&-(\delta_1+\delta_2+2\vartheta)\left(A_P^*+A_F^*\right)))),
\end{aligned}
$$

$$
\begin{aligned}
K_{11}=&\frac{I^*}{(1+\varepsilon I^*)^3(1+bI^*)^2}(\beta_1\vartheta(\beta_3((1+\varepsilon I^*)aI^*\\
&-\vartheta A_P^*(1+bI^*)^2\\
&+(1+bI^*)^2(I^*(1+\varepsilon I^*)(d+\theta+\vartheta)\\
&-S^*\vartheta)-\delta_1S^*(bI^*+1)^2)+\beta_2(aI^*(\varepsilon I^*+1)\\
&-\vartheta A_F^*(1+bI^*)^2\\
&+(1+bI^*)^2(I^*(1+\varepsilon I^*)(\theta+d+\vartheta)\\
&-S^*\vartheta)-\delta_2S^*(bI^*+1)^2))\\
&+\beta_2\beta_3\left(\delta_1+\delta_2+\vartheta\right)(I^*(1+\varepsilon I^*)\\
&\left(a+(1+bI^*)^2(\theta+d+\vartheta)\right)\\
&-\vartheta(1+bI^*)^2\left(A_P^*+A_F^*\right))),
\end{aligned}
$$

$$
\begin{aligned}
&\frac{1}{(\varepsilon I^*+1)^3}\left(I^*(\beta_2\beta_3(-\vartheta)\right.\\
&\times(\delta_1+\delta_2+\vartheta)\left(A_P^*+A_F^*\right)\\
&-\beta_1(\beta_2\left(\vartheta^2A_F^*+S^*\left(\vartheta\left(\delta_2+\vartheta\right)-\beta_3I^{*2}(d+\theta+\vartheta)\right)\right)\\
&\left.+\beta_3\vartheta\left(\vartheta A_P^*+S^*\left(\delta_1+\vartheta\right)\right)\right)))\\
&\frac{1}{(\varepsilon I^*+1)^2}\left(\left(\delta_1+\delta_2+\vartheta\right)(\beta_2\vartheta^2\left(-A_F^*\right)\right.\\
&-\beta_3\left(\vartheta^2A_P^*-\beta_2I^{*2}(d+\theta+\vartheta)\right))\\
&+\beta_1\vartheta\left(\beta_2I^{*2}(d+\theta+\vartheta)+\beta_3I^{*2}(d+\theta+\vartheta)\right.\\
&\left.\left.-S^*\vartheta\left(\delta_1+\delta_2+\vartheta\right)\right)\right)\\
&-\frac{\beta_1\beta_2\beta_3S^*I^{*2}\vartheta\left(A_P^*+A_F^*+S^*\right)}{(\varepsilon I^*+1)^4}\\
&+\frac{1}{(bI^*+1)^2(\varepsilon I^*+1)^3}\left(a(\beta_1I^*\left(\beta_2I^*\left(\beta_3S^*I^*\right.\right.\right.\\
&\left.\left.+\varepsilon I^*\vartheta+\vartheta\right)+\vartheta(\varepsilon I^*+1)\left(\varepsilon I^*\vartheta+\beta_3I^*+\vartheta\right)\right)\\
&+(\delta_1+\delta_2+\vartheta)(\varepsilon I^*+1)\left(I^*\vartheta\varepsilon+\vartheta+\beta_2I^*\right)\\
&\left.\left(\varepsilon\vartheta I^*+\vartheta+\beta_3I^*\right)\right)),
\end{aligned}
$$

$$
\begin{aligned}
Q_1=&\frac{1}{(bI^*+1)^2(1+\varepsilon I^*)^4}\left(\beta_1(\beta_2I^*((1+\varepsilon I^*)(aI^*(1+\varepsilon I^*)\right.\\
&+(1+bI^*)^2(I^*(1+\varepsilon I^*)(d+\theta+2\vartheta)-2\vartheta S^*))\\
&-(bI^*+1)^2(\beta_3S^*I^*(A_P^*+S^*-I^*(\varepsilon I^*+1))\\
&+A_F^*\left(\beta_3S^*I^*+2(\varepsilon I^*\vartheta+\vartheta)\right)+\delta_2S^*(\varepsilon I^*+1)))\\
&+(\varepsilon I^*+1)(\beta_3I^*((1+\varepsilon I^*)aI^*\\
&-2\vartheta A_P^*(1+bI^*)^2+(1+bI^*)^2(I^*(1+\varepsilon I^*)\\
&(\theta+d+2\vartheta)-2S^*\vartheta)\\
&-(1+bI^*)^2\delta_1S^*)+\vartheta(1+\varepsilon I^*)(2aI^*(1+\varepsilon I^*)\\
&+(1+bI^*)^2(I^*(1+\varepsilon I^*)(2(d+\theta)+3\vartheta)-3\vartheta S^*)\\
&-2\left(\delta_1+\delta_2\right)S^*(1+bI^*)^2)))\\
&+(1+\varepsilon I^*)(\beta_2(I^*(\beta_3(I^*(\varepsilon I^*+1)\\
&\left(a+(1+bI^*)^2(\theta+d+2\vartheta)\right)\\
&+(1+bI^*)^2(1+\varepsilon I^*)((\delta_1+\delta_2)I^*\\
&-A_P^*\left(\delta_1+\delta_2+2\vartheta\right)))+(\varepsilon I^*+1)^2((\delta_1+\delta_2)\\
&\times\left(a+(1+bI^*)^2(\theta+d+2\vartheta)\right)
\end{aligned}
$$

$$
K_{12}=-\frac{\beta_1\beta_2\beta_3S^*I^{*2}\left(A_P^*+A_F^*+S^*-I^*(1+\varepsilon I^*)\right)}{(1+\varepsilon I^*)^4},
$$

$$
K_{13}=\frac{\beta_1\beta_2\beta_3S^*I^{*2}\left((1+\varepsilon I^*)aI^*-\vartheta(1+bI^*)^2\left(A_P^*+A_F^*\right)+(1+bI^*)^2(I^*(\theta+d+\vartheta)(1+\varepsilon I^*)-S^*\vartheta)\right)}{(bI^*+1)^2(\varepsilon I^*+1)^4}.
$$

The coefficients of the characteristic equation (36) are obtained as follows:

$$
\begin{aligned}
Q_0=&\vartheta^2\left(\delta_1+\delta_2+\vartheta\right)(d+\theta+\vartheta)\\
&+\frac{I^*\vartheta\left(\left(\beta_2+\beta_3\right)\left(\delta_1+\delta_2+\vartheta\right)+\beta_1\vartheta\right)(d+\theta+\vartheta)}{\varepsilon I^*+1}\\
&+\vartheta(2a+(bI^*+1)^2(2(d+\theta)+3\vartheta))))\\
&-A_F^*(bI^*+1)^2(\vartheta\left(2\delta_1+2\delta_2+3\vartheta\right)(\varepsilon I^*+1)\\
&+\beta_3I^*(\delta_1+\delta_2+2\vartheta)))\\
&+(\varepsilon I^*+1)(\beta_3(I^*(\varepsilon I^*+1)\\
&\times((\delta_1+\delta_2)(a+(1+bI^*)^2(d+\theta+2\vartheta))
\end{aligned}
$$

$$
\begin{aligned}
&+\vartheta\left(2a+(1+bI^*)^2(2(d+\theta)+3\vartheta)\right))\\
&-\vartheta A_P^*(bI^*+1)^2\\
&\times(2\delta_1+2\delta_2+3\vartheta))\\
&+\vartheta(\varepsilon I^*+1)^2\times((\delta_1+\delta_2)\\
&\left(2a+(1+bI^*)^2(2(d+\theta)+3\vartheta)\right)\\
&+\vartheta\left(3a+(1+bI^*)^2(4\vartheta+3(\theta+d))\right)))))\Big),
\end{aligned}
$$

$$
\begin{aligned}
Q_2=&\frac{1}{(1+\varepsilon I^*)^3(1+bI^*)^2}\Big(\beta_1((1+\varepsilon I^*)\\
&\left((1+\varepsilon I^*)aI^*+(1+bI^*)^2(I^*(1+\varepsilon I^*)\right.\\
&\left.(\theta+d+3\vartheta)-3S^*\vartheta\right)\\
&-(bI^*+1)^2\left(\beta_3 I^*\left(A_P^*+S^*-I^*(\varepsilon I^*+1)\right)\right.\\
&\left.+\beta_2 I^*\left(A_F^*+S^*-I^*(\varepsilon I^*+1)\right)\right.\\
&\left.+(\delta_1+\delta_2)S^*(\varepsilon I^*+1)\right))\\
&+\beta_2(I^*(1+\varepsilon I^*)^2\left(a+(1+bI^*)^2(d+\theta+3\vartheta)\right)\\
&+(1+bI^*)^2(I^*(\beta_3\left(-A_P^*+\varepsilon I^{*2}+I^*\right)\\
&+(\delta_1+\delta_2)(\varepsilon I^*+1)^2)\\
&-A_F^*\left((\delta_1+\delta_2+3\vartheta)(\varepsilon I^*+1)+\beta_3 I^*\right)))\\
&+(\varepsilon I^*+1)\\
&\times(\beta_3(I^*(1+\varepsilon I^*)\left(a+(1+bI^*)^2(\theta+3\vartheta+d)\right)\\
&+(1+bI^*)^2((\delta_1+\delta_2)I^*(\varepsilon I^*+1)\\
&-A_P^*(\delta_1+\delta_2+3\vartheta)))+(\delta_1+\delta_2)(\varepsilon I^*+1)^2\\
&\left(a+(1+bI^*)^2(\theta+d+3\vartheta)\right)))\\
&+3\vartheta\left(\frac{a}{(1+bI^*)^2}+d+\theta+2\vartheta\right),
\end{aligned}
$$

$$
\begin{aligned}
Q_3=&\frac{a}{(bI^*+1)^2}\\
&+\frac{\beta_3\left(-A_P^*+\varepsilon I^{*2}+I^*\right)+\beta_2\left(-A_F^*+\varepsilon I^{*2}+I^*\right)+\beta_1\left(-S+\varepsilon I^{*2}+I^*\right)}{(\varepsilon I^*+1)^2}\\
&+d+\delta_1+\delta_2+\theta+4\vartheta.
\end{aligned}
$$

References

1. Kermack, W.O., Mckendrick, A.G.: A contribution to the mathematical theory of epidemic. Proc. R. Soc. Lond. **115**(772), 700–721 (1927)
2. Capasso, V., Serio, G.: A generalization of the Kermack Mckendrick deterministic epidemic model. Math. Biosci. **42**(1–2), 43–61 (1978)
3. Anderson, R.M., May, R.M.: Regulation and stability of host-parasite population interactions: I. regulatory processes. J. Anim. Ecol. **47**(1), 219–267 (1978)
4. Cooke, K.L.: Stability analysis for a vector disease model. Rock. Mount. J. Math. **9**, 31–42 (1979)
5. Guckenheimer, J., Holmes, P.: Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields. Springer, Berlin (1983)
6. Hale, J.K., Lunel, S.M.V.: Introduction to Functional Differential Equations. Springer, New York (1993)
7. Hethcote, H.W., van den Driessche, P.: An SIS epidemic model with variable population size and a delay. J. Math. Biol. **34**, 177–194 (1995)
8. van den Driessche, P., Watmough, J.: Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math. Biosci. **180**(1–2), 29–48 (2002)
9. Murray, J.D.: Introduction to Mathematical Biology: I. An Introduction. Springer, Berlin (2002)
10. Wang, W., Ruan, S.: Bifurcation in an epidemic model with constant removal rates of the infectives. J. Math. Anal. Appl. **21**, 775–793 (2004)
11. Castillo-Chavez, C., Song, B.: Dynamical models of tuberculosis and their applications. Math. Biosci. Eng. **1**(2), 361–404 (2004)
12. Wang, X.: A simple proof of Descartes's rule of signs. Am. Math. Mon. **111**(6), 525–526 (2004)
13. Song, X., Cheng, S.: A delay-differential equation model of HIV infection of CD4+ T-cells. J. Kor. Math. Soc. **42**(5), 1071–1086 (2005)
14. Liu, Y., Cui, J.: The impact of media convergence on the dynamics of infectious diseases. Int. J. Biomath. **1**(1), 65–74 (2008)
15. Zhang, J.Z., Jin, Z., Liu, Q.X., Zhang, Z.Y.: Analysis of a delayed SIR model with nonlinear incidence rate. Discret. Dyn. Nat. Soc. **2008**, 636153 (2008)
16. Li, X.Z., Li, W.S., Ghosh, M.: Stability and bifurcation of an SIR epidemic model with nonlinear incidence and treatment. Appl. Math. Comput. **210**(1), 141–150 (2009)
17. Jones, J.H., Salathé, M.: Early assessment of anxiety and behavioral response to novel swine-origin influenza A(H1N1). PLoS One **4**(12), e8032 (2009)
18. Pappas, G., Kiriaze, I.J., Giannakis, P., Falagas, M.E.: Psychosocial consequences of infectious diseases. Clin. Microbiol. Infect. **15**(8), 743–747 (2009)
19. Funk, S., Gilad, E., Jansen, V.A.A.: Endemic disease, awareness, and local behavioral response. J. Theor. Biol. **264**(2), 501–509 (2010)
20. Kiss, I.Z., Cassell, J., Recker, M., Simon, P.L.: The impact of information transmission on epidemic outbreaks. Math. Biosci. **225**(1), 1–10 (2010)
21. Misra, A.K., Sharma, A., Singh, V.: Effect of awareness programs in controlling the prevalence of an epidemic with time delay. J. Biol. Syst. **19**(2), 389–402 (2011)
22. Cai, L., Li, X., Ghosh, M.: Global dynamics of a mathematical model for HTLV-I infection of CD4+ T-cells. Appl. Math. Model. **35**(7), 3587–3595 (2011)
23. Hatzopoulos, V., Taylor, M., Simon, P.L., Kiss, I.Z.: Multiple sources and routes of information transmission: implications for epidemic dynamics. Math. Biosci. **231**, 197–209 (2011)
24. Zhou, L., Fan, M.: Dynamics of an SIR epidemic model with limited medical resources revisited. Nonlinear Anal.: Real World Appl. **13**(1), 312–324 (2012)
25. Hattaf, K., Lashari, A.A., Louartassi, Y., Yousfi, N.: A delayed SIR epidemic model with general incidence rate. Electron. J. Qual. Theory Differ. Equ. **3**, 1–9 (2013)
26. Dubey, B., Patra, A., Srivastava, P.K., Dubey, U.S.: Modeling and analysis of an SEIR model with different types of nonlinear treatment rates. J. Biol. Syst. **21**(03), 1350023 (2013)

27. Zhao, H., Lin, Y., Dai, Y.: An SIRS epidemic model incorporating media coverage with time delay. Comput. Math. Methods Med. **2014**, 680743 (2014)
28. Zuo, L., Liu, M.: Effect of awareness programs on the epidemic outbreaks with time delay. Abstr. Appl. Anal. **2014**, 940841 (2014)
29. Zuo, L., Liu, M., Wang, J.: The impact of awareness programs with recruitment and delay on the spread of an epidemic. Math. Problems Eng. **2015**, 235935 (2015)
30. Buonomo, B., Cerasuolo, M.: The effect of time delay in plant-pathogen interactions with host demography. Math. Biosci. Eng. **12**(3), 473–490 (2015)
31. Pruyt, E., Auping, W.L., Kwakkel, J.H.: Ebola in West Africa: model-based explo- ration of social psychological effects and interventions. Syst. Res. Behav. Sci. **32**, 2–14 (2015)
32. Yang, M., Sun, F.: Global stability of SIR models with nonlinear Incidence and discontinuous treatment. Electron. J. Differ. Equ. **2015**(304), 1–8 (2015)
33. Greenhalgh, D., Rana, S., Samanta, S., Sardar, T., Bhattacharya, S., Chattopadhyay, J.: Awareness programs control infectious disease - multiple delay induced mathematical model. Appl. Math. Comput. **251**, 539–563 (2015)
34. Dubey, B., Dubey, P., Dubey, U.S.: Role of media and treatment on an SIR model. Nonlinear Anal.: Modell. Control **21**(2), 185–200 (2015)
35. Kumar, A.N.: Stability of a time delayed SIR epidemic model along with nonlinear incidence rate and Holling Type-II treatmemt rate. Int. J. Comput. Methods **15**(6), 1850055 (2018)
36. Goel, K., Nilam, : A mathematical and numerical study of a SIR epidemic model with time delay, nonlinear incidence and treatment rates. Theory Biosci. **138**(2), 203–213 (2019)
37. Kumar, A., Nilam, : Mathematical analysis of a delayed epidemic model with nonlinear incidence and treatment rates. J. Eng. Math. **115**, 1–20 (2019)
38. Upadhyay, R.K., Pal, A.K., Kumari, S., Roy, P.: Dynamics of an SEIR epidemic model with nonlinear incidence and treatment rates. Nonlinear Dyn. **96**(4), 2351–2368 (2019)
39. Goel, K., Nilam, : Stability behavior of a nonlinear mathematical epidemic transmission model with time delay. Nonlinear Dyn. **98**, 1501–1518 (2019)
40. Goel, K., Kumar, A., Nilam, : A deterministic time-delayed SVIRS epidemic model with incidences and saturated treatment. J. Eng. Math. **121**, 19–38 (2020)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/342676330

Numerical heat transfer analysis & predicting thermal performance of fins for a novel heat exchanger using machine learning

Article *in* Case Studies in Thermal Engineering · July 2020

DOI: 10.1016/j.csite.2020.100706

CITATIONS
0

READS
7

3 authors, including:



Gaurav Krishnayatra
Delhi Technological University
3 PUBLICATIONS **0** CITATIONS
SEE PROFILE



Sulekh Tokas
Delhi Technological University
2 PUBLICATIONS **0** CITATIONS
SEE PROFILE

All content following this page was uploaded by Sulekh Tokas on 14 July 2020.

The user has requested enhancement of the downloaded file.

Case Studies in Thermal Engineering 21 (2020) 100706

Contents lists available at ScienceDirect

Case Studies in Thermal Engineering

journal homepage: http://www.elsevier.com/locate/csite

Numerical heat transfer analysis & predicting thermal performance of fins for a novel heat exchanger using machine learning

Gaurav Krishnayatra *, Sulekh Tokas, Rajesh Kumar

Department of Mechanical Engineering, Delhi Technological University, Delhi, India

ARTICLE INFO

Keywords:
Convection heat transfer
Fin effectiveness
Fin efficiency
Novel fins
Machine learning
K-nearest neighbor

ABSTRACT

In the present case study, the thermal performance of fins for a novel axial finned-tube heat exchanger is investigated and predicted using machine learning regression technique. The effects of variation in the fin spacing, fin thickness, material, and the convective heat transfer coefficient on the overall efficiency and total effectiveness have been analyzed and commented upon. The *k*-Nearest Neighbor (*k*-NN), a machine learning algorithm, is used for regression analysis to predict the thermal performance outputs and the results showed high prediction accuracies. The *k*-NN algorithm is robust and precise which can be used by thermal system design engineers for predicting output variables. The temperature profiles of various geometries have been depicted and compared in the results. It was concluded that the efficiency is increasing with fin thickness & decreasing with fin spacing and the maximum efficiency $\eta_{max} = 0.99975$ is achieved at $\delta^* = 0.1$ & $t^* = 0.0133$ having $h = 5\ W/m.K$ for copper material. The effectiveness is increasing with fin spacing & fin thickness and the maximum effectiveness $\varepsilon_{max} = 122.766$ is for $\delta^* = 8$ & $t^* = 0.4$ having $h = 5\ W/m.K$.

1. Introduction

Extended surfaces (fins) are widely employed in various heat transfer components used in a wide range of applications like heat sinks, microprocessor cooling & latent thermal energy storage devices [1,2] and thus the fins have always been a notable domain for research. Increasing the exposed surface area by employing the fins is a prominent method to enhance the rate of heat transfer with the advantage of being economically feasible. Extensive studies are available which analyze the thermal performance and costing of different types of fins. Harper and Brown [3] presented two-dimensional analytical solutions for circumferential and longitudinal fins having rectangular and wedge type shape. It was deduced that the one-dimensional model is sufficient to evaluate the performance of the fin. Schmidt [4] examined the performance of various profiles for longitudinal fins and concluded that the parabolic profile is the optimum type of profile for longitudinal fins. Heggs & Ooi [5] evaluated the performance of radial rectangular fins and presented variations in thermal performance for different designs. Alam & Ghoshdastidar [6] performed numerical study using FDM on a circular tube with internal longitudinal fins having a tapered lateral profile, it was observed that there is a significant improvement in the net heat transfer rate with the installation of fins. Nagarani et al. [7] performed optimization using genetic algorithm on the geometrical parameters of the elliptical annular fin (EAF). The result showed that the EAF is more effective than the conventional circular fin for a

* Corresponding author.
E-mail address: gauravkrishnayatra_mt2k18@dtu.ac.in (G. Krishnayatra).

https://doi.org/10.1016/j.csite.2020.100706
Received 3 June 2020; Received in revised form 25 June 2020; Accepted 27 June 2020
Available online 4 July 2020
2214-157X/© 2020 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

fixed cross-sectional area irrespective of the efficiency when the shape factor (SF) is less than 0.5. Cuce & Cuce [8] used CFD to investigate the performance of novel rectangular fin designs with single-step change (RFSSC) and double-step change (RFDSC). The RFSSC fins were found to be more efficient than the conventional rectangular fin (CRF). Aziz and Khani [9] developed analytical solutions to examine the performance of rectangular and various convex parabolic fins installed on a rotating shaft using homotopy analysis method (HAM). The highest rate of heat transfer was recorded for uniform fin thickness and the effectiveness increased with the rotational speed of the shaft. Okiy [10] investigated the effect of two-dimensional heat transfer in radial and planar fins in comparison to the assumption of one-dimensional heat transfer in fin analysis. It was concluded that one-dimensional heat transfer assumption provided inaccurate solutions with respect to two-dimensional heat transfer for short planar fins. Samana et al. [11] experimentally examined the fin efficiency of a solid wire fin used in a wire-on-tube heat exchanger and the fin efficiency increased with the use of oscillating heat pipe in comparison to solid wire fin. Li et al. [12] performed experiments to study the effects of various thermal boundary conditions and thermal conductivity on the thermal performance of a pin fin array using liquid crystal sheet, and it was concluded that thermal conductivity is directly influencing the rate of heat transfer.

After a thorough literature review, it was deduced that some designs of finned-tube heat exchangers are particularly overlooked and many studies have utilized the artificial neural network [13,14] however, very limited studies have utilized the *k*-Nearest Neighbor for regression analysis. A novel type of fins is designed and its thermal performance has been analyzed in this computational investigation. The finned-tube heat exchanger consists of a hollow tube with an isothermal boundary condition at the inner wall (in order to simulate some of the industrial applications) having the novel rectangular fins attached axially at the outer wall. The novel design of fins consists of a fin-on-fin approach, where in Fig. 1 it is visible that the tube has primary fins (colored in black), this is a simple case of longitudinal fins attached to the outer surface of the tube. But consequently another set of fins, the secondary fins (colored in blue), are attached on to the primary fins in order to increase the total finned area that proportionately enhance the total rate of heat transfer. In this study, the variations of the overall Efficiency and the Effectiveness of the novel finned-tube system has been analyzed with respect to the varying fin thickness, fin spacing, material of the system, and the convective heat transfer coefficient using Ansys Fluent. Moreover, Machine learning technique called *k*-Nearest Neighbor (*k*-NN) has been employed for regression analysis to predict the thermal performance of the novel finned-tube system.

2. Physical model & governing equations

The design consists of a tube with the inner diameter (D_i) and the outer diameter (D_o) of 5 & 6 mm respectively with the fin thickness (t) being varied from 0.2 to 2.4 mm and having the fin spacing (δ) from 0.6 to 48 mm as shown in Fig. 1. The two materials of the finned-tube system chosen are Copper $(K_{cu} = 387.6\ W/m.K)$ and Steel $(K_{ss} = 16.7\ W/m.K)$. The conjugated heat transfer is taking place from the inner isothermal wall $(T_w = 400\ K)$ to the ambient air $(T_\infty = 300\ K)$. To reduce computational cost and time, the simulated flow of air on the outer surface of the finned-tube system has not been considered. However, a range of convective heat transfer coefficient has been taken which covers almost all the practical applicated values found in heat exchangers in order to visualize the impact of its variation on the efficiency and the effectiveness of the system. The convective heat transfer coefficient (h) at the outer surface is being varied from 5 to 200 W/m.K.

The governing partial differential equations used to solve the conjugated heat transfer problem are linearized and computed in the Ansys Fluent using with the following assumptions.

1. The material is homogenous and isotropic.
2. The fin tips are adiabatic.
3. The convective heat transfer coefficient is uniform over the entire outer surface area.
4. The radiation phenomenon is neglected.

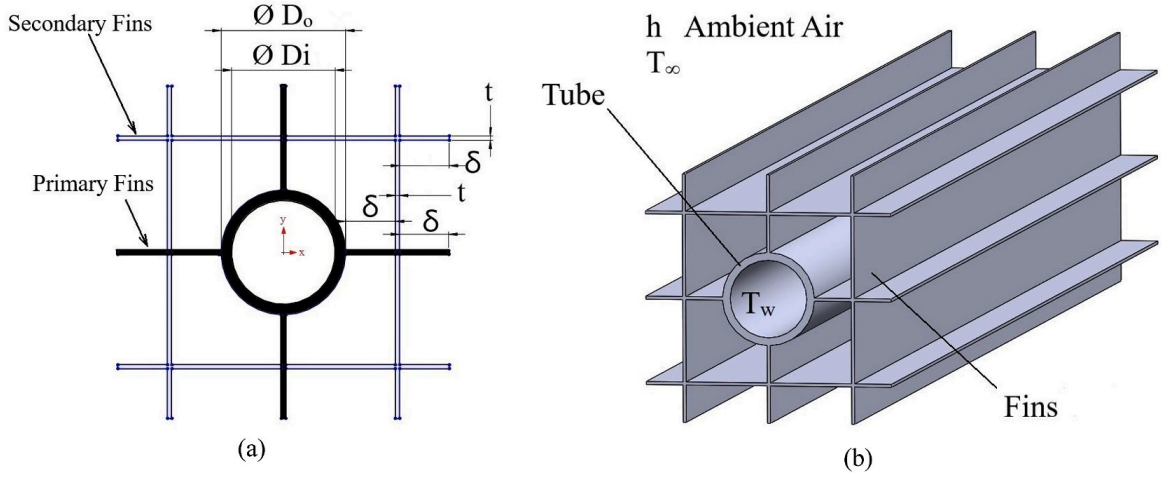


Fig. 1. a) Schematic model of the novel finned-tube heat exchanger b) 3-Dimensional Representation of the model.

The energy equation for the finned-tube system [1]:

$$\left\{\frac{d^2T}{dx^2}+\frac{d^2T}{dy^2}\right\}=0 \tag{1}$$

The heat transfer equation for the outer surface [1]:

$$Q=\frac{(T_w-T_\infty)}{\frac{\ln\left(D_o/D_i\right)}{2\pi kL}+\frac{1}{\eta_o A_t h}} \tag{2}$$

Where, k is the thermal conductivity of the material, h is the convective heat transfer coefficient and η_o is the overall efficiency of the finned-tube [1].

$$\eta_o=1-\frac{NA_f}{A_t}\left(1-\eta_f\right) \tag{3}$$

Where, η_f is the efficiency of a single fin, A_f the surface area of single fin and A_t the total surface area of the finned-tube system. The total surface area of the novel finned-tube system is calculate by:

$$A_t=L_{exp}\times L \tag{4}$$

$$L_{exp}=48\delta+8(D_o-t) \tag{5}$$

Where, axial length (L) of the system is unity and L_{exp} is the total perimetric length of the outer surface of the novel finned-tube system.

Equations (2) and (3) are used to validate the computational results for a simple finned-tube system having 4 equidistant axial fins attached at the outer surface of a hollow tube having same diameters with fin length $L_f = 10$ *mm* and fin thickness $t = 0.5$ *mm* as shown in Fig. 2 along with the heat flux variation with convective heat transfer coefficient which is showing remarkable agreement.

The overall Efficiency for the system is given as:

$$\eta=\frac{q}{h(T_w-T_\infty)} \tag{6}$$

Where, q is the area weighted average of the heat flux at the outer surface of the novel finned-tube system computed from the simulations. And the total effectiveness of the system is:

$$\varepsilon=\frac{Q}{hA_o(T_w-T_\infty)} \tag{7}$$

Where, Q is the area surface integral of the heat flux at the outer surface computed by simulation.

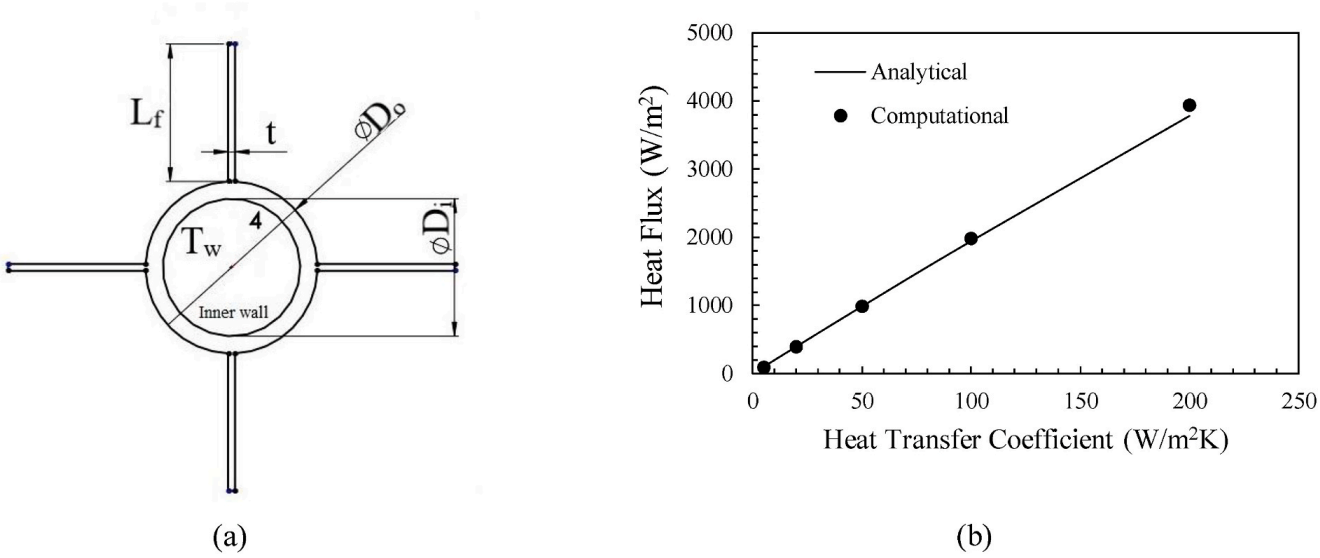


Fig. 2. a) Schematic finned-tube system used for convective heat transfer validation with 4 axial fins at the outer surface of the tube b) Variation of heat flux with convective heat transfer coefficient representing the comparing between analytical and computational heat flux values.

3. Computational & machine learning modelling

3.1. Computational modelling

A 2-Dimensional analysis is performed in this study in order to reduce the computational time. Moreover, the 2D model of the finned-tube system is symmetrical about 4 axes, therefore the $1/8^{th}$ 2D model is selected for the study as shown in Fig. 3. The 2D models were developed using Solidworks software and then were modeled & solved in Ansys Fluent. The models were discretized into a finite number of control volumes on which the governing equations were solved using appropriate approximations. Second-order upwind was used for energy equation and the residual value was set to 10^{-9}. Tetrahedron meshing was done on the solid domain as shown in Fig. 3 and the mesh convergence test was also performed to make the result independent of the size of the elements. Table 1 is showing the variation of heat flux with the element size and it can be inferred that the heat flux is getting independent when the size of the element is at 0.0125 mm.

3.2. Machine learning modelling

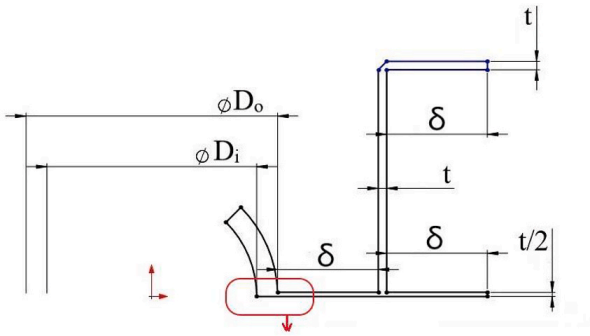
This section explains the utilization of supervised machine learning technique called *k*-Nearest Neighbor (*k*-NN) which has been applied as a regression tool [15]. The *k*-NN algorithm is robust, precise and computationally fast when used as a regression model, moreover it works well with a fewer number of input variables and relatively small data set which is used in designing of thermal systems. The dataset for the regression analysis is procured from the simulations result for all the geometries considered in the study, the dataset includes the output variables namely the efficiency (η) & the effectiveness (ε) of the finned-tube heat exchanger and the outer diameter of tube (D_o), fin spacing (δ), fin thickness (t), thermal conductivity of the material (k), and the convective heat transfer coefficient (h) are the input variables. Therefore, the variables are represented in the forms of following functions.

$$\eta = f_1(D_o,\ \delta,\ t,\ k,\ h) \tag{8}$$

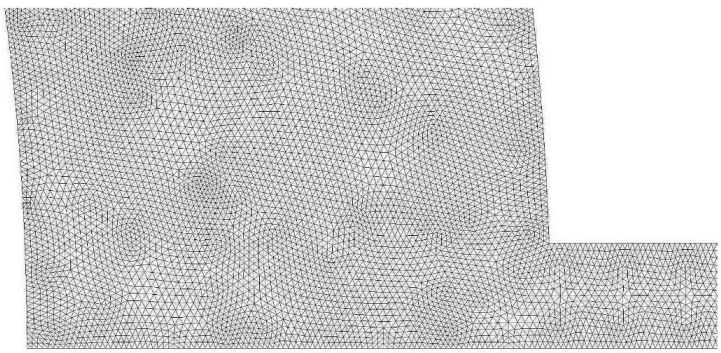
$$\varepsilon = f_2(D_o,\ \delta,\ t,\ k,\ h) \tag{9}$$

In order to dimensionally balance the equation, the input variables are converted into non-dimensional forms subsequently decreasing the number of input variables.

$$\eta = f_3(\delta^*,\ t^*,\ M) \tag{10}$$



(a)



(b)

Fig. 3. a) $1/8^{th}$ model designed for the simulation b) The red encircled portion of the $1/8^{th}$ model is shown after descritization

Table 1
Variation of heat flux with number of elements and element size.

Max. Element Size (mm)	Number of Elements	Heat Flux $q(W/m^2)$	Difference in q (%)
0.05	31392	498.187	0.212
0.04	51936	497.129	1.913
0.03	96128	487.796	2.154
0.02	191072	498.536	-0.266
0.015	344592	497.210	0.057
0.0125	461472	497.494	-0.012
0.01	601844	497.555	–

$$\varepsilon = f_4(\delta^*,\ t^*,\ M) \tag{11}$$

Where, δ^* is the non-dimensional fin spacing, t^* is the non-dimensional fin thickness and M is a new non-dimensional parameter used for this study, relations given below:

$$\delta^* = \frac{\delta}{D_o} \tag{12}$$

$$t^* = \frac{t}{D_o} \tag{13}$$

$$M = \frac{h}{8k} L_{exp} \tag{14}$$

The k-NN algorithm is a "black-box" regression tool where the actual functions are not derived instead they are predicted from the dataset. The dataset is divided into two parts: 65% Testing data & 35% Training data, the training dataset is used for "learning" and the testing dataset is used for validating the predictions with actual output values. The prediction accuracy of k-NN regression is calibrated by two factors namely, the "k" number of close neighbors and the type of metric distance between the outputs [16]. In Fig. 4, the output y_i is scattered with respect to the inputs x_i & z_i for $k = 3$ number of neighbors. The p_i is the prediction computed on the basis of the median of Euclidean distances $(d_{12},\ d_{13}\ \&\ d_{14})$ of y_1 from $y_2,\ y_3\ \&\ y_4$, which are the 3 closest neighboring outputs. The Euclidean distance is calculated by

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (z_i - z_j)^2} \tag{15}$$

Where, d_{ij} calculates the Euclidean distance between the ith and jth elements in the input data.

The predicted values of the outputs are compared to the computed outputs from the simulation results in terms of R^2 values to check the accuracy of the regression model and the selection of "k" is based on the accuracy of the regression model. The R^2 value is the

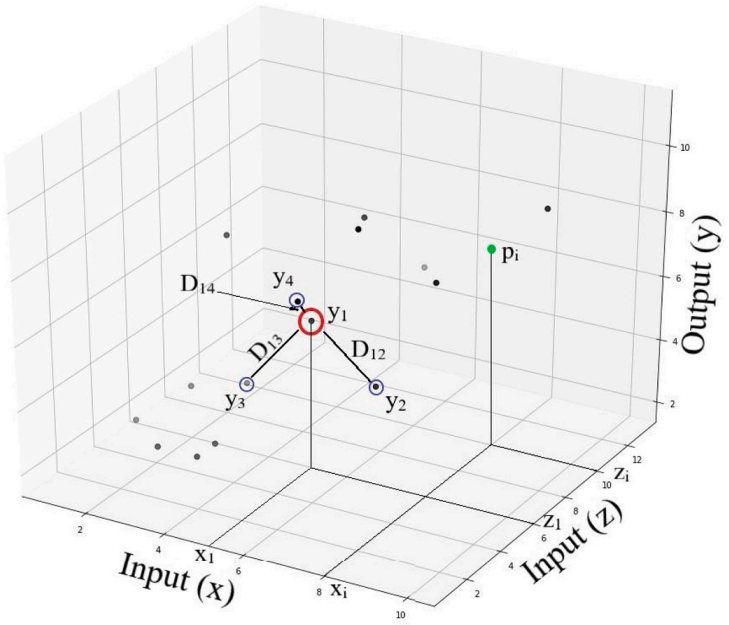


Fig. 4. Three-Dimensional scattering of output variable shown with variation of inputs 1 & 2 for $k = 3$ number of neighbors.

coefficient of determination which is given by the following relation.

$$R^2 = 1 - \frac{\sum_i (y_i - p_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{16}$$

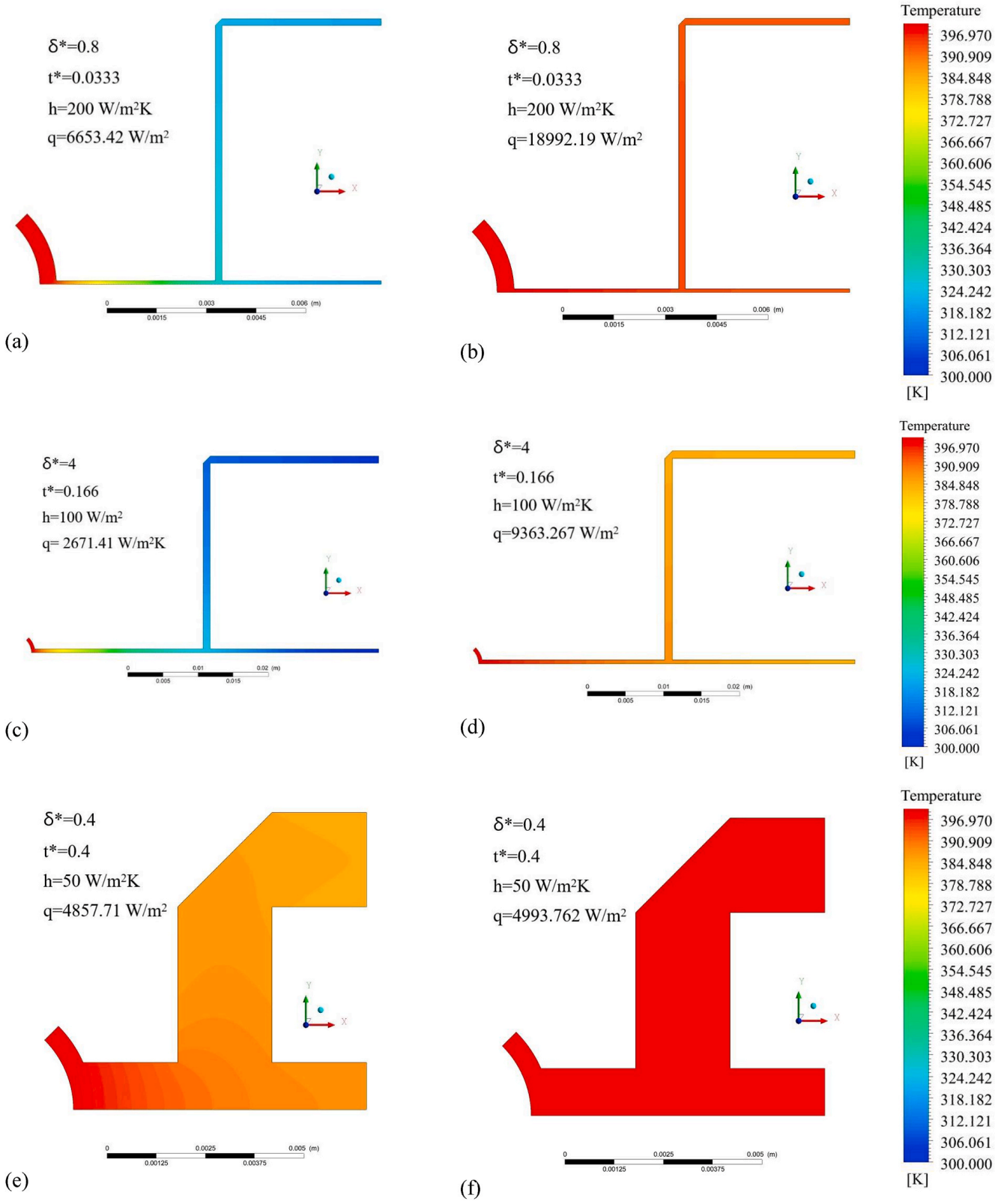
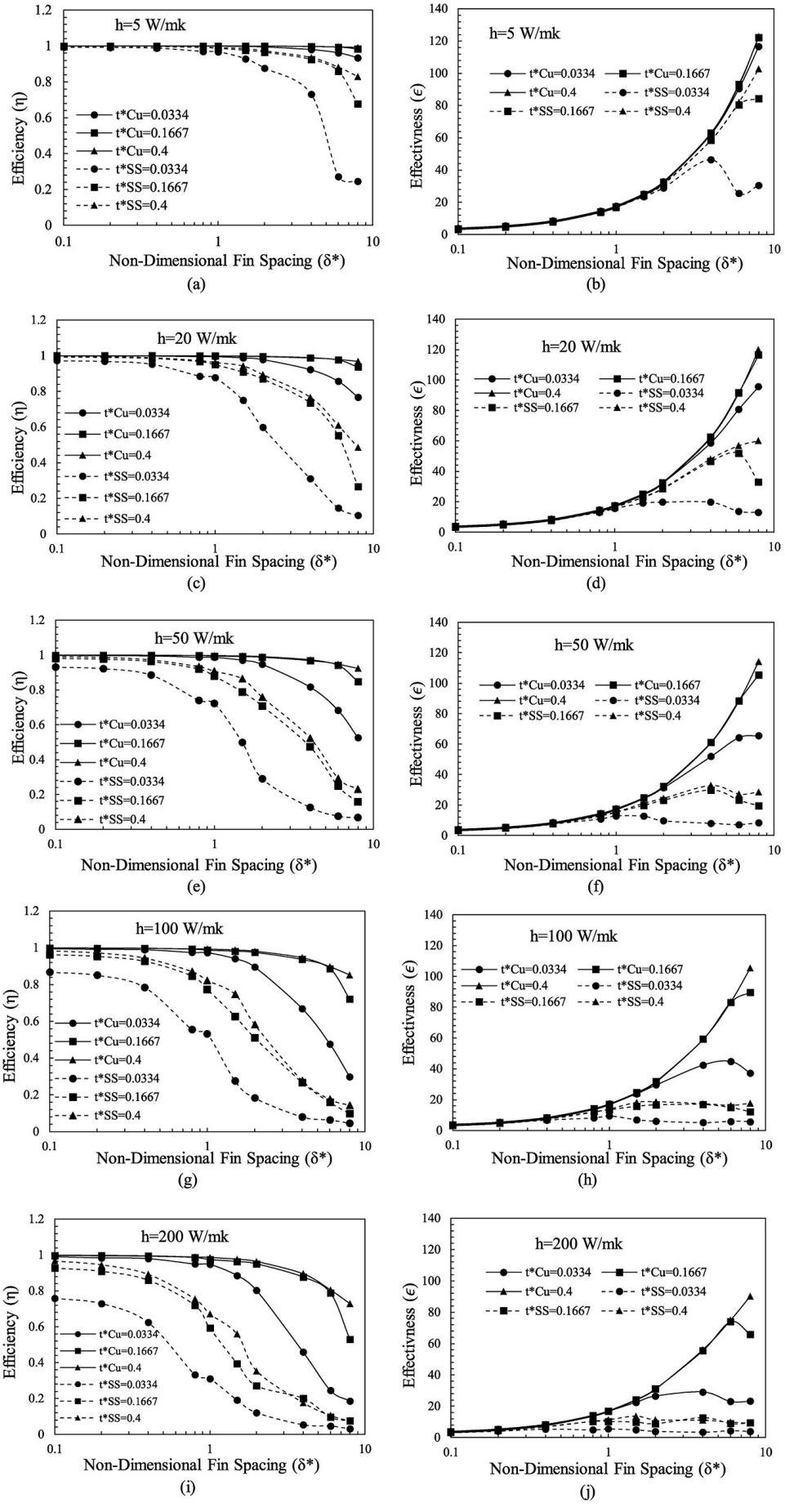


Fig. 5. Temperature plumes of cross-section for different geometries and different convective heat transfer coefficients.



(caption on next page)

Fig. 6. Thermal performance of the novel finned-tube heat exchangers with respect to non-dimensional fin spacing for various non-dimensional fin thickness and convective heat transfer coefficients.

Where, y_i & p_i are the output variable and the corresponding predicted value for i^{th} element in the data and $\bar{y}$ is the mean.

4. Results & discussion

This study provides insights which explain the variations of the overall efficiency (η) and the total effectiveness (ε) of a novel finned-tube system with respect to the fin spacing, the fin thickness, thermal conductivity of the material and the convective heat transfer coefficient. The number of parameters is reduced and converted to non-dimensional forms. Therefore, the non-dimensional fin spacing $\left(\delta^* = \frac{\delta}{D_o}\right)$, non-dimensional fin thickness $\left(t^* = \frac{t}{D_o}\right)$ & the convective heat transfer coefficient is being varied from 0.1 to 8, from 0.0333 to 0.4 and from 5 to 200 W/m^{2}K respectively for copper and steel fins. Fig. 5 shows the temperature profiles having same temperature bound for various geometrical models along with the heat flux values. It can be inferred that the system having copper material is showing more rate of heat transfer due to the fact that copper has higher thermal conductivity. The temperature gradients developed along the fin length for copper fins is lower than that for steel fins, consequently increasing the efficiency of copper fins. The drop in temperature along the fins length is increasing with the increase in fin thickness. In Fig. 6, the overall efficiency and total effectiveness of the novel finned-tube system is represented with respect to the δ^* for different values of t^*. For steel fins the decrease in fin efficiency is gradual till $\delta^* = 1$, the it is decreasing rather quickly with δ^* in contrast to copper fins. Moreover, the rate of decrease of efficiency is also directly proportional to the convective heat transfer coefficient. It can be observed that for same values of δ^* & t^*, higher values of h are showing larger decrease in the efficiency. It can be asserted that overall efficiency of finned-tube system decreases with increase in δ^* & h, and it increases with increase in t^*. The maximum efficiency $\eta_{max} = 0.99975$ is achieved at $\delta^* = 0.1$ & $t^* = 0.0133$ for $h = 5\ W/m.K$ with copper material. The total effectiveness of the system for copper fins is seen to be insensitive to the

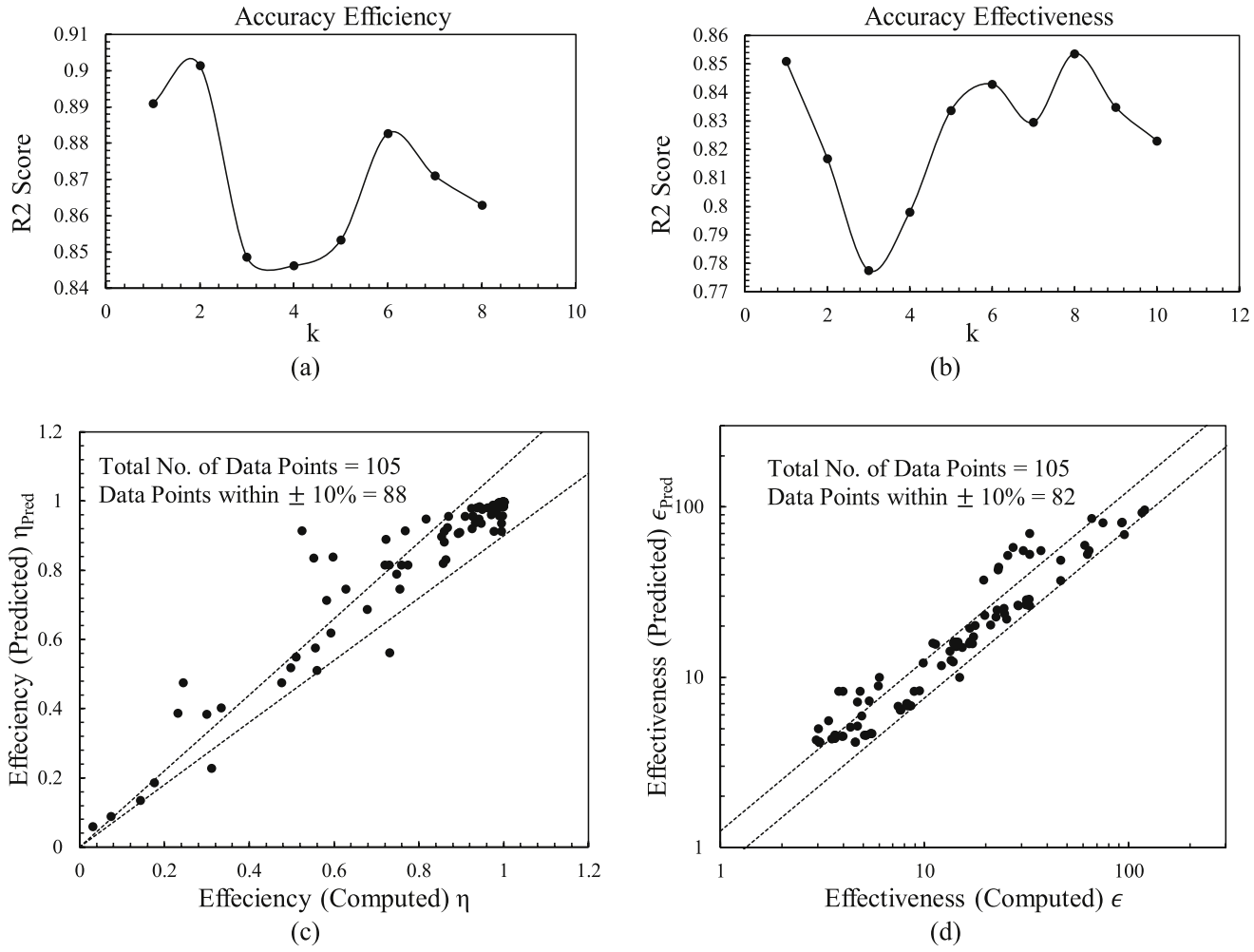


Fig. 7. Results of k-NN regression showing variation of R^2 score with k value (a & b) along with predicted and computed outputs for efficiency and effectiveness (c & d).

fin thickness for $h = 5$ & 20 W/m^2K, however, it is decreasing with the increase in δ^* for $h = 50$, 100 & 200 W/m^2K. The effectiveness of steel fins is observed to be decreasing with δ^* for all values of h. It can be concluded that the total effectiveness is increasing with increase in δ^* & t^* and is decreasing with h. The maximum effectiveness $\varepsilon_{max} = 122.766$ is for $\delta^* = 8$ & $t^* = 0.4$ for $h = 5\ W/\ m.K$.

The variation of R^2 values with "k" has been recorded and presented in Fig. 7 and it is clear that $k = 2$ is predicting the highest accuracy for Efficiency and $k = 8$ is giving the highest accuracy for Effectiveness predictions. The R^2 value for efficiency at $k = 2$ is 0.9014 and for effectiveness at $k = 8$ is 0.8537, which implies that the regression model predictions are in good agreement with the actual computed outputs as shown in Fig. 7 c & d. It can be observed that for Efficiency, 88 out of 105 data points are lying within the 10% error range and for effectiveness, 85 out of 105 data points are lying with the 10% error range. The k-NN regression algorithm is showing remarkable accuracies therefore, it is an exceptional tool for thermal system designer for predicting outputs without computationally solving the thermal system at its entirety.

5. Conclusion

The convective heat transfer coefficient (h) is being varied along with variation in fin geometry $(\delta$ & $t)$ for two materials copper and steel to analyze the effects on the thermal performance of this novel finned-tube heat exchanger. It is concluded that copper is the obvious choice of material for the finned-tube system. The efficiency has direct proportionality with fin thickness and inverse proportionality with fin spacing & heat transfer coefficient. The maximum efficiency is achieved with the shortest & thickest fin for the lowest heat transfer coefficient. The effectiveness is showing direct relation with fin spacing & fin thickness and inverse relation with the heat transfer coefficient. The maximum overall effectiveness of the system is achieved for the longest and thickest fin with the lowest heat transfer coefficient. The k-NN regression model resulted in high prediction accuracies and because of its advantages, k-NN can be an extremely useful tool for thermal system design engineers.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Gaurav Krishnayatra: Data curation, Conceptualization, Writing - original draft. **Sulekh Tokas:** Conceptualization, Writing - review & editing. **Rajesh Kumar:** Supervision, Writing - review & editing.

Nomenclature

D_o	Outer diameter of the tube, m
D_i	Inner diameter of the tube, m
δ	Fin spacing, m
L_f	Length of fin, m
t	Thickness of fin, m
T	Temperature (general), $^\circ C$
T_w	Temperature of inner wall, $^\circ C$
T_∞	Temperature of the free stream, $^\circ C$
L	Axial length of the hollow cylinder, m
t^*	Non-dimensional fin thickness, $t^* = t/D_o$
δ^*	Non-dimensional fin spacing, $\delta^* = \delta/D_o$
A_t	Total surface area, m^2
h	Convective heat transfer coefficient, $W/m^2.K$
q	Heat transfer per unit area, simulations result, W/m^2
Q	Total rate of heat transfer, W, $Q = \frac{(T_w - T_\infty)}{\frac{\ln\left(D_o/D_i\right)}{2\pi kL} + \frac{1}{\eta_o A_t h}}$
k_{ss}	Thermal conductivity of Steel, $W/m.K$
k_{cu}	Thermal conductivity of Copper, $W/m.K$
k	Thermal conductivity (general), $W/m.K$
η_o	Overall Efficiency of finned-tube system, $\eta_o = 1 - \frac{NA_f}{A_t}(1 - \eta_f)$
η_f	Efficiency of single fin
N	Number of Fins
A_f	Area of single fins, m^2
L_{exp}	Outer surface perimetric length of the finned-tube system, m, $L_{exp} = 112\delta + 16(D_o - t)$

η	Efficiency computed by computational results, $\eta = \frac{q}{h(T_w - T_\infty)}$
ε	Effectiveness computed by computational results, $\varepsilon = \frac{Q}{hA_o(T_w - T_\infty)}$
M	Non-dimensional parameter "M", $M = \frac{h}{8k}L_{exp}$
η_{pred}	Predicted Efficiency
ε_{pred}	Predicted Effectiveness
η_{max}	Maximum value of the Efficiency
ε_{max}	Maximum value of the Effectiveness
d_{ij}	Euclidean distance between ith and jth elements of the predictors, $d_{ij} = \sqrt{(x_i - x_j)^2 + (z_i - z_j)^2}$
x_i, z_i	Predictor variables for ith element
k	Number of neighbors used in k-NN algorithm
y_i	ith Output variable used for regression
p_i	Predicted Output for ith element
R^2	R-squared values, $R^2 = 1 - \frac{\sum_i (y_i - p_i)^2}{\sum_i (y_i - \bar{y})^2}$

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csite.2020.100706.

References

[1] A.D. Kraus, A. Aziz, J. Welty, Extended Surface Heat Transfer, John Wiley & sons, 2001.
[2] N. Nagarani, K. Mayilsamy, A. Murugesan, G.S. Kumar, Review of utilization of extended surfaces in heat transfer problems, Renew. Sustain. Energy Rev. 29 (2014) 604–613, https://doi.org/10.1016/j.rser.2013.08.068.
[3] D.R. Harper, W.B. Brown, Mathematical Equations for Heat Conduction in the Fins of Air Cooled Engines, vol. 158, National Advisory Committee on Aeronautics, Washington, DC, 1922. NACA Rep. https://ntrs.nasa.gov/search.jsp?R=19930091223.
[4] E. Schmidt, Die warmeuebertragung durch rippen, Z. Des. Vereines Dtsch. Ingenieure 70 (1926) 885–951.
[5] P.J. Heggs, T.H. Ooi, Design charts for radial rectangular fins in terms of performance ratio and maximum effectiveness, Appl. Therm. Eng. 24 (2004) 1341–1351, https://doi.org/10.1016/j.applthermaleng.2003.12.021.
[6] A. Alam, P.S. Ghoshdastidar, A study of heat transfer effectiveness of circular tubes with internal longitudinal fins having tapered profiles, Int. J. Heat Mass Tran. 45 (2002) 1371–1376, https://doi.org/10.1016/S0017-9310(01)00240-X.
[7] N. Nagarani, K. Mayilsamy, A. Murugesan, G.S. Kumar, Review of utilization of extended surfaces in heat transfer problems, Renew. Sustain. Energy Rev. 29 (2014) 604–613, https://doi.org/10.1016/j.proeng.2012.06.343.
[8] P.M. Cuce, E. Cuce, Optimization of configurations to enhance heat transfer from a longitudinal fin exposed to natural convection and radiation, Int. J. Low Carbon Technol. 9 (2014) 305–310, https://doi.org/10.1093/ijlct/ctt005.
[9] A. Aziz, F. Khani, Analytical solutions for a rotating radial fin of rectangular and various convex parabolic profiles, Commun. Nonlinear Sci. Numer. Simulat. 15 (2010) 1565–1574, https://doi.org/10.1016/j.cnsns.2009.07.008.
[10] R.E.K.V. Okiy, An assessment of extended surfaces-two dimensional effects, Int. J. Eng. Res. Afr. 15 (2015) 71–85. https://doi.org/10.4028/www.scientific.net/JERA.15.71.
[11] T. Samana, T. Kiatsiriroat, A. Nuntaphan, Enhancement of fins efficiency of a solid wire fin by oscillating heat pipe under forced convection, Case Studies Thermal Engg 2 (2014) 36–41, https://doi.org/10.1016/j.csite.2013.10.003.
[12] W. Li, L. Yang, J. Ren, H. Jiang, Effect of thermal boundary conditions and thermal conductivity on conjugate heat transfer performance in pin fin arrays, Int. J. Heat Mass Tran. 95 (2016) 579–592, https://doi.org/10.1016/j.ijheatmasstransfer.2015.12.010.
[13] M. Ramezanizadeh, M.A. Nazari, Modeling thermal conductivity of Ag/water nanofluid by applying a mathematical correlation and artificial neural network, Int. J. Low Carbon Technol. 14 (2019) 468–474, https://doi.org/10.1093/ijlct/ctz030.
[14] A. Maleki, A. Haghighi, M.I. Shahrestani, Z. Abdelmalek, Applying different type of artificial neural network for modelling thermal conductivity of nanofluids containing silica particles, J. Therm. Anal. Calorim. (2020), https://doi.org/10.1007/s10973-020-09541-x.
[15] P. Harrington, Machine Learning in Action, Manning Publications Co., New York, USA, 2012.
[16] N.S. Altman, An introduction to kernel and nearest-neighbor nonparametric regression, Am. Statistician 46 (1992) 175–185, https://doi.org/10.1080/00031305.1992.10475879.

View publication stats

arXiv:2007.06069v1 [math.CV] 12 Jul 2020

ON CERTAIN GENERALIZATIONS OF $\mathcal{S}^*(\psi)$

S. SIVAPRASAD KUMAR AND KAMALJEET GANGANIA

Abstract. We deal with different kinds of generalizations of $\mathcal{S}^*(\psi)$, the class of Ma-Minda starlike functions, in addition to a majorization result of $\mathcal{C}(\psi)$, the class of Ma-Minda convex functions, which are enlisted as follows: 1. Let h be an analytic function, f be in $\mathcal{C}(\psi)$ and h be majorized by f in the unit disk $\mathbb{D}$, then for a given ψ, we derive a general equation, which yields the radius constant r_ψ such that $|h'(z)| \leq |f'(z)|$ in $|z| \leq r_\psi$. Consequently, obtain results associating $\mathcal{S}^*(\psi)$ and others. 2. We find the largest radius r_0 so that the product function $g(z)h(z)/z$ belongs to a desired class for $|z| < r_0$ whenever $g \in \mathcal{S}^*(\psi_1)$ and $h \in \mathcal{S}^*(\psi_2)$. Also we obtain a condition for the functions to be in $\mathcal{S}^*(\psi)$ 3. We obtain the modified distortion theorem for $\mathcal{S}^*(\psi)$ with a general perspective. 4. For a fixed $f \in \mathcal{S}^*(\psi)$, the class of subordinants $S_f(\psi) := \{g : g \prec f\}$ is introduced and studied for the Bohr-phenomenon and a couple of conjectures are also proposed.

2010 AMS Subject Classification. Primary 30C80, Secondary 30C45.
Keywords and Phrases. Subordination, Bohr-Radius, Majorization, Distortion theorem.

1. Introduction

Let $\mathcal{A}$ denote the class of analytic functions of the form $f(z) = z + \sum_{k=2}^{\infty} a_k z^k$ in the open unit disk $\mathbb{D} := \{z : |z| < 1\}$. Using subordination, Ma-Minda [13] introduced the unified class of starlike and convex functions defined as follows:

$$\mathcal{S}^*(\psi) := \left\{ f \in \mathcal{A} : \frac{zf'(z)}{f(z)} \prec \psi(z) \right\} \text{ and } \mathcal{C}(\psi) := \left\{ f \in \mathcal{A} : 1 + \frac{zf''(z)}{f'(z)} \prec \psi(z) \right\}, \tag{1.1}$$

where ψ is a Ma-Minda function, which is analytic and univalent with $\Re\psi(z) > 0$, $\psi'(0) > 0$, $\psi(0) = 1$ and $\psi(\mathbb{D})$ is symmetric about real axis. Note that $\psi \in \mathcal{P}$, the class of normalized Carathéodory functions. The class $\mathcal{S}^*(\psi)$ unifies various subclasses of starlike functions, which are obtained for an appropriate choice of ψ. Ma-Minda discussed many properties of the class $\mathcal{S}^*(\psi)$, in particular, they proved the distortion theorem [13, Theorem 2, p.162] with some restriction on ψ, namely

$$\min_{|z|=r} |\psi(z)| = \psi(-r) \quad \text{and} \quad \max_{|z|=r} |\psi(z)| = \psi(r). \tag{1.2}$$

In section 4, we modify the distortion theorem by relaxing this restriction on ψ to obtain a more general result. In 1914, Harald Bohr [5] proved the following remarkable result related to the power series:

Theorem 1.1 ([5]). *Let $g(z) = \sum_{k=0}^{\infty} a_k z^k$ be an analytic function in $\mathbb{D}$ and $|g(z)| < 1$ for all $z \in \mathbb{D}$, then $\sum_{k=0}^{\infty} |a_k| r^k \leq 1$ for all $z \in \mathbb{D}$ with $|z| = r \leq 1/3$.*

The work of the second author is supported by University Grant Commission, New-Delhi, India under UGC-Ref. No.:1051/(CSIR-UGC NET JUNE 2017).

Bohr actually proved the above result for $r \leq 1/6$. Further Wiener, Riesz and Shur independently sharpened the result for $r \leq 1/3$. Presently the Bohr inequality for functions mapping unit disk onto different domains, other than unit disk is an active area of research. For the recent development on Bohr-phenomenon, see the articles [1, 2, 3, 4, 10, 11, 17, 18, 19, 20, 22] and references therein. The concept of Bohr-phenomenon in terms of subordination can be described as: Let $f(z) = \sum_{k=0}^{\infty} a_k z^k$ and $g(z) = \sum_{k=0}^{\infty} b_k z^k$ are analytic in $\mathbb{D}$ and $f(\mathbb{D}) = \Omega$. For a fixed f, consider a class of analytic functions $S(f) := \{g : g \prec f\}$ or equivalently $S(\Omega) := \{g : g(z) \in \Omega\}$. Then the class $S(f)$ is said to satisfy Bohr-phenomenon, if there exists a constant $r_0 \in (0, 1]$ satisfying the inequality $\sum_{k=1}^{\infty} |b_k| r^k \leq d(f(0), \partial\Omega)$ for all $|z| = r \leq r_0$ and $g(z) \in S(f)$, where $d(f(0), \partial\Omega)$ denotes the Euclidean distance between $f(0)$ and the boundary of $\Omega = f(\mathbb{D})$. The largest such r_0 for which the inequality holds, is called the Bohr-radius.

In 2014, Muhanna et al. [19] proved the Bohr-phenomenon for $S(W_\alpha)$, where $W_\alpha := \{w \in \mathbb{C} : |\arg w| < \alpha\pi/2, 1 \leq \alpha \leq 2\}$, which is a Concave-wedge domain (or exterior of a compact convex set) and the class $R(\alpha, \beta, h)$ defined by $R(\alpha, \beta, h) := \{g \in \mathcal{A} : g(z) + \alpha z g'(z) + \beta z^2 g''(z) \prec h(z)\}$, where h is a convex function (or starlike) and $R(\alpha, \beta, h) \subset S(h)$. In 2018, Bhowmik and Das [3] proved the Bohr-phenomenon for the classes given by $S(f) = \{g \in \mathcal{A} : g \prec f \text{ and } f \in \mu(\lambda)\}$, where $\mu(\lambda) = \{f \in \mathcal{A} : |(z/f(z))^2 f'(z) - 1| < \lambda, 0 < \lambda \leq 1\}$ and $S(f) = \{g \in \mathcal{A} : g \prec f \text{ and } f \in \mathcal{S}^*(\alpha), 0 \leq \alpha \leq 1/2\}$, where $\mathcal{S}^*(\alpha)$ is the well-known class of starlike functions of order α. In Section 5, for any fixed $f \in \mathcal{S}^*(\psi)$, we introduce and study the Bohr-phenomenon inside the disk $|z| \leq 1/3$ for the following class of analytic subordinants:

$$S_f(\psi) := \left\{ g(z) = \sum_{k=1}^{\infty} b_k z^k : g \prec f \right\}. \tag{1.3}$$

Note that $\mathcal{S}^*(\psi) \subset \bigcup_{f \in \mathcal{S}^*(\psi)} S_f(\psi)$. As an application, we obtain the Bohr-radius for the class $S(f)$, where $f \in \mathcal{S}^*((1 + Dz)/(1 + Ez))$, the class of Janowski starlike functions, with some additional restriction on D and E apart from $-1 \leq E < D \leq 1$. Now recall the following definition and a result due to T. H. MacGregor [14]:

Definition 1.2. ([14]) Let f and g be analytic in $\mathbb{D}$. A function $g(z)$ is said to be majorized by $f(z)$, denoted by $g << f$, if there exists an analytic function $\Phi(z)$ in $\mathbb{D}$ satisfying $|\Phi(z)| \leq 1$ and $g(z) = \Phi(z) f(z)$ for all $z \in \mathbb{D}$.

Theorem 1.3 ([14])**.** *Let g be majorized by f in $\mathbb{D}$ and $g(0) = 0$. If $f(z)$ is univalent in $\mathbb{D}$, then $|g'(z)| \leq |f'(z)|$ in $|z| \leq 2 - \sqrt{3}$. The constant $2 - \sqrt{3}$ is sharp.*

Recently Tang and Deng [25] obtained the majorization results for $\mathcal{S}^*(\psi)$ for some specific choices of ψ, motivated by this in section 2, we devise a general approach to handle the same for $\mathcal{C}(\psi)$, which is precisely stated as: if $g \in \mathcal{A}$, $f \in \mathcal{C}(\psi)$ and g is majorized by f in $\mathbb{D}$, then we find the largest radius $r_\psi \leq 1$ such that $|g'(z)| \leq |f'(z)|$ in $|z| \leq r_\psi$. Several other results in this direction are also obtained. In section 3, we consider the radius problem posed by Obradović and Ponnusamy [21] namely: Let $g \in \mathcal{S}^*(\psi_1)$ and $h \in \mathcal{S}^*(\psi_2)$, then find the largest radius $r_0 \leq 1$ such that the function $F(z) = (g(z)h(z))/z$ belongs to certain well-known class of starlike functions in $|z| < r_0$. As a special case, we also obtain a result of Obradović and Ponnusamy [21]. Further we obtain the condition for functions to be in $\mathcal{S}^*(\psi)$ which is an extention of the Bulboacă and Tuneski [6]. Throughout this paper we shall assume that the function ϕ has real coefficients in it's power series expansion.

2. Majorization

Let us consider the analytic function $\psi(z) := 1+B_1z+B_2z^2+\cdots$. Here $B_1 = \psi'(0)$, the coefficient of z, plays a major role in deciding the orientation of the function ψ. Thus ψ is positively or negatively oriented depends on whether B_1 is positive or negative. Ma-Minda only considered the case $\psi'(0) > 0$, as it may be possible that for the case when $\psi'(0) < 0$, many postulates for the class $\mathcal{S}^*(\psi)$ need not remain same. With this perspective, we begin with the following:

Theorem 2.1. *Let $\Re\phi(z) > 0$ and ϕ be convex in $\mathbb{D}$ with $\phi(0) = 1$. Suppose ψ be the function such that $m_r := \min\limits_{|z|=r}|\psi(z)|$ and also satisfies the differential equation*

$$\psi(z) + \frac{z\psi'(z)}{\psi(z)} = \phi(z). \tag{2.1}$$

Let $g \in \mathcal{A}$ and $f \in \mathcal{C}(\phi)$. If g is majorized by f in $\mathbb{D}$, then

$$|g'(z)| \leq |f'(z)| \quad \text{in } |z| \leq r_\psi, \tag{2.2}$$

where r_ψ is the least positive root of the equation

$$(1-r^2)m_r - 2r = 0. \tag{2.3}$$

Proof. Let us define $p(z) := zf'(z)/f(z)$. Since $f \in \mathcal{C}(\phi)$, therefore we have $1 + zf''(z)/f'(z) \prec \phi(z)$, which can be equivalently written as

$$p(z) + \frac{zp'(z)}{p(z)} = 1 + \frac{zf''(z)}{f'(z)} \prec \phi(z). \tag{2.4}$$

Since $\Re\phi(z) > 0$ and ϕ is convex in $\mathbb{D}$, therefore the solution ψ of the differential equation (2.1) is analytic in $\mathbb{D}$ with $\Re\psi(z) > 0$ and has the following integral form given by

$$\psi(z) := h(z)\left(\int_0^z \frac{h(t)}{t}dt\right)^{-1},$$

where

$$h(z) = z\exp\int_0^z \frac{\phi(t)-1}{t}dt.$$

Since $\Re\psi(z) > 0$ and p satisfies the subordination (2.4), therefore we conclude that ψ is univalent and $p \prec \psi$, where ψ is the best dominant. Thus we have obtained that $f \in \mathcal{C}(\phi)$ implies $zf'(z)/f(z) \prec \psi(z)$ and ψ is the best dominant, which is a univalent Caratheódory function. Now as $g \in \mathcal{A}$ and $f \in \mathcal{C}(\phi)$, therefore we obtain the following well defined equality

$$\frac{f(z)}{f'(z)} = \frac{z}{\psi(\omega(z))}, \quad (z \in \mathbb{D})$$

where ω is a Schwarz function. Hence, using $\min_{|z|=r}|\psi(\omega(z))| \geq \min_{|z|=r}|\psi(z)|$ and the hypothesis $\min\limits_{|z|=r}|\psi(z)| = m_r$, we obtain that

$$\left|\frac{f(z)}{f'(z)}\right| \leq \frac{r}{m_r}, \quad (0 < r < 1). \tag{2.5}$$

Now if g is majorized by f, then by definition, we have $g(z) = \Psi(z)f(z)$, where Ψ is analytic and satisfies $|\Psi(z)| \leq 1$ in $\mathbb{D}$ such that $g'(z) = \Psi(z)f'(z) + \Psi'(z)f(z)$.

Thus using (2.5) together with the following Schwarz-Pick inequality

$$|\Psi'(z)| \leq \frac{1-|\Psi(z)|^2}{1-|z|^2},$$

we obtain

$$|g'(z)| \leq |f'(z)| \left(|\Psi(z)| + \frac{1-|\Psi(z)|^2}{1-r^2} \frac{r}{m_r} \right) = |f'(z)| h(\beta, r), \tag{2.6}$$

where $|\Psi(z)| := \beta$ and

$$h(\beta, r) = \beta + \frac{1-\beta^2}{1-r^2} \frac{r}{m_r}.$$

Thus to arrive at (2.2), it suffices to show that $h(\beta, r) \leq 1$, which is equivalent to show that

$$k(\beta, r) := (1-r^2) m_r - (\beta+1) r \geq 0. \tag{2.7}$$

Since $\frac{\partial}{\partial \beta} k(\beta, r) = -r < 0$, Therefore, (2.7) holds whenever

$$k(r) := \min_{\beta} k(\beta, r) = k(1, r) \geq 0.$$

Note that $k(r)$ is a continuous function of r and further $k(0) = m_0 = \psi(0) = 1 > 0$ and $k(1) < 0$. Thus there exists a point $r_\psi \in (0,1)$ such that $k(r) \geq 0$ for all $r \in [0, r_\psi]$, where r_ψ is the least positive root of (2.3). □

The following result was proved by MacGregor [14]:

Remark 2.1. Let $g \in \mathcal{A}$ and $f \in \mathcal{C}$. If g is majorized by f in $\mathbb{D}$, then $|g'(z)| \leq |f'(z)|$ in $|z| \leq 1/3$. The result is sharp.

In our next result, we show the application to the Janowski class, which covers many well-known classes:

Corollary 2.2. *Let f belongs to $\mathcal{C}[D, E]$, where $-1 \leq E < D \leq 1$ along with $1 + D/E \geq 0$ and $-1 \leq E < 0$. If g is majorized by f, then*

$$|g'(z)| \leq |f'(z)| \quad \textit{in} \quad |z| \leq r_0,$$

where r_0 is the smallest positive root of the equation

$$(1-r^2) \left({}_2F_1 \left(1 - \frac{D}{E}, 1, 2; \frac{-Er}{1-Er} \right) \right)^{-1} - 2r = 0.$$

Proof. In Theorem 2.1, put $\phi(z) = (1+Dz)/(1+Ez)$. Then we have $\psi(z) := 1/q(z)$, where

$$q(z) = \begin{cases} \int_0^1 \left(\frac{1+Etz}{1+Ez} \right)^{\frac{D-E}{E}} dt, & \text{if } E \neq 0; \\ \\ \int_0^1 e^{D(t-1)z} dt, & \text{if } E = 0, \end{cases}$$

which further can be represented in terms of confluent and Gaussian hypergeometric functions, respectively as follows:

$$q(z) = \begin{cases} {}_2F_1 \left(1 - \frac{D}{E}, 1, 2; \frac{Ez}{1+Ez} \right), & \text{if } E \neq 0; \\ {}_1F_1 \left(1, 2; -Dz \right), & \text{if } E = 0. \end{cases}$$

Since $1 + D/E \geq 0$ and $-1 \leq E < 0$, therefore we have

$$\min_{|z|=r} \Re \psi(z) = \psi(-r) = \frac{1}{q(-r)} = \left({}_2F_1 \left(1 - \frac{D}{E}, 1, 2; \frac{-Er}{1-Er} \right) \right)^{-1}.$$

Since $\Re\psi(z) > 0$ and $\min_{|z|=r} \Re\psi(z) = \psi(-r)$, therefore we conclude that $\min_{|z|=r} |\psi(z)| = \psi(-r)$ and hence, the result follows from Theorem 2.1. □

Now we have the result for the class of convex functions of order α using Corollary 2.2:

Corollary 2.3. *Let f belongs to $\mathcal{C}[1-2\alpha, -1]$, where $0 \leq \alpha < 1$. If g is majorized by f, then $|g'(z)| \leq |f'(z)|$ in $|z| \leq r_0$, where r_0 is the smallest positive root of the equation*

$$(1-r^2)\left({}_2F_1\left(2(1-\alpha), 1, 2; \frac{r}{1+r}\right)\right)^{-1} - 2r = 0.$$

Corollary 2.4. *Let f belongs to $\mathcal{C}[D, 0]$. If g is majorized by f, then $|g'(z)| \leq |f'(z)|$ in $|z| \leq r_0$, where r_0 is the smallest positive root of the equation*

$$(1-r^2)(Dre^{-Dr}/(e^{-Dr}-1)) + 2r = 0.$$

Proof. From the proof of Corollary 2.2, we obtain that $\psi(z) = Dze^{Dz}/(e^{Dz}-1)$, when $\phi(z) = 1 + Dz$. Now with a little computation, we find that the function $l(z) = ze^z/(e^z-1)$ is convex univalent in $\mathbb{D}$. Therefore, the function $\psi(z) = l(Dz)$ is also convex in $\mathbb{D}$ for each fixed $0 < D \leq 1$. Since ψ is also symmetric about the real axis, we conclude that $\min_{|z|=r} |\psi(z)| = \psi(-r)$. Hence the result. □

Theorem 2.5. *Let ϕ be convex in $\mathbb{D}$, with $\Re\phi(z) > 0$, $\phi(0) = 1$ and suppose $f \in \mathcal{A}$ satisfies the differential subordination*

$$\frac{zf'(z)}{f(z)} + z\left(\frac{zf'(z)}{f(z)}\right)' \prec \phi(z). \tag{2.8}$$

If g is majorized by f, then $|g'(z)| \leq |f'(z)|$ in $|z| \leq r_0$, where r_0 is the least positive root of the equation

$$(1-r^2)\min_{|z|=r} \Re\psi(z) - 2r = 0,$$

where

$$\psi(z) := \frac{1}{z}\int_0^z \phi(t)dt.$$

Proof. Let $p(z) = zf'(z)/f(z)$. Then the subordination (2.8) can equivalently be written as:

$$p(z) + zp'(z) \prec \phi(z).$$

A simple calculation show that the analytic function $\psi(z) := (1/z)\int_0^z \phi(t)dt$ satisfies

$$\psi(z) + z\psi'(z) = \phi(z).$$

Now from the Hallenbeck and Ruscheweyh result [16, Theorem 3.1b, p. 71], we have $p \prec \psi$, where ψ is the best dominant and also convex. Further, since $\Re\phi(z) > 0$, using the integral operator [16, Theorem 4.2a, p. 202] preserving functions with positive real part, we see that ψ is a Caratheódory function. Thus we have

$$\frac{f(z)}{zf'(z)} \prec \frac{1}{\psi(z)} \quad \text{which implies} \quad \left|\frac{f(z)}{f'(z)}\right| \leq \frac{r}{\min_{|z|=r}|\psi(z)|} = \frac{r}{\min_{|z|=r}\Re\psi(z)}.$$

Now proceeding same as in the Theorem 2.1 result follows. □

Corollary 2.6. *Suppose $f \in \mathcal{A}$ satisfies the differential subordination*

$$\frac{zf'(z)}{f(z)} + z\left(\frac{zf'(z)}{f(z)}\right)' \prec \frac{1+z}{1-z}.$$

If g is majorized by f, then $|g'(z)| \leq |f'(z)|$ in $|z| \leq r_0$, where r_0 is the least positive root of the equation

$$(1-r^2)(2\log(1+r) - r) - 2r^2 = 0.$$

Corollary 2.7. *Suppose $f \in \mathcal{A}$ satisfies the differential subordination*

$$\frac{zf'(z)}{f(z)} + z\left(\frac{zf'(z)}{f(z)}\right)' \prec e^z.$$

If g is majorized by f, then $|g'(z)| \leq |f'(z)|$ in $|z| \leq r_0$, where r_0 is the least positive root of the equation

$$(1-r^2)(1-e^{-r}) - 2r^2 = 0.$$

Theorem 2.8. *Let ϕ be convex in $\mathbb{D}$, with $\Re\phi(z) > 0$, $\phi(0) = 1$ and suppose $f \in \mathcal{A}$ satisfies the differential subordination*

$$\frac{zf'(z)}{f(z)}\left(\frac{zf'(z)}{f(z)} + 2z\left(\frac{zf'(z)}{f(z)}\right)'\right) \prec \phi(z), \quad \alpha \in [0,1). \tag{2.9}$$

If g is majorized by f, then $|g'(z)| \leq |f'(z)|$ in $|z| \leq r_0$, where r_0 is the least positive root of the equation

$$(1-r^2)\min_{|z|=r}|\sqrt{\psi(z)}| - 2r = 0,$$

where

$$\psi(z) := \frac{1}{z}\int_0^z \phi(t)dt.$$

Proof. Let $p(z) = zf'(z)/f(z)$. Then the subordination (2.9) can be equivalently written as:

$$p^2(z) + 2zp(z)p'(z) \prec \phi(z),$$

which using the change of variable $P(z) = p^2(z)$ becomes

$$P(z) + zP'(z) \prec \phi(z).$$

Now proceeding as in Theorem 2.5, we see that $p(z) \prec \sqrt{\psi(z)}$ and $\sqrt{\psi(z)}$ is the best dominant. Further, since $\Re\phi(z) > 0$, using [16, Theorem 4.2a, p. 202], we see that ψ is a Caratheódory function. Therefore,

$$|\arg\sqrt{\psi(z)}| = \frac{1}{2}|\arg\psi(z)| \leq \frac{\pi}{4},$$

which implies $\Re\sqrt{\psi(z)} > 0$. Thus we have

$$\frac{f(z)}{zf'(z)} \prec \frac{1}{\sqrt{\psi(z)}} \quad \text{which implies} \quad \left|\frac{f(z)}{f'(z)}\right| \leq \frac{r}{\min_{|z|=r}|\sqrt{\psi(z)}|}.$$

Now proceeding same as in the Theorem 2.1 result follows. □

Corollary 2.9. *Suppose $f \in \mathcal{A}$ satisfies the differential subordination*

$$\frac{zf'(z)}{f(z)}\left(\frac{zf'(z)}{f(z)}+2z\left(\frac{zf'(z)}{f(z)}\right)'\right) \prec \frac{1+(2\alpha-1)z}{1+z}.$$

If g is majorized by f, then $|g'(z)| \leq |f'(z)|$ in $|z| \leq r_0$, where r_0 is the least positive root of the equation

$$(1-r^2)\min_{|z|=r} \Re\sqrt{\psi(z)} - 2r = 0,$$

where

$$\psi(z) := \frac{1}{z}\int_0^z \frac{1+(2\alpha-1)t}{1+t}dt.$$

Corollary 2.10. *Suppose $f \in \mathcal{A}$ satisfies the differential subordination*

$$\frac{zf'(z)}{f(z)}\left(\frac{zf'(z)}{f(z)}+2z\left(\frac{zf'(z)}{f(z)}\right)'\right) \prec 1+\alpha z, \quad (\alpha \in (0,1]).$$

If g is majorized by f, then $|g'(z)| \leq |f'(z)|$ in $|z| \leq r_0$, where r_0 is the least positive root of the equation

$$(1-r^2)\sqrt{1-\beta r} - 2r = 0, \quad \text{where} \quad \beta = \alpha/2.$$

Now we state the following result without proof as it follows from Theorem 2.1:

Theorem 2.11. *Let $\psi \in \mathcal{P}$ be a univalent function such that*

$$m_r := \min_{|z|=r}|\psi(z)| = \begin{cases} \psi(-r), & \text{if } \psi'(0) > 0; \\ \psi(r), & \text{if } \psi'(0) < 0. \end{cases}$$

Let $g \in \mathcal{A}$ and $f \in \mathcal{S}^*(\psi)$. If g is majorized by f in $\mathbb{D}$, then

$$|g'(z)| \leq |f'(z)| \quad \text{in } |z| \leq r_\psi,$$

where r_ψ is the least positive root of the equation

$$(1-r^2)m_r - 2r = 0.$$

Example 2.12. *Let us consider the analytic functions $\psi_1(z) = \sqrt{1-z}$ and $\psi_2(z) = \sqrt{1+z}$. Note that $\psi_1'(0) < 0$, $\psi_2'(0) > 0$ and for $|z| = r$,*

$$m_{r_1} = \min_{|z|=r}|\psi_1(z)| = \psi_1(r) = \sqrt{1-r} = \psi_2(-r) = \min_{|z|=r}|\psi_2(z)| = m_{r_2}.$$

Now from Theorem 2.11, we obtain the following result:
If $g \in \mathcal{A}$, $f \in \mathcal{S}^*(\psi_i)$, where $i = 1, 2$ and g is majorized by f, then $|g'(z)| \leq |f'(z)|$ in $|z| \leq r_0$, where r_0 is the least positive root of the equation

$$(1-r^2)\sqrt{1-r} - 2r = 0.$$

Interestingly, the desired radius in both the cases remain same as $\psi_1(\mathbb{D}) = \psi_2(\mathbb{D})$, though ψ_1 and ψ_2 are oppositely oriented.

Remark 2.2. Taking $\alpha = 0$ or $\eta = 1$ in corollary 2.13, case (ii) and (iii), respectively, we obtain the result proved by T. H. MacGregor [14], namely: *Let $g \in \mathcal{A}$ and $f \in \mathcal{S}^*$. If $g << f$ in $\mathbb{D}$, then $|g'(z)| \leq |f'(z)|$ in $|z| \leq 2-\sqrt{3}$.* The result is sharp.

Now we obtain the following majorization results for some known classes as well those introduced and studied in [7, 8, 15, 23].

Corollary 2.13. *Let $g \in \mathcal{A}$ and $f \in \mathcal{S}^*(\psi)$. If $g << f$ in $\mathbb{D}$, then $|g'(z)| \leq |f'(z)|$ in $|z| \leq r_\psi$, where r_ψ is the least positive root of the equation $P(r) = 0$ and the result follows for each one of the following cases:*

(*i*) $P(r) = (1-r^2)((1-Dr)/(1-Er)) - 2r$ *when* $\psi(z) = \frac{1+Dz}{1+Ez}$, *where* $-1 \leq E < D \leq 1$.
(*ii*) $P(r) = (1-r)(1-(1-2\alpha)r) - 2r$ *when* $\psi(z) = \frac{1+(1-2\alpha)z}{1+z}$, *where* $0 \leq \alpha < 1$.
(*iii*) $P(r) = (1-r^2)((1-r)/(1+r))^\eta - 2r$ *when* $\psi(z) = \left(\frac{1+z}{1-z}\right)^\eta$, *where* $0 < \eta \leq 1$.
(*iv*) $P(r) = (1-r^2)\left(\sqrt{2} - (\sqrt{2}-1)\sqrt{\frac{1+r}{1-2(\sqrt{2}-1)r}}\right) - 2r$ *when* $\psi(z) = \sqrt{2} - (\sqrt{2}-1)\sqrt{\frac{1-z}{1+2(\sqrt{2}-1)z}}$.
(*v*) $P(r) = (1-r^2)(b(1-r))^{1/a} - 2r$ *when* $\psi(z) = (b(1+z))^{1/a}$, *where* $a \geq 1$ *and* $b \geq 1/2$.
(*vi*) $P(r) = (1-r^2) - 2re^r$ *when* $\psi(z) = e^z$.
(*vii*) $P(r) = (1-r^2)(\sqrt{1+r^2} - r) - 2r$ *when* $\psi(z) = z + \sqrt{1+z^2}$.
(*viii*) $P(r) = (1-r^2) - r(1+e^r)$ *when* $\psi(z) = \frac{2}{1+e^{-z}}$.
(*ix*) $P(r) = (1-r^2)(1-\sin r) - 2r$ *when* $\psi(z) = 1 + \sin z$.

Remark 2.3. In Corollary 2.13, case (*ix*), we obtained the radius $r_\psi \approx 0.312478$ which improves the majorization-radius $r_s \approx 0.309757$ obtained in [26].

Let $\psi(z) = 1+z/(1-\alpha z^2)$, $0 \leq \alpha < 1$, introduced and studied by Kargar et al. [9]. Clearly $\psi \in \mathcal{P}$ only when $\alpha = 0$ and hence Theorem 2.11 holds when $\psi(z) = 1+z$. Moreover, for some $r > 0$, the quantity $z/\psi(z)$ does not exist for all $|z| = r$. In view of the same, the result proved by Tang and Deng [25], needs correction and the corrected version is stated in the following corollaries:

Corollary 2.14. *Let $g \in \mathcal{A}$ and $f \in \mathcal{S}^*(1+\beta z)$, $0 < \beta \leq 1$. If $g << f$ in $\mathbb{D}$, then*

$$|g'(z)| \leq |f'(z)| \quad in \quad |z| \leq r_\beta,$$

where r_β is the least positive root of the equation

$$(1-r^2)(1-\beta r) - 2r = 0.$$

Now we obtain the result related to $\mathcal{BS}(\alpha)$, the class of Booth lemniscate starlike functions, when $\alpha \neq 0$.

Corollary 2.15. *Let $0 < \alpha < 1$ and r_α be the unique root of the equation*

$$\alpha r^2 + r - 1 = 0. \tag{2.10}$$

Let $g \in \mathcal{A}$ and $g << f$ in $\mathbb{D}$, where $f \in \mathcal{BS}(\alpha)$. Then

$$|g'(z)| \leq |f'(z)| \quad in \quad |z| \leq r_{B(\alpha)} := \min\{r_\alpha, r_0\},$$

where r_0 is the least positive root of the equation

$$(1-r^2)\left(1 - \frac{r}{1-\alpha r^2}\right) - 2r = 0.$$

Proof. Observe that $\Re\left(1 + \frac{z}{1-\alpha z^2}\right) > 0$ for $|z| < r_\alpha$, where r_α is the unique root of (2.10). Thus the inequality in (2.5) holds for $|z| = r < r_\alpha$ and the result follows at once. □

3. **Product of starlike functions and a sufficient condition**

Assume that ψ_1 and ψ_2 belong to $\mathcal{P}$ and satisfy the following conditions for $|z| = r$ and $i = 1, 2$

$$\max_{|z|=r} \Re\psi_i(z) = \psi_i(r) \quad \text{and} \quad \min_{|z|=r} \Re\psi_i(z) = \psi_i(-r). \tag{3.1}$$

Motivated by Obradović and Ponnusamy [21], in this section, we consider the radius problem to generalize their result and also establish a similar result for the Urlagaddi class $\mathcal{M}(\beta) := \{f \in \mathcal{A} : \Re(zf'(z)/f(z)) < \beta,\ \beta > 1\}$. Also we extend a result of the Bulboacă and Tuneski [6] for the class $\mathcal{S}^*(\psi)$.

Theorem 3.1. *Let $g \in \mathcal{S}^*(\psi_1)$ and $h \in \mathcal{S}^*(\psi_2)$, where ψ_i satisfy the first condition in (3.4). Then the function F defined by*

$$F(z) = \frac{g(z)h(z)}{z} \tag{3.2}$$

belongs to $\mathcal{M}(\beta)$ in the disk $|z| < r_\beta = \min\{1, r_0(\beta)\}$, where $r_0(\beta)$ is the least positive root of the equation

$$\psi_1(r) + \psi_2(r) - 1 - \beta = 0. \tag{3.3}$$

The radius r_β is sharp.

Proof. Let $g \in \mathcal{S}^*(\psi_1)$ and $h \in \mathcal{S}^*(\psi_2)$. Then in view of (3.4) and subordination principle, it follows that

$$\Re\frac{zg'(z)}{g(z)} \le \psi_1(r) \quad \text{and} \quad \Re\frac{zh'(z)}{h(z)} \le \psi_2(r)$$

in $|z| \le r < 1$. Since

$$\frac{zF'(z)}{F(z)} = \frac{zg'(z)}{g(z)} + \frac{zh'(z)}{h(z)} - 1,$$

we have for $|z| = r$,

$$\Re\frac{zF'(z)}{F(z)} \le \psi_1(r) + \psi_2(r) - 1 \le \beta,$$

whenever $r \le \min\{1, r_0(\beta)\}$, where $r_0(\beta)$ is the least positive root of the equation (3.3). The sharpness follows by considering the functions

$$g(z) = z\exp\int_0^z \frac{\psi_1(t) - 1}{t}dt \quad \text{and} \quad h(z) = z\exp\int_0^z \frac{\psi_2(t) - 1}{t}dt.$$

Corollary 3.2. *Let $g \in \mathcal{S}^*(\gamma)$ and $h \in \mathcal{S}^*(\tau)$. Then the function F defined in (3.2) belongs to $\mathcal{M}(\beta)$ in the disk $|z| < \min\{1, r_0(\beta)\}$, where*

$$r_0(\beta) = \frac{\beta - 1}{3 + \beta - 2(\gamma + \tau)}.$$

The proof of the following result is much akin to Theorem 3.1, so omitted here.

Theorem 3.3. *Let $g \in \mathcal{S}^*(\psi_1)$ and $h \in \mathcal{S}^*(\psi_2)$, where ψ_i satisfy the second condition in (3.4). Then the function F defined in (3.2) is starlike of order γ in the disk $|z| < r_\gamma$, where r_γ is the least positive root of the equation*

$$\psi_1(-r) + \psi_2(-r) - 1 - \gamma = 0.$$

The radius r_γ is sharp.

We obtain the following result proved by Obradović and Ponnusamy [21]:

Remark 3.1. Let $g \in \mathcal{S}^*(\gamma)$ and $h \in \mathcal{S}^*(\tau)$. Then the function F defined in (3.2) is starlike of order γ_0 in the disk

$$|z| < \frac{1-\gamma_0}{\gamma_0 + 3 - 2(\gamma+\tau)}.$$

Remark 3.2. Note that the identity function $z \in \mathcal{S}^*(\psi)$. Thus if we choose $g(z) = z$ (or $h(z) = z$) in (3.2), then the problem reduces to obtaining the $\mathcal{M}(\beta)$-radius (or $\mathcal{S}^*(\gamma)$-radius) of the class $\mathcal{S}^*(\psi_2)$ (or $\mathcal{S}^*(\psi_1)$). It is also evident that the conditions given in (3.4) establish the inclusion relations $\mathcal{S}^*(\psi) \subseteq \mathcal{M}(\psi(1))$ and $\mathcal{S}^*(\psi) \subseteq \mathcal{S}^*(\psi(-1))$, respectively.

In the following, we extend the results of the Bulboacă and Tuneski [6]:

Theorem 3.4. *Let h be analytic with $h(0) = 0$, $h'(0) \neq 0$. Suppose that h satisfies*

$$\Re\left(1 + \frac{zh''(z)}{h(z)}\right) \geq -\frac{1}{2}$$

and

$$\frac{1}{z}\int_0^z h(t)dt \prec \frac{\psi(z)-1}{\psi(z)}. \tag{3.4}$$

If $f \in \mathcal{A}$, then

$$\frac{f(z)f''(z)}{(f'(z))^2} \prec h(z) \quad \textit{implies} \quad f \in \mathcal{S}^*(\psi).$$

Proof. Using the result [6, Theorem 3.1, p. 3], we see that $f(z)f''(z)/(f'(z))^2 \prec h(z)$ implies

$$\frac{1}{z}\int_0^z \left(1 - \left(\frac{f(t)}{f'(t)}\right)'\right) dt = 1 - \frac{f(z)}{zf'(z)} \prec \frac{1}{z}\int_0^z h(t)dt.$$

From the above subordination, we have

$$\frac{f(z)}{zf'(z)} \prec 1 - \frac{1}{z}\int_0^z h(t)dt.$$

Now to prove that $f \in \mathcal{S}^*(\psi)$, it suffices to consider

$$1 - \frac{1}{z}\int_0^z h(t)dt \prec \frac{1}{\psi(z)},$$

which is equivalent to (3.4). This completes the proof. □

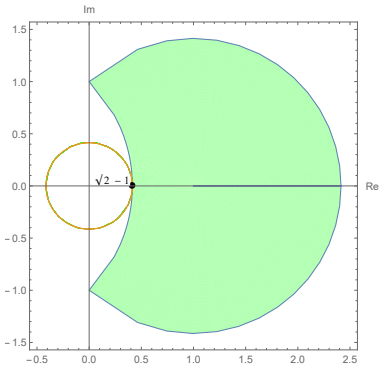
4. **Distortion theorem**

Ma-Minda [13] proved the distortion theorem for the class $\mathcal{S}^*(\psi)$ with some restriction on ψ, namely $|\psi(z)|$ attains its maximum and minimum value respectively at $z = r$ and $z = -r$, see eq. (1.2). Now what if ψ does not satisfy the condition (1.2) and why the condition (1.2) is so important? To answer this, we first need to recall the following result:

Lemma 4.1. ([13]) *Let $f \in \mathcal{S}^*(\psi)$ and $|z_0| = r < 1$. Then $-f_0(-r) \leq |f(z_0)| \leq f_0(r)$. Equality holds for some $z_0 \neq 0$ if and only if f is a rotation of f_0, where $zf_0(z)/f_0(z) = \psi(z)$ such that*

$$f_0(z) = z\exp\int_0^z \frac{\psi(t)-1}{t}dt. \tag{4.1}$$

We see that a Ma-Minda starlike function, in general, need not satisfy the condition (1.2). To examine the same, let us consider two different Ma-Minda starlike functions, namely $\psi_1(z) := z + \sqrt{1+z^2}$ and $\psi_2(z) := 1 + ze^z$. The unit disk images of ψ_1 and ψ_2 are displayed in figure 1 and figure 2.



Legend for figure 3 -

Let $\min_{|z|=r}|\psi_2(z)| =: \gamma_i(r)$, where
$z = r_i e^{i\theta}$, $0 \le \theta \le \pi$,
then from table 1, we have
$\gamma_1(1) = 0.372412$,
$\gamma_2(4/5) = 0.527912$,
$\gamma_3(2/3) = 0.611553$,
$\gamma_4(1/2) = 0.693287$,
$\gamma_5(r) = 1 - re^{-r}$, where $r \le (3-\sqrt{5})/2$.

FIGURE 1. $\psi_1(z) := z + \sqrt{1+z^2}$

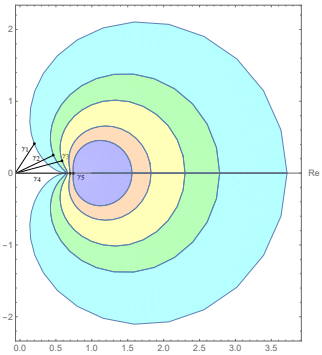


FIGURE 2. $\psi_2(z) := 1 + ze^z$

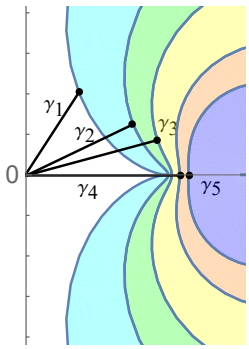


FIGURE 3. A zoom of figure 2

We know that the radius of a circle, centered at origin and touching only the boundary points of a image domain of a complex function, yields the optimal values of the modulus of the function. For example, see figure 1 to locate the lower bound of the modulus for a crescent function. Therefore it is evident from figure 1 that both the bounds $\psi_1(-r)$ and $\psi_1(r)$ of $|\psi_1|$ are attained on the real line and we have $\psi_1(-r) \le |\psi_1(z)| \le \psi_1(r)$ for each $|z| = r$. Whereas, from figure 2, we see

that only the upper bound $\psi_2(r)$ of $|\psi_2|$ is attained on the real line and $|\psi_2(z)| \leq \psi_2(r)$ for each $|z| = r$. Although both ψ_1 and ψ_2 are Ma-Minda functions but the distortion theorem of Ma-Minda [13, theorem 2, p. 162] does not accommodate the function ψ_2, as the lower bound for $|\psi_2(z)|$ is not attained on the real line for all $|z| = r > (3 - \sqrt{5})/2$, see figure 3. To overcome this limitation, we modify the distortion theorem, wherein we theoretically assume the modulus bounds of the function and obtain a more general result. Thus the Ma-Minda functions, for which modulus bounds are not attained on the real line but could be estimated, can now be entertained for distortion theorem using the following result:

Theorem 4.1 (Modified Distortion Theorem)**.** *Let ψ be a Ma-Minda function. Assume that $\min_{|z|=r}|\psi(z)| = |\psi(z_1)|$ and $\max_{|z|=r}|\psi(z)| = |\psi(z_2)|$, where $z_1 = re^{i\theta_1}$ and $z_2 = re^{i\theta_2}$ for some $\theta_1, \theta_2 \in [0, \pi]$. Let $f \in \mathcal{S}^*(\psi)$ and $|z_0| = r < 1$. Then*

$$|\psi(z_1)| \left(\frac{-f_0(-r)}{r}\right) \leq |f'(z_0)| \leq \left(\frac{f_0(r)}{r}\right) |\psi(z_2)|. \tag{4.2}$$

Proof. Let $p(z) = zf'(z)/f(z)$. Then $f \in \mathcal{S}^*(\psi)$ if and only if $p(z) \prec \psi(z)$. Using a result [13, Theorem 1, p.161], we have

$$\frac{f(z)}{z} \prec \frac{f_0(z)}{z}, \tag{4.3}$$

where f_0 is given by (4.1). Now using Maximum-Minimum principle of modulus, (4.3) and by Lemma 4.1, $-f_0(-r)/r \leq |f(z_0)/z| \leq f_0(r)/r$, we easily obtain for $|z_0| = r$

$$\begin{aligned}
|\psi(z_1)| \left(\frac{-f_0(-r)}{r}\right) &= \min_{|z|=r} |\psi(z)| \min_{|z|=r} \left|\frac{f_0(z)}{z}\right| \\
&\leq \left|p(z_0)\frac{f(z_0)}{z_0}\right| \\
&= |f'(z_0)| \leq \max_{|z|=r} |\psi(z)| \max_{|z|=r} \left|\frac{f_0(z)}{z}\right| \\
&= \left(\frac{f_0(r)}{r}\right) |\psi(z_2)|,
\end{aligned}$$

that is,

$$|\psi(z_1)| \left(\frac{-f_0(-r)}{r}\right) \leq |f'(z_0)| \leq \left(\frac{f_0(r)}{r}\right) |\psi(z_2)|,$$

where z_1 and z_2 are as defined in the hypothesis. Hence the result. □

To illustrate Theorem 4.1, we consider the function $\psi(z) = 1 + ze^z$. Then we have the following expression for its modulus:

$$|\psi(z)| = \sqrt{1 + re^{r\cos\theta}(re^{r\cos\theta} + 2\cos(\theta + r\sin\theta))}. \tag{4.4}$$

Using equation (4.4) and Theorem 4.1, we obtain the following table 1, providing the minimum for various choices of r.

Now using Theorem 4.1, we obtain the following distortion theorem for the class $\mathcal{S}^*(1 + ze^z)$:

TABLE 1. The lower bounds for $|1+ze^z|$ for different choices of $r=|z|$.

r	$0\leq\theta_1\leq\pi$	$\lvert\psi(re^{i\theta_1})\rvert$	$m(r,\theta_1)=\lvert\psi(re^{i\theta_1})\rvert(-f_0(-r)/r)$
1	1.88438	0.372412	0.197923
4/5	2.01859	0.527912	0.304374
2/3	2.17677	0.611553	0.375966
1/2	2.58169	0.693287	0.467769
$r\leq(3-\sqrt{5})/2$	π	$\psi_2(-r)$	$f_0'(-r)$

Example 4.2. *Let $\psi(z)=1+ze^z$ and $\min\limits_{|z|=r}|\psi(z)|=|\psi(z_1)|$, where $z_1=re^{i\theta_1}$ for some $\theta_1\in[0,\pi]$. Let $f\in\mathcal{S}^*(\psi)$ and $|z_0|=r<1$. Then*

$$m(r,\theta_1)\leq|f'(z_0)|\leq f_0'(r),\quad\left(r>\tfrac{3-\sqrt{5}}{2}\right)$$

and

$$f_0'(-r)\leq|f'(z_0)|\leq f_0'(r),\quad\left(r\leq\tfrac{3-\sqrt{5}}{2}\right),$$

where $f_0(z)=z\exp(e^z-1)$ and $m(r,\theta_1)$ is provided in table 1 for some specific values of r.

Remark 4.1. In Theorem 4.1, if we assume that $\theta_1=\pi$ and $\theta_2=0$, then extremes in equation (4.2) simplifies to $f_0'(-r)$ and $f_0'(r)$, respectively since $zf_0'(z)/f_0(z)=\psi(z)$. Thus the extremes in the equation (4.2) are in terms of r alone and also lead to the sharp bounds. Consequently, we obtain the following distortion theorem of Ma-Minda [13] as a special case of Theorem 4.1:

Let $\min_{|z|=r}|\psi(z)|=\psi(-r)$ and $\max_{|z|=r}|\psi(z)|=\psi(r)$. If $f\in\mathcal{S}^*(\psi)$ and $|z_0|=r<1$. Then

$$f_0'(-r)=\psi(-r)\frac{f_0(-r)}{-r}\leq|f'(z_0)|\leq\frac{f_0(r)}{r}\psi(r)=f_0'(r).$$

Equality holds for some $z_0\neq0$ if and only if f is a rotation of f_0.

5. Bohr-Phenomenon for functions in $S_f(\psi)$

Note that "the Bohr radius of the class $\mathcal{X}$ is at least r_x", this holds for every result in this section. In general, Bohr radius is estimated for a specific class provided the sharp coefficients bounds of the functions in that class are known. For instance, consider the class of starlike univalent functions, where we have the sharp coefficient bounds: $|a_n|\leq n$. However, for most of the Ma-Minda subclasses, the better coefficients bounds are yet not known. Hence, we encounter the following problem, especially in context of Ma-minda classes, which we deal here to a certain extent:
Problem: *If coefficients bounds are not known, how one can find a good lower estimate for the Bohr radius of a given class?*

To readily understand the above problem, consider the class $\mathcal{S}^*(1+ze^z)$, where the sharp coefficients bounds for functions in this class are unknown. In this situation, how one can find the Bohr radius for this class or is there any way out with the lower bounds all alone? Here below we state Theorem 5.1, where we find a solution for this problem. Note that the Bohr radius $3-2\sqrt{2}\approx0.1713$ for the class $\mathcal{S}^*$ serves as a lower bound for the class $S_f(\psi)$ and is also a special case of Theorem 5.1.

Let $\mathbb{B}(0,r) := \{z \in \mathbb{C} : |z| < r\}$, $g(z) = \sum_{k=1}^{\infty} b_k z^k$, $\mathcal{S}^*(\psi)$ and $S_f(\psi)$ as defined in (1.1) and (1.3) respectively. For any $g \in S_f(\psi)$, we find the radius r_b so that $S_f(\psi)$ obey the following Bohr-phenomenon:

$$\sum_{k=1}^{\infty} |b_k| r^k \leq d(f(0), \partial\Omega) \quad \text{for} \quad |z| = r \leq r_b, \tag{5.1}$$

where $d(f(0), \partial\Omega)$ denotes the Euclidean distance between $f(0)$ and the boundary of $\Omega = f(\mathbb{D})$. Now we prove our main result:

Theorem 5.1. *Let r_* be the Koebe-radius for the class $\mathcal{S}^*(\psi)$, $f_0(z)$ be given by the equation (4.1) and $g(z) = \sum_{k=1}^{\infty} b_k z^k \in S_f(\psi)$. Assume $f_0(z) = z + \sum_{n=2}^{\infty} t_n z^n$ and $\hat{f}_0(r) = r + \sum_{n=2}^{\infty} |t_n| r^n$. Then $S_f(\psi)$ satisfies the Bohr-phenomenon*

$$\sum_{k=1}^{\infty} |b_k| r^k \leq d(f(0), \partial\Omega), \quad \textit{for } |z| = r \leq r_b, \tag{5.2}$$

where $r_b = \min\{r_0, 1/3\}$, $\Omega = f(\mathbb{D})$ and r_0 is the least positive root of the equation

$$\hat{f}_0(r) = r_*.$$

Proof. Since $g \in S_f(\psi)$, we have $g \prec f$ for a fixed $f \in \mathcal{S}^*(\psi)$. By letting r tends to 1 in Lemma 4.1, we obtain the Koebe-radius $r_* = -f_0(-1)$. Therefore $\mathbb{B}(0, r_*) \subset f(\mathbb{D})$, which implies $r_* \leq d(0, \partial\Omega) = |f(z)|$ for $|z| = 1$. Also using [13, Theorem 1, p.161], we have

$$\frac{f(z)}{z} \prec \frac{f_0(z)}{z}. \tag{5.3}$$

Recall the result [3, Lemma 1, p.1090], which reads as: let f and g be analytic in $\mathbb{D}$ with $g \prec f$, where $f(z) = \sum_{n=0}^{\infty} a_n z^n$ and $g(z) = \sum_{k=0}^{\infty} b_k z^k$. Then $\sum_{k=0}^{\infty} |b_k| r^k \leq \sum_{n=0}^{\infty} |a_n| r^n$ for $|z| = r \leq 1/3$. Now using the result for $g \prec f$ and (5.3), we have

$$\sum_{k=1}^{\infty} |b_k| r^k \leq r + \sum_{n=2}^{\infty} |a_n| r^n \leq \hat{f}_0(r) \quad \text{for } |z| = r \leq 1/3.$$

Finally, to establish the inequality (5.2), it is enough to show $\hat{f}_0(r) \leq r_*$. But this holds whenever $r \leq r_0$, where r_0 is the least positive root of the equation $\hat{f}_0(r) = r_*$. The existence of the root r_0 is ensured by the relations $\hat{f}_0(1) \geq |f_0(1)| \geq r_*$ and $\hat{f}_0(0) < r_*$. Thus, if $r_b = \min\{r_0, 1/3\}$ then $\sum_{k=1}^{\infty} |b_k| r^k \leq d(0, \partial\Omega)$ holds. Hence the result. □

Remark 5.1. Let us further assume that the coefficients B_n of ψ are positive. Then the function $f_0(z) = z + \sum_{n=2}^{\infty} t_n z^n$ defined by integral representation (4.1) can be written as

$$f_0(z) = z \exp\left(\sum_{n=1}^{\infty} \frac{B_n}{n} z^n \right),$$

which implies $f_0(r) = \hat{f}_0(r)$ for $|z| = r$.

From the proof of Theorem 5.1, we have the following:

Theorem 5.2. *Let r_* be the Koebe-radius for the class $\mathcal{S}^*(\psi)$, $f_0(z)$ be given by the equation (4.1) and $f(z) = z + \sum_{n=2}^{\infty} a_n z^n \in \mathcal{S}^*(\psi)$. Assume $f_0(z) = z + \sum_{n=2}^{\infty} t_n z^n$ and $\hat{f}_0(r) = r + \sum_{n=2}^{\infty} |t_n| r^n$. Then $\mathcal{S}^*(\psi)$ satisfies the Bohr-phenomenon*

$$r + \sum_{n=2}^{\infty} |a_n| r^n \leq d(f(0), \partial\Omega), \quad \text{for } |z| = r \leq r_b,$$

where $r_b = \min\{r_0, 1/3\}$, $\Omega = f(\mathbb{D})$ and r_0 is the least positive root of the equation

$$\hat{f}_0(r) = r_*.$$

Some Applications:

(a). If we choose $\psi(z) = (1 + Dz)/(1 + Ez)$, $-1 \leq E < D \leq 1$, then $\mathcal{S}^*(\psi)$ denotes the class of Janowski starlike functions and we have

$$r_* = \begin{cases} (1-E)^{\frac{D-E}{E}}, & E \neq 0; \\ e^{-D}, & E = 0. \end{cases} \tag{5.4}$$

and

$$f_0(z) = \begin{cases} z(1+Ez)^{\frac{D-E}{E}}, & E \neq 0; \\ z\exp(Dz), & E = 0. \end{cases} \tag{5.5}$$

Observe that if $E \neq 0$, the n-th ($n \geq 2$) coefficients of $f_0(z)$ is given by

$$t_n = \prod_{k=2}^{\infty} \frac{D - (k-1)E}{(k-1)!}. \tag{5.6}$$

Thus from Theorem 5.1, we have the following result:

Corollary 5.3. *Let $\psi(z) = (1 + Dz)/(1 + Ez)$, $-1 \leq E < D \leq 1$. Then $S_f(\psi)$ (and $\mathcal{S}^*(\psi)$) satisfies the Bohr-phenomenon (5.1) for $|z| = r \leq r_b$, where $r_b = \min\{r_0, 1/3\}$ and r_0 is the least positive root of the equation*

$$r + \sum_{n=2}^{\infty} |t_n| r^n - (1-E)^{\frac{D-E}{E}} = 0,$$

where t_n is as defined in (5.6).

Note that for the Janowski class, sharp coefficients bounds in general are not known. Now as an application of Corollary 5.3, we obtain the following result when $t_n > 0$:

Corollary 5.4. (Bohr-radius with Janowski class) *Let $\psi(z) = (1 + Dz)/(1 + Ez)$, $-1 \leq E < D \leq 1$.*

(*i*) *If $E = 0$ and $D \geq \frac{3}{4}\log 3$. Then $S_f(\psi)$ (and $\mathcal{S}^*(\psi)$) satisfies the Bohr-phenomenon (5.1) for $|z| = r \leq r_0$, where r_0 is the only real root of the equation*

$$1 - re^{D(1+r)} = 0. \tag{5.7}$$

(*ii*) *If $E \neq 0$ and further satisfies*

$$3(1-E)^{\frac{D-E}{E}} \leq (1 + E/3)^{\frac{D-E}{E}}. \tag{5.8}$$

Then $S_f(\psi)$ (and $\mathcal{S}^*(\psi)$) satisfies the Bohr-phenomenon (5.1) for $|z| = r \leq r_0$, where r_0 is the only real root of the equation

$$(1-E)^{\frac{D-E}{E}} - r(1+Er)^{\frac{D-E}{E}} = 0. \tag{5.9}$$

The result is sharp for the function f_0 defined in (5.5).

Proof. (i): Since $E = 0$, we have $r_* = e^{-D}$. Moreover $\hat{f}_0(r) = f_0(r) = r\exp(Dr)$. Now we need to show

$$r\exp(Dr) \leq e^{-D} \tag{5.10}$$

or equivalently $T(r) := 1 - re^{D(1+r)} \geq 0$ holds for $r \leq r_0$. Which obviously holds for $\frac{3}{4}\log 3 \leq D \leq 1$. Since $d(f_0(0), \partial f_0(\mathbb{D})) = r_*$, therefore we see from inequality (5.10) that Bohr-radius is sharp for the function f_0 given by (5.5).

(ii): Proceeding as in case (i), it is sufficient to show the inequality

$$r(1+Er)^{\frac{D-E}{E}} \leq (1-E)^{\frac{D-E}{E}} \tag{5.11}$$

or equivalently $g(r) := (1-E)^{\frac{D-E}{E}} - r(1+Er)^{\frac{D-E}{E}} \geq 0$ holds for $r \leq r_0$. Which obviously follows whenever D and E satisfies (5.8). In view of the inequality (5.11), the sharp Bohr-radius is achieved for the function f_0 given by (5.5). □

Remark 5.2. **(Bohr-radius with starlike functions of order α)** Let $\psi(z) := (1+(1-2\alpha)z)/(1-z)$, where $0 \leq \alpha < 1$. We see $\mathcal{S}^*(\psi) := \mathcal{S}^*(\alpha)$ and for this class, we have

$$r_* = \frac{1}{2^{2(1-\alpha)}} \quad \text{and} \quad f_0(z) = \frac{z}{(1-z)^{2(1-\alpha)}}.$$

Observe that here $\hat{f}_0(r) = f_0(r)$. Now as an application of Corrollary 5.4, we obtain the result due to Bhowmik et al. [3], namely: *If* $0 \leq \alpha \leq 1/2$. *Then* $S_f(\psi)$ *satisfies the Bohr-phenomenon* $\sum_{k=1}^{\infty}|b_k|r^k \leq d(f(0), \partial\Omega)$, *for* $|z| = r \leq r_b$, *where* $r_b = \min\{r_0, 1/3\} = r_0$ *and* r_0 *is the only real root of the equation* $(1-r)^{2(1-\alpha)}/r = 2^{2(1-\alpha)}$. *The result is sharp.*

Now form the above remark, in particular, we have:

Corollary 5.5. *If* $0 \leq \alpha \leq 1/2$. *Then the class* $\mathcal{S}^*((1+(1-2\alpha)z)/(1-z))$ *satisfies the Bohr-phenomenon* (5.1) *for* $|z| = r \leq r_0$, *where* r_0 *is the only real root of the equation*

$$(1-r)^{2(1-\alpha)}/r = 2^{2(1-\alpha)}.$$

The result is sharp. In particular, the Bohr radius for the class $\mathcal{S}^*$ *is* $3 - 2\sqrt{2} \approx 0.1713$.

(b). If we choose $\psi(z) = \sqrt{1+z}$, then $\mathcal{S}^*(\psi) := \mathcal{SL}^*$, the class of lemniscate starlike functions and for this class we have:

$$f_0(z) = \frac{4z\exp(2\sqrt{1+z}-2)}{(1+\sqrt{1+z})^2} \quad \text{and} \quad r_* = -f_0(-1) \approx 0.541341. \tag{5.12}$$

Also in this case $\hat{f}_0(r) = f_0(r)$ and therefore, we obtain the following corollary:

Corollary 5.6. *The class* $S_f(\psi)$ *(and* $\mathcal{SL}^*$*), where* $\psi(z) = \sqrt{1+z}$ *satisfies the Bohr-phenomenon* (5.1) *for* $|z| = r \leq 1/3$.

(c). If we consider $\psi(z) = 1 + ze^z$, then $\mathcal{S}^*(\psi) := \mathcal{S}^*_{\wp}$, the class of cardioid starlike functions introduced in [12] and for this class, we have:

$$f_0(z) = z\exp(e^z - 1) \quad \text{and} \quad r_* = -f_0(-1) \approx 0.531464. \tag{5.13}$$

Here we can also see that $\hat{f}_0(r) = f_0(r)$ and we obtain the following corollary:

Corollary 5.7. *The class $S_f(\psi)$ (and $\mathcal{S}^*_\wp$), where $\psi(z) = 1 + ze^z$ satisfies the Bohr-phenomenon* (5.1) *for $|z| = r \leq 1/3$.*

Ali et al. [2] also showed that the coefficient bound of a function in a class have a role in the estimation of the Bohr-radius. Observed that for each $f \in \mathcal{S}^*(\psi)$, the class $S_f(\psi)$ satisfies the Bohr-phenomenon for $r \leq \min(1/3, r_0)$, where r_0 is the least positive root of $\hat{f}_0(r) - r_* = 0$. Since $\mathcal{S}^*(\psi) \subset \bigcup_{f\in\mathcal{S}^*(\psi)} S_f(\psi)$, therefore the Bohr-radius for the class $\mathcal{S}^*(\psi)$ is $r \geq \min(1/3, r_0)$. In Corollary 5.7, we find $r_0 \approx 0.349681$ (an upper bound for Bohr radius), which is almost close to $1/3 \approx 0.33333$ and is the unique root of $f_0(r) - r_* = 0$. Moreover, the bound for the coefficients of the functions belonging to $\mathcal{S}^*_\wp$ and $\mathcal{SL}^*$ have been conjectured [12, 24] with the extremals given in (5.13) and (5.12) respectively. Thus by using Theorem 5.1 and the approach dealt in [2] (assuming that conjectures are true), we propose the following conjectures:

Conjecture. The Bohr-radius for the class $\mathcal{S}^*_\wp$ is $r_0 \approx 0.349681$ which is the unique root in $(0, 1)$ of the equation

$$re^{e^r} = e^{1/e}.$$

Conjecture. The Bohr-radius for the class $\mathcal{SL}^*$ is $r_0 \approx 0.439229$, which the unique root in $(0, 1)$ of the equation

$$e^2 r \exp(2\sqrt{1+r} - 2) = (1 + \sqrt{1+r})^2.$$

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] Ali, R. M., Barnard, R. W., Solynin, A. Yu.: A note on the Bohrs phenomenon for power series. J. Math. Anal. Appl. **449**, 154–167 (2017)

[2] Ali, R. M., Jain, N. K., Ravichandran, V.: Bohr radius for classes of analytic functions. Results Math. **74**, Art. 179, 13 pp. (2019)

[3] Bhowmik, B., Das, N.: Bohr phenomenon for subordinating families of certain univalent functions. J. Math. Anal. Appl. **462**, 1087–1098 (2018)

[4] Boas, H. P., Khavinson, D.: Bohr's power series theorem in several variables. Proc. Amer. Math. Soc. **125**, 2975–2979 (1997)

[5] Bohr, H.: A theorem concerning power series. Proc. London Math. Soc. **13**, 1–5 (1914)

[6] Bulboacă, T., Tuneski, N.: New criteria for starlikeness and strongly starlikeness. Mathematica. **43**, 11–22 (2001)

[7] Cho, N. E., Kumar, V., Kumar, S. S., Ravichandran, V.: Radius problems for starlike functions associated with the sine function. Bull. Iranian Math. Soc. **45**, 213–232 (2019)

[8] Goel, P., Kumar, S. S.: Certain Class of Starlike Functions Associated with Modified Sigmoid Function. Bull. Malays. Math. Sci. Soc. (2019) https://doi.org/10.1007/s40840-019-00784-y.

[9] Kargar, R., Ebadian, A., Sokół, J.: On Booth lemniscate and starlike functions. Anal. Math. Phys. **9**(1), 143–154 (2019)

[10] Kayumov, I. R., Ponnusamy, S.: Bohr inequality for odd analytic functions. Comput. Methods Funct. Theory. **17**, 679–688 (2017)

[11] Kayumov, I. R., Ponnusamy, S.: Improved version of Bohr's inequality. C. R. Math. Acad. Sci. Paris. **356**, 272–277 (2018)

[12] Kumar, S. S., Kamaljeet, G.: A Cardioid Domain and Starlike Functions. Analysis and Mathematical Physics. (2020), (Communicated)

[13] Ma, W.C., Minda, D.: A unified treatment of some special classes of univalent functions. *Proceedings of the Conference on Complex Analysis (Tianjin, 1992)*. 157–169, Conf. Proc. Lecture Notes Anal., I Int. Press, Cambridge, MA.

[14] MacGregor, T. H.: Majorization by univalent functions. Duke Math. J. **34**, 95–102 (1967)

[15] Mendiratta, R., Nagpal, S., Ravichandran, V.: On a subclass of strongly starlike functions associated with exponential function. Bull. Malays. Math. Sci. Soc. **38**, 365–386 (2015)

[16] Miller S. S., Mocanu, P. T.: *Differential subordinations*, Monographs and Textbooks in Pure and Applied Mathematics. Marcel Dekker, Inc, New York (2000)

[17] Muhanna, Y. A.: Bohr's phenomenon in subordination and bounded harmonic classes. Complex Var. Elliptic Equ. **55**, 1071–1078 (2010)

[18] Muhanna, Y. A., Ali, R. M.: Bohr's phenomenon for analytic functions into the exterior of a compact convex body. J. Math. Anal. Appl. **379**, 512–517 (2011)

[19] Muhanna, Y. A., Ali, R. M., Ng, Z. C., Hansi, S. F. M.: Bohr radius for subordinating families of analytic functions and bounded harmonic mappings. J. Math. Anal. Appl. **420**, 124–136 (2014)

[20] Muhanna, Y. A., Ali R. M., Ponnusamy, S.: On the Bohr inequality. In Progress in Approximation Theory and Applicable Complex Analysis (Edited by N.K. Govil et al.), Springer Optimization and Its Applications. **117**, 265295 (2016)

[21] Obradović, M., Ponnusamy, S.: Product of univalent functions. Math. Comput. Modelling. **57**, 793–799 (2013)

[22] Paulsen, V. I., Popescu, G., Singh, D.: On Bohr's inequality. Proc. London Math. Soc. **85**, 493–512 (2002)

[23] Raina, R. K., Sokół, J.: Some properties related to a certain class of starlike functions. C. R. Math. Acad. Sci. Paris. **353**, 973–978 (2015)

[24] Sokół, J.: Coefficient estimates in a class of strongly starlike functions. Kyungpook Math. J. **49**, 349–353 (2009)

[25] Tang H., Deng, G.: Majorization problems for some subclasses of starlike functions. J. Math. Res. Appl. **39**, 153–159 (2019)

[26] Tang, H., Srivastava, H. M., Li, Sh., Deng, G. T.: Majorization results for subclasses of starlike functions based on the sine and cosine functions. Bull. Iran. Math. Soc. (2019). https://doi.org/10.1007/s41980-019-00262-y.

Department of Applied Mathematics, Delhi Technological University, Delhi–110042, India

E-mail address: spkumar@dce.ac.in

Department of Applied Mathematics, Delhi Technological University, Delhi–110042, India

E-mail address: gangania.m1991@gmail.com

Journal Pre-proofs



Regular paper

On the Outage Probability and BER of a DF-Based Multi-Hop System Over - - and - - -Extreme Fading Channels

Tau Raphael Rasethuntsa, Manpreet Kaur, Sandeep Kumar

PII: S1434-8411(20)30238-7
DOI: https://doi.org/10.1016/j.aeue.2020.153324
Reference: AEUE 153324

To appear in: *International Journal of Electronics and Communications*

Received Date: 30 January 2020
Accepted Date: 20 June 2020

Please cite this article as: T.R. Rasethuntsa, M. Kaur, S. Kumar, On the Outage Probability and BER of a DF-Based Multi-Hop System Over - - and - - -Extreme Fading Channels, *International Journal of Electronics and Communications* (2020), doi: https://doi.org/10.1016/j.aeue.2020.153324

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier GmbH.

Journal Pre-proofs

On the Outage Probability and BER of a DF-Based Multi-Hop System Over $\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme Fading Channels

Tau Raphael Rasethuntsa, Manpreet Kaur*, *Sandeep Kumar†

Abstract

This paper analyzes the performance of decode-and-forward (DF)-based multi-hop relaying communication systems over $\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme fading channels. Analytical formulas for the probability density function (PDF) and the cumulative distribution function (CDF) of the received signal-to-noise ratio (SNR) as well as its moment generating function (MGF) are derived. Based on the derived PDFs, novel closed-form expressions for the outage probability (OP), bit error rate (BER) under coherent and non-coherent modulation schemes are derived. Additionally, asymptotic analyses of BER based on Poincare series expansions of SNR PDFs are carried out and results show good approximations for low SNR regimes. Simulations are performed to verify the accuracy and to show the tightness of the theoretical analysis.

Index Terms

Multi-hop system, decode-and-forward relay, non-linear generalized fading, outage Probability, Bit error rate

Tau Raphael Rasethuntsa is based in, Lenasia Extension 9, Gauteng Province, Johannesburg, South Africa and Ha Abia/Mahlabatheng, Moreneng, Maseru, Lesotho, e-mail: tauzand@gmail.com

Manpreet Kaur is with Department of Computer Science & Engineering, Delhi Technological University, Delhi, India, e-mail: manpreettiet@gmail.com*

Corresponding Author: *Sandeep Kumar is with the Central Research Laboratory, Bharat Electronics Limited, Ghaziadbad, Uttar Pradesh, India, e-mail: sann.kaushik@gmail.com†

Journal Pre-proofs

Journal Pre-proofs

I. Introduction

The use of multi-hop communication systems as a way of increasing network coverage area and energy-efficiency with high data rates is gaining importance in next-generation wireless networks. In a multi-hop wireless communication system, information is transmitted from a source device to a destination device through several intermediate nodes, referred to as relays [1]. The relaying systems are classified in three categories 1) decode and forward (DF); 2) amplify and forward (AF); and 3) compress and forward (CF). In the DF systems, the relay device decodes the received data and then re-sends the decoded data to the next relay, therefore preventing the propagation of noise [2]. In AF, the received signal is amplified without decoding and forwarded to the next node. In CF, the received data is compressed, re-encoded and forwarded to the next-hop device [3].

The performance analysis of multi-hop communication systems over numerous fading distributions has been well documented in the literature [4], [5], [6], [7], [8], [9], [10]. The ergodic capacity of a multi-hop relaying system over Rayleigh fading channels has been studied in [4]. The end-to-end performance of a DF relaying system based on Ricean and Nakagami-m fading channels was explored in [5]. Yang et al. [6] studied the unified performance analysis of a DF-based multi-hop system based on second order statistics in average outage duration (AOD) and level crossing rate (LCR) over Nakagami, Rayleigh and Rician fading channels. The performance study of a multi-hop communication system with regenerative relays has been investigated over generalized $\eta-\mu$, $\alpha-\mu$ and $\kappa-\mu$ fading channels in [7], [8] and [9], respectively. Using traditional performance metrics as well as LCR and amount of fade duration (AFD), Cao et al. [10] presented the performance analysis of end-to-end wireless links over the Generalized-K fading channel.

In wireless communication, multi-path and shadowing highly deteriorate the strength of transmitting signals. A wide range of arithmetical distributions are available to characterize the fading effect in various signal propagation scenarios [11]. For instance, motivated by the pioneering works in [12] and [13] the generalized $\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme fading models were proposed in [14] and [15]. The $\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme are the generalized distributions which accounts for small-scale variations of mobile radio signals in line-of-sight (LOS) communications. The

$\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme models are very flexible models that can account well for non-linearity of multi-path propagation. The $\alpha-\kappa-\mu$ distribution is a popular model as it includes some realistic fading channels such as $\kappa-\mu$, Rayleigh, Nakagami-m, Weibull, and Rice as its special cases. The $\alpha-\kappa-\mu$-Extreme fading distribution is derived from the $\alpha-\kappa-\mu$ distribution by following the same methodology as in the derivation of the $\kappa-\mu$-Extreme model from the $\kappa-\mu$ distribution [15].

The $\alpha-\kappa-\mu$-Extreme fading distribution is used to characterize the mobile radio propagation under nonlinear severe multi-path fading effects usually encountered in enclosed environments such as airplanes, trains, and shopping malls etc. [16]. Furthermore, there are only a few studies that deal with the performance of wireless communications over $\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme fading channels. In [17] and [18] analytical expressions for OP, average error rate, and channel capacity of a single-input–single-output system were derived in the presence of $\alpha-\kappa-\mu$ fading. The second-order statistics of $\alpha-\kappa-\mu$ distribution were considered in [19] whereas [20] studied the spectrum sensing over $\alpha-\kappa-\mu$ fading channels. A comprehensive performance evaluation of cooperative spectrum sensing over $\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme fading channels has been presented in [16]. Recently in [21] and [22], energy detector performance over $\alpha-\kappa-\mu$ based composite fading channels were studied and the effect of all the fading parameters on the system performance was discussed.

In summary, the performance analysis of DF cooperative systems over many types of fading channels has been extensively studied, however, the performance of the DF over that $\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme fading channels in multi-hop systems has not been considered. Moreover, given the several practical application scenarios for the multi-hop wireless communication systems and the fact that$\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme fading model fits empirical data of wireless channels much better than previously studied fading models, it is important to investigate the different performance metrics of the multi-hop wireless systems operating over these channels. Henceforth, this paper investigates the outage probability (OP) and (bit error rate) BER performance of the DF multi-hop communication systems over the generalized fading model. The end-to-end performance of DF multi-hop systems is studied through numerical evaluation, which validate the correctness of our derived expressions compared to well-known fading distributions, and

Journal Pre-proofs

study the impacts of fading parameters and path length on the system performance.

The remainder of this paper is organized as follows. The system and channel model are described in Section II. In Section III, we provide the closed-form expressions for the statistical parameters over the channels. Different performance metrics and their approximations for the DF-based multi-hop wireless systems over $\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme fading channels has been presented in Section IV and V respectively. Performance evaluation, results, and discussion are provided in Section VI. Finally, we conclude this paper in Section VI.

II. Model Formulation and Description

A. System and Channel Models

Consider an n-hop system shown in Figure 1 transmitting information from the source S to the destination D via $n-1$ intermediate relays $R_1, R_2, \ldots, R_{n-1}$ in time-sharing principles. The relay link of the present model has a property of regenerative transmission and the channels of the n hops are assumed independent. As a result, if the signal-to-noise ratio (SNR) of any of the $n-1$ hops falls below a predetermined threshold γ_{th}, the system in regarded as failed [10], [8].

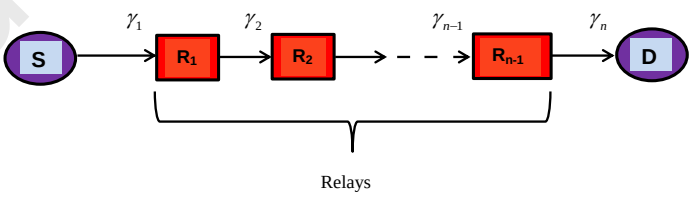


Fig. 1. A n-link DF-based multi-hop system.

Journal Pre-proofs

TABLE I
LIST OF SYMBOLS

Notation	Name of the Function	References
$\delta(\cdot)$	Dirac delta function	[23]
$Q(\cdot,\cdot)$	Generalized Marcum Q-function	[24, Eq. 2.23]
$\mathbb{E}(\cdot)$	Expectation operator	[16]
$\tilde{\Phi}(.;.;.)$	Regularized confluent hypergeometric function	[25, Eq. 9.210(1)]
$G(\cdot)$	Meijer's functions	[25, Eq. 9.301]
$H(\cdot)$	Fox-H function	[26, Eq. 2.57]
$I(\cdot)$	Modified Bessel functions of first kind	[27, Eq. 2.3.31)]
$\Gamma(.)$	Gamma function	[27, Eq. 2.3.22)]
$erfc(.)$	Complementary error function	[27, Eq. 2.3.17)]

If the SNR of the i-th hop is denoted by γ_i, with probability density function (PDF) f_{γ_i}, $i = 1, \dots n$, the statistical properties of the end-to-end system SNR are then determined by the SNR of the last hop γ_n. According to [23], the PDF of the end-to-end system SNR γ can be written as

$$p_\gamma(\gamma) = \mathcal{A}\delta(\gamma_n) + \bar{\mathcal{A}} p_{\gamma_n}(\gamma_n), \tag{1}$$

where $\mathcal{A}$ denotes the probability of the first outage occurring for the first $n-1$ hops and $\bar{\mathcal{A}} = 1 - \mathcal{A}$. The detail of all the symbols used in this work has been shown in Table I. The probability $\mathcal{A}$ is given by [23]

$$\begin{aligned} \mathcal{A} &= 1 - Pr\left(\gamma_1 > \gamma_{th}, \gamma_2 > \gamma_{th}, \cdots, \gamma_{m-1} > \gamma_{th}\right) \\ &= 1 - \prod_{k=1}^{m-1} Pr\left(\gamma_n > \gamma_{th}\right) = 1 - \prod_{k=1}^{m-1} \left[1 - P_{\gamma_n}(\gamma_{th})\right], \end{aligned} \tag{2}$$

where P_{γ_n} is the cumulative distribution function (CDF) of γ_n.

1) The $\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme fading models: The PDF of the instantaneous SNR of the n-th hop $\gamma_n > 0$ following the $\alpha_n - \kappa_n - \mu_n$ distribution is given by

$$p_{\gamma_n}(\gamma_n) = \mathcal{K}_n \gamma_n^{(\frac{\alpha_n}{2})\omega_n - 1} e^{-\sigma_n \gamma_n^{\frac{\alpha_n}{2}}} I_{\mu_n - 1}\left(\sqrt{\theta_n \gamma_n^{\frac{\alpha_n}{2}}}\right), \tag{3}$$

Journal Pre-proofs

where, in order to improve clarity in presentation, we have set

$$\mathcal{K}_n = \frac{\alpha_n \sigma_n^{\omega_n} e^{-\mu_n \kappa_n}}{2(\mu_n \kappa_n)^{\omega_n - 1}},\ \omega_n = \frac{\mu_n + 1}{2},\ \sigma_n = \frac{\mu_n(\kappa_n + 1)}{\bar{\gamma}_n^{\frac{\alpha_n}{2}}},$$
$$\theta_n = 4\mu_n \kappa_n \sigma_n \tag{4}$$

with $\alpha_n, \mu_n, \kappa_n > 0$. The average SNR per symbol $\bar{\gamma}_n$ is given by

$$\bar{\gamma}_n = \frac{e^{\mu_n \kappa_n} \left[\mu_n(\kappa_n + 1)\right]^{\frac{2}{\alpha_n}}}{\Gamma\left(\frac{2}{\alpha_n} + \mu_n\right) \tilde{\Phi}\left(\frac{2}{\alpha_n} + \mu_n; \mu_n; \mu_n \kappa_n\right)} \frac{E_b}{N_0}, \tag{5}$$

where E_b denotes the energy per symbol and N_0 is the single-sided power spectral density. Using the method of [28], one can easily show that the CDF of γ_n is written as

$$P_{\gamma_n}(\gamma_n) = 1 - Q_{\mu_n}\left(\sqrt{2\kappa_n \mu_n}, \sqrt{2\sigma_n \gamma_n^{\frac{\alpha_n}{2}}}\right),\ \gamma_n > 0, \tag{6}$$

The PDF of the instantaneous SNR of the n-th hop $\gamma_n > 0$ following the $\alpha_n - \kappa_n - \mu_n$-Extreme distribution is given by

$$p_{\gamma_n}^e(\gamma) = \frac{a_n \gamma_n^{\alpha_n/4 - 1}}{e^{b_n \gamma_n^{\frac{\alpha_n}{2}}}} I_1\left(\sqrt{c_n \gamma_n^{\frac{\alpha_n}{2}}}\right) + \frac{e^{-2m_n}}{2\sqrt{\gamma_n \bar{\gamma}_n}} \delta\left(\sqrt{\frac{\gamma_n}{\bar{\gamma}}}\right), \tag{7}$$

where

$$a_n = \frac{\alpha_n m_n e^{-2m_n}}{\bar{\gamma}_n^{\alpha_n/4}}, b = \frac{2m_n}{\bar{\gamma}_n^{\frac{\alpha_n}{2}}}, c_n = \frac{(16 m_n^2)}{\bar{\gamma}_n^{\frac{\alpha_n}{2}}}, \tag{8}$$

with $\alpha_n, \mu_n, \kappa_n > 0$ and $m_n = \mu_n(\kappa_n + 1)^2/(2\kappa_n + 1)$ as the Nakagami parameter which is inversely proportional to the fading severity. The average SNR per symbol is given by

$$\bar{\gamma}_n^e = \frac{e^{2m_n} (2m_n)^{\frac{2}{\alpha_n} - 1}}{\Gamma\left(\frac{2}{\alpha_n} + \mu_n\right) \tilde{\Phi}\left(\frac{2}{\alpha_n} + \mu_n; \mu_n; \mu_n \kappa_n\right)} \frac{E_b}{N_0}. \tag{9}$$

The CDF of the $\alpha - \kappa - \mu$-Extreme fading model can also be derived using the method of [28] resulting in

$$P_{\gamma_n}^e(\gamma) = 1 - Q_0\left(2\sqrt{m_n}, 2\sqrt{m_n(\gamma/\bar{\gamma})^{\frac{\alpha}{2}}}\right). \tag{10}$$

The symbols in (4) and (8) will be employed throughout the paper unless the original notation produces shorter expressions.

Journal Pre-proofs

Remark 1. It is also worth noting that the proposed PDFs representations renders, purely on a symbolic level, the $\alpha-\kappa-\mu$-Extreme fading model PDF without the extra $\delta(\cdot)$ term as a special case of the PDF of the $\alpha-\kappa-\mu$ model in a sense that $\mathcal{K}_n = a_n$, $\omega_n = 1/2$, $\sigma_n = b_n$, $\theta_n = c_n$ and $\mu_n = 2$. This observation will be particularly useful for reducing redundancy in the forthcoming derivations.

III. Statistical Properties

Two fading channels namely $\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme are considered in this study. For the convenience of readers, the analysis of $\alpha-\kappa-\mu$ channel is performed under **Case I** and that of $\alpha-\kappa-\mu$-Extreme channel under **Case II** in each section.

A. PDF and CDF of End-to-End SNR

Case I : Assume that the SNRs of the n hops are independent and not necessarily identically distributed (INID) random variables following the $\alpha-\kappa-\mu$ distribution. Then we have from (3) and (6) that the PDF and CDF of the end-to-end SNR γ must be given respectively by

$$p_\gamma(\gamma) = \mathcal{A}\delta(\gamma) + \mathcal{K}_n\bar{\mathcal{A}}\frac{\gamma^{(\frac{\alpha_n}{2})\omega_n-1}}{e^{\sigma_n\gamma^{\frac{\alpha_n}{2}}}} I_{\mu_n-1}\left(\sqrt{\theta_n\gamma^{\frac{\alpha_n}{2}}}\right) \tag{11}$$

where $\mathcal{A} = 1-\prod_{k=1}^{n-1} Q_{\mu_k}\left(\sqrt{2\kappa_k\mu_k}, \sqrt{2\sigma_k\gamma_{th}^{\frac{\alpha_k}{2}}}\right)$ and

$$P_\gamma(\gamma) = 1-\bar{\mathcal{A}}Q_{\mu_n}\left(\sqrt{2\kappa_n\mu_n}, \sqrt{2\sigma_n\gamma^{\frac{\alpha_n}{2}}}\right). \tag{12}$$

Case II : On the other hand if the SNRs of the n hops are INID random variables following the $\alpha-\kappa-\mu$-Extreme distribution, then we have from (7) and (10) that the PDF of the end-to-end SNR γ must be given by

$$p_\gamma^e(\gamma) = \mathcal{A}^e\delta(\gamma) + \bar{\mathcal{A}}^e\left[\frac{a_n\gamma^{\frac{\alpha_n}{4}-1}}{e^{b_n\gamma^{\frac{\alpha_n}{2}}}} I_1\left(\sqrt{c_n\gamma^{\frac{\alpha_n}{2}}}\right) + \delta^*(\gamma)\right] \tag{13}$$

Journal Pre-proofs

where $\delta^*(\gamma) = \dfrac{e^{-2m_n}}{2\sqrt{\gamma\bar{\gamma}_n}}\delta\left(\sqrt{\dfrac{\gamma_n}{\bar{\gamma}}}\right)$
and $\mathcal{A}^e = 1 - \prod_{k=1}^{n-1} Q_0\left(2\sqrt{m_k}, 2\sqrt{m_k(\gamma_k/\bar{\gamma}_k)^{\frac{\alpha_k}{2}}}\right)$. The CDF of γ can be written as

$$P_\gamma^e(\gamma) = 1 - \bar{\mathcal{A}}^e Q_0\left(2\sqrt{m_n}, 2\sqrt{m_n(\gamma/\bar{\gamma})^{\frac{\alpha_n}{2}}}\right). \tag{14}$$

B. Moment Generating Function

Another useful statistical characteristic of random variables is the moment generating function (MGF) used to obtain BER under non-coherent modulation, see Section IV. The MGF of γ, denoted by $\mathcal{M}_\gamma$, is defined as

$$\mathcal{M}_\gamma(s) \triangleq \int_0^\infty e^{-s\gamma} f_\gamma(\gamma) d\gamma. \tag{15}$$

For **Case I**, we substitute (11) into (15) to get the integral

$$\mathcal{M}_\gamma(s) = \mathcal{A} + \mathcal{K}_n \bar{\mathcal{A}} \overbrace{\int_0^\infty \frac{\gamma^{(\frac{\alpha_n}{2})\omega_n - 1}}{e^{s\gamma} e^{\sigma_n \gamma^{\frac{\alpha_n}{2}}}} I_{\mu_n - 1}\left(\sqrt{\theta_n \gamma^{\frac{\alpha_n}{2}}}\right) d\gamma}^{\mathcal{I}_1} \tag{16}$$

In order to evaluate $\mathcal{I}_1$, we set $y = \gamma^{\alpha_n/2}$ and employ [26, Eq. 1.25] and [29, Eq. 8.4.22(1)] as well as the relationship between the Meijer-G and the Fox-H functions as

$$\begin{aligned} \mathcal{I}_1 =& \pi \int_0^\infty y^{\omega_n - 1} e^{-\sigma_n y} H_{0,1}^{1,0}\left(s^{\frac{\alpha_n}{2}} y \,\middle|\, \begin{matrix} - \\ \left(0, \frac{\alpha_n}{2}\right) \end{matrix}\right) \\ & \times H_{1,3}^{1,0}\left(\frac{\theta_n}{4} y \,\middle|\, \begin{matrix} \left(\frac{\mu_n}{2}, 1\right) \\ (\omega_n - 1, 1), (1 - \omega_n, 1), \left(\frac{\mu_n}{2}, 1\right) \end{matrix}\right) dy \end{aligned} \tag{17}$$

Now, expanding the two Fox H-functions in terms of their definition as given in [30, Eq. 1.1.1] and reversing the order of integration, we have that

$$\begin{aligned} \mathcal{I}_1 =& \pi \frac{1}{(2\pi i)^2} \int_{\mathcal{L}_r} \int_{\mathcal{L}_t} \overbrace{\int_0^\infty y^{\omega_n + r + t - 1} e^{-\sigma_n y} dy}^{\mathcal{G}} \\ & \times \frac{\Gamma\left[-\left(\frac{\alpha_n}{2}\right) r\right] \Gamma(\omega_n - 1 - t) \left(s^{\frac{\alpha_n}{2}}\right)^r \left(\frac{\theta_n}{4}\right)^t dr dt}{\Gamma(\omega_n + t) \Gamma\left(1 - \frac{\mu_n}{2} + t\right) \Gamma\left(\frac{\mu_n}{2} - t\right)} \end{aligned} \tag{18}$$

Journal Pre-proofs

where $\mathcal{L}_r$ and $\mathcal{L}_t$ are two suitable contours in the complex r and t planes respectively. The integral $\mathcal{G}$ can be evaluated with the aid of [25, Eq. 3.326(2)] to yield

$$\begin{aligned}\mathcal{I}_1 =& \frac{\pi}{\sigma_n^{\omega_n}(2\pi i)^2}\int_{\mathcal{L}_r}\int_{\mathcal{L}_t}\frac{\Gamma\left[-\left(\frac{\alpha_n}{2}\right)r\right]\Gamma(\omega_n-1-t)}{\Gamma(\omega_n+t)} \\ &\times\frac{\Gamma\left(\omega_n+r+t\right)\left(s^{\frac{\alpha_n}{2}}/\sigma_n\right)^r\left(\frac{\theta_n}{4\sigma_n}\right)^t drdt}{\Gamma\left(1-\frac{\mu_n}{2}+t\right)\Gamma\left(\frac{\mu_n}{2}-t\right)}\end{aligned} \tag{19}$$

where $\Re(\omega_n+r+t)>0$. Now comparing (19) with the definition of the bivariate Fox H-function given in [26, Eq. 2.57] yields a closed-form expression for the MGF as shown in (22) of Table II. Remark 1 suggests that the MGF for **Case II** can be obtained from (22) by inspection resulting in the MGF for **Case II** being as shown in (23) of Table II. It is also worth pointing out that (15) can be written with the aid of integration by parts as

$$\mathcal{M}_\gamma(s) \triangleq s\int_0^\infty e^{-s\gamma}F_\gamma(\gamma)d\gamma. \tag{20}$$

Substituting (12) into (20), a more restricted alternative form for the MGF of γ for **Case I** valid for $2/\alpha_2, \mu_n \in \mathbb{N}$, can be obtained from [31, Theorem 2] for **Case I** as

$$\begin{aligned}\mathcal{M}_\gamma(s) =& 1-\frac{\alpha_n}{2}\bar{\mathcal{A}}+\frac{\bar{\mathcal{A}}\exp(-\kappa_n\mu_n)\left(\frac{2}{\alpha_n\sigma_n}\right)^{\frac{2}{\alpha_n}}s}{\left(\sqrt{\frac{2}{\alpha_n}}\right)(2\pi)^{\frac{1}{\alpha_n}-\frac{1}{2}}} \\ &\times\sum_{r=\mu_n}^{\infty}\sum_{n=0}^{\infty}\frac{(2\kappa_n\mu_n/\alpha_n)^n}{n!\Gamma(n+r+1)}\left(\frac{2}{\alpha_n}\right)^r \\ &\times G_{\frac{2}{\alpha_n},1}^{1,\frac{2}{\alpha_n}}\left[s\left(\frac{2}{\alpha_n\sigma_n}\right)^{\frac{2}{\alpha_n}}\middle|\begin{matrix}\Delta(\frac{2}{\alpha_n},1-n-r-\frac{2}{\alpha_n})\\ 0\end{matrix}\right].\end{aligned} \tag{21}$$

where $\Delta(k,\vartheta)=\frac{\vartheta}{k},\ldots,\frac{\vartheta+k-1}{k}$ comprises k elements. A similar expression can be obtained for **Case II**.

Journal Pre-proofs

TABLE II
MOMENT GENERATING FUNCTOINS

$$\mathcal{M}_{\gamma}(s) = \mathcal{A} + \bar{\mathcal{A}}\mathcal{K}_n \frac{\pi}{\sigma_n^{\omega_n}} H_{1,0:1,0;1,3}^{0,1:1,0;1,0} \left(\begin{array}{c|c} \frac{s^{\frac{\alpha_n}{2}}}{\sigma_n} & \mathbf{\Delta_1}\left(\alpha_n, \mu_n, \omega_n\right) \\ \frac{\theta_n}{4\sigma_n} & \mathbf{\Delta_2}\left(\alpha_n, \mu_n, \omega_n\right) \end{array} \right) \tag{22}$$

$$\mathcal{M}_{\gamma}^{e}(s) = \mathcal{A}^{e} + \bar{\mathcal{A}}^{e} \frac{a_n \pi}{\sqrt{b_n}} H_{1,0:1,0;1,3}^{0,1:1,0;1,0} \left(\begin{array}{c|c} \frac{s^{\frac{\alpha_n}{2}}}{b_n} & \mathbf{\Delta_1}\left(\alpha_n, 2, \frac{1}{2}\right) \\ \frac{c_n}{4b_n} & \mathbf{\Delta_2}\left(\alpha_n, 2, \frac{1}{2}\right) \end{array} \right). \tag{23}$$

$$\mathbf{\Delta_1} \equiv (1-\omega_n; 1, 1) : \; - \; ; \left(\frac{\mu_n}{2}, 1\right)$$

$$\mathbf{\Delta_2} \equiv - \; : \left(0, \frac{\alpha_n}{2}\right); \; (\omega_n - 1, 1), (1-\omega_n, 1), \left(\frac{\mu_n}{2}, 1\right)$$

IV. Performance Analysis of Multi-Hop DF-Based Relay Systems

A. Outage probability

The OP is defined as the probability that the SNR per symbol falls below a certain threshold γ_{th}. By employing (12) and (14), this probability can be written for **Case I** as

$$\begin{aligned} P_{out} &= P\left(\gamma \leq \gamma_{th}\right) \\ &= 1 - \bar{\mathcal{A}} Q_{\mu_n}\left(\sqrt{2\kappa_n \mu_n}, \sqrt{2\sigma_n \gamma_{th}^{\alpha_n/2}}\right). \end{aligned} \tag{24}$$

Similarly, for **Case II**, we have that OP can be written as

$$P_{out}^{e} = 1 - \bar{\mathcal{A}}^{e} Q_0\left(2\sqrt{m_n}, 2\sqrt{m_n(\gamma_{th}/\bar{\gamma})^{\alpha_n/2}}\right). \tag{25}$$

B. Bit Error Rate

1) Coherent Modulation Schemes: For a given SNR γ, the BER for a variety of coherent modulation schemes can be obtained by averaging the PDF f_{γ} over the conditional additive white Gaussian noise (AWGN) BER given as [32, Eq. 17]

$$P_b(e|\gamma) \triangleq (\phi/2)\text{erfc}\left(\rho\sqrt{\gamma/2}\right) \tag{26}$$

Journal Pre-proofs

where erfc$(\cdot)$ is the complementary error function [25, Eq. 8.250(4)]. The constants ϕ and ρ vary depending on the type of modulation scheme. For BFSK, $\phi = 1$ and $\rho = 1$. For BPSK, $\phi = 1$ and $\rho = \sqrt{2}$. For 4-QAM and QPSK, $\phi = 2$ and $\rho = 1$. Finally, for M-PAM, $\phi = 2(1 - 1/M)$ and $\rho = \sqrt{6/(M^2 - 1)}$. Thus, for **Case I** we average (11) over (26) to obtain the integral

$$P_b = \frac{\phi}{2} \int_0^\infty p_\gamma(\gamma) \text{erfc}\left(\rho\sqrt{\gamma/2}\right) d\gamma. \tag{27}$$

Substituting (11) into (27), one has that

$$P_b = \frac{\phi\mathcal{A}}{2} \int_0^\infty \delta(\gamma) \text{erfc}\left(\rho\sqrt{\gamma/2}\right) d\gamma + \frac{\phi\mathcal{K}_n\bar{\mathcal{A}}}{2}\mathcal{I}_2 \tag{28}$$

where

$$\mathcal{I}_2 = \int_0^\infty \frac{\gamma^{(\alpha_n/2)\omega_n - 1}}{e^{\sigma_n\gamma^{\alpha_n/2}}} I_{\mu_n - 1}\left(\sqrt{\theta_n\gamma^{\frac{\alpha_n}{2}}}\right) \text{erfc}\left(\rho\sqrt{\frac{\gamma}{2}}\right) d\gamma. \tag{29}$$

Following the same method as in Section III, using [29, Eq. 8.4.14(2)] to express the complementary error function in terms of its Mellin-Barnes representation, we obtain after some simple mathematical manipulations the BER under coherent modulation schemes for **Case I** and **Case II** as shown in (30) and (31) of Table III, respectively.

2) Non-Coherent Modulation Schemes: Under a range of different non-coherent modulation schemes, the BER of the end-to-end SNR γ can be written as [32, Eq. 27]

$$P_{bn} \triangleq \phi \int_0^\infty e^{-\rho\gamma} f_\gamma(\gamma) d\gamma. \tag{32}$$

The values of ϕ and ρ also vary depending on the modulation scheme. Specifically, for BFSK, $\phi = \frac{1}{2}$ and $\rho = \frac{1}{2}$. For DBPSK, $\phi = \frac{1}{2}$ and $\rho = 1$ and for M-FSK $\phi = (M - 1)/2$ and $\rho = \frac{1}{2}$. Upon substituting (11) into (32), it is clear that the resulting integral is of the form of the MGF integral given in (16). It is therefore straightforward to show that

$$P_{bn} = \phi\mathcal{M}_\gamma(\rho). \tag{33}$$

Journal Pre-proofs

TABLE III
BER UNDER COHERENT MODULATION

$$P_b = \frac{\phi}{2}\mathcal{A} + \bar{\mathcal{A}}\mathcal{K}_n \frac{\phi\sqrt{\pi}}{2}\left(\frac{(\sqrt{2})^{\alpha_n}}{\rho^{\alpha_n}}\right)^{\omega_n} \times H^{0,2:1,0;1,0}_{2,1:0,1;1,3}\left(\begin{array}{c|c} \sigma_n\dfrac{(\sqrt{2})^{\alpha_n}}{\rho^{\alpha_n}} & \boldsymbol{\nabla_1}(\alpha_n,\mu_n,\omega_n) \\ \theta_n\dfrac{(\sqrt{2})^{\alpha_n}}{4\rho^{\alpha_n}} & \boldsymbol{\nabla_2}(\alpha_n,\mu_n,\omega_n) \end{array}\right) \tag{30}$$

$$P_b^e = \frac{\phi}{2}\mathcal{A}^e + a_n\bar{\mathcal{A}}^e \frac{\phi\sqrt{\pi}}{2}\left(\frac{(\sqrt{2})^{\alpha_n}}{\rho^{\alpha_n}}\right)^{\frac{1}{2}} \times H^{0,2:1,0;1,0}_{2,1:0,1;1,3}\left(\begin{array}{c|c} b_n\dfrac{(\sqrt{2})^{\alpha_n}}{\rho^{\alpha_n}} & \boldsymbol{\nabla_1}\left(\alpha_n,2,\frac{1}{2}\right) \\ c_n\dfrac{(\sqrt{2})^{\alpha_n}}{4\rho^{\alpha_n}} & \boldsymbol{\nabla_2}\left(\alpha_n,2,\frac{1}{2}\right) \end{array}\right). \tag{31}$$

$$\boldsymbol{\nabla_1} \equiv \left(1-\frac{\alpha_n}{2}\omega_n;\frac{\alpha_n}{2},\frac{\alpha_n}{2}\right),\left(\frac{1}{2}-\frac{\alpha_n}{2}\omega_n;\frac{\alpha_n}{2},\frac{\alpha_n}{2}\right):-;\left(\frac{\mu_n}{2},1\right)$$

$$\boldsymbol{\nabla_2} \equiv \left(-\frac{\alpha_n}{2}\omega_n;\frac{\alpha_n}{2},\frac{\alpha_n}{2}\right):(0,1);\,(\omega_n-1,1),(1-\omega_n,1),\left(\frac{\mu_n}{2},1\right)$$

Similarly, the BER for non-coherent modulation schemes under **Case II**, can be derived using (23) resulting in

$$P_{bn}^e = P_{bn} = \phi\mathcal{M}_\gamma^e(\rho). \tag{34}$$

V. BER AND APPROXIMATIONS FOR LOW SNR REGIMES

In this section, we explore the asymptotic behaviour of the BER by employing asymptotic Poincare series expansions for both the $\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme fading models in (11) and (13). Following the idea from [33] of using [25, Eq. 8.445] and [25, Eq. 0.316], we have that the PDFs of the SNRs for the n-th hop of **Case I** and **Case II** fading channels have Poincare asymptotic series expansion near the origin given respectively by

$$p_\gamma(\gamma) \dot{\sim} \mathcal{A}\delta(\gamma) + \bar{\mathcal{A}}\mathcal{K}_n\left(\frac{\theta_n}{4}\right)^{\omega_n-1}\sum_{k=0}^{N-1} d_k\gamma^{t_k} + \mathcal{O}\left(\gamma^{t+N-1}\right) \tag{35}$$

Journal Pre-proofs

where $\mathcal{O}$ denotes the order of a term in the asymptotic series and

$$d_k = \sum_{j=0}^{k} \frac{(-1)^j \sigma_n^j \left(\frac{\theta_n}{4}\right)^{k-j}}{j!(k-j)!\Gamma(k+\mu_n-j)},\ t_k = \alpha_n(k+\mu_n)/2 - 1. \tag{36}$$

$$p_\gamma^e(\gamma) \dot{\sim} \mathcal{A}^e \delta(\gamma) + \bar{\mathcal{A}}^e \frac{a_n\sqrt{c_n}}{2} \sum_{k=0}^{N-1} d_k^e \gamma^{t_k^e} + \mathcal{O}\left(\gamma^{t+N-1}\right) \tag{37}$$

where

$$d_k^e = \sum_{j=0}^{k} \frac{(-1)^j b_n^j \left(\frac{c_n}{4}\right)^{k-j}}{j!(k-j)!\Gamma(k+2-j)},\ t_k^e = \alpha_n(k+1)/2 - 1. \tag{38}$$

We will now attempt to derive asymptotic BER expressions based on the series expansions in (35) and (37). Due to the form of the series, the process of deriving BER equates to a straightforward evaluation of Mellin transforms of appropriate functions.

1) Coherent Modulation Schemes: Substituting (35) and (37) into (26) and employing [29, Eq. 8.4.15(2)], we have that the asymptotic BER under coherent modulation for **Case I** and **Case II** are given respectively by

$$\mathcal{P}_b \dot{\sim} \frac{\phi}{2}\mathcal{A} + \bar{\mathcal{A}}\mathcal{K}_n \left(\frac{\theta_n}{4}\right)^{\omega_n - 1} \frac{\phi}{\sqrt{\pi}} \sum_{k=0}^{N-1} d_k \frac{2^{t_k}\Gamma\left(\frac{3}{2}+t_k\right)}{(t_k+1)\rho^{2(t_k+1)}}, \tag{39}$$

$$\mathcal{P}_b^e \dot{\sim} \frac{\phi}{2}\mathcal{A}^e + \phi\bar{\mathcal{A}}^e \frac{a_n\sqrt{c_n}}{2\sqrt{\pi}} \sum_{k=0}^{N-1} d_k^e \frac{2^{t_k^e}\Gamma\left(\frac{3}{2}+t_k^e\right)}{(t_k^e+1)\rho^{2(t_k^e+1)}}. \tag{40}$$

2) Non-coherent Modulation Schemes: Similarly, by substituting (35) and (37) into (32) and employing [29, Eq. 8.4.3(1)], we have that the asymptotic BER under non-coherent modulation for **Case I** and **Case II** are given respectively by

$$\mathcal{P}_{bn} \dot{\sim} \phi\mathcal{A} + \phi\bar{\mathcal{A}}\mathcal{K}_n \left(\frac{\theta_n}{4}\right)^{\omega_n - 1} \sum_{k=0}^{N-1} d_k \frac{\Gamma(t_k+1)}{\rho^{t_k+1}} \tag{41}$$

$$\mathcal{P}_{bn}^e \dot{\sim} \phi\mathcal{A}^e + \phi\bar{\mathcal{A}}^e \frac{a_n\sqrt{c_n}}{2} \sum_{k=0}^{N-1} d_k^e \frac{\Gamma(t_k^e+1)}{\rho^{t_k^e+1}}. \tag{42}$$

VI. Numerical Experiments and Simulations

In this section, the comprehensive performance evaluation of our DF-based multi-hop system has been obtained over $\alpha-\kappa-\mu$ and $\alpha-\kappa-\mu$-Extreme fading channel for **Case I** and **Case II**,

Journal Pre-proofs

respectively. Specifically, we consider equally spaced n hops between a source S and destination D. Although the results produced here are applicable for arbitrary values of the fading parameters α, κ, μ and m that are positive real numbers, and are present in terms of standard bivariate Fox H-function, still the choice of the parameters are made as described in [15], [16], [34], [35] and to give better visibility of the results. The Matlab code given in [36] has been used for the computation of bivariate Fox H-function in this work. To further ensure fairness in comparisons, $\bar{\gamma}_n$ in (5) and (9) have to be multiplied by $(n/n_{max})^{\beta}$, where β is the propagation loss exponent which is taken as 3 for the current setup. Monte Carlo simulations are included to verify and investigate the correctness of the performance metrics derived in the preceding sections.

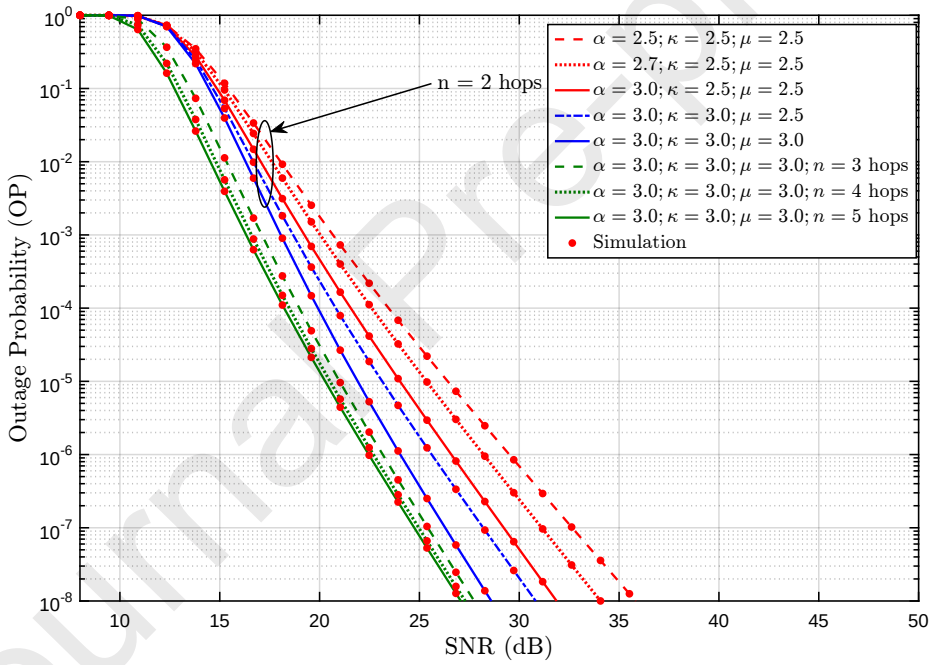


Fig. 2. OP curves for the $\alpha - \kappa - \mu$ case under moderate fading showing the effects of increments in hop number and fading model parameter values.

Figure 2 illustrates the OP vs average SNR curves for **Case I** under moderate fading conditions. From the figure, we can deduce that the increment in the value of any of the fading parameters results in notable system performance enhancement, with steeper increments noted with the increase in μ, value. Increasing the number of hops also improves system performance, especially

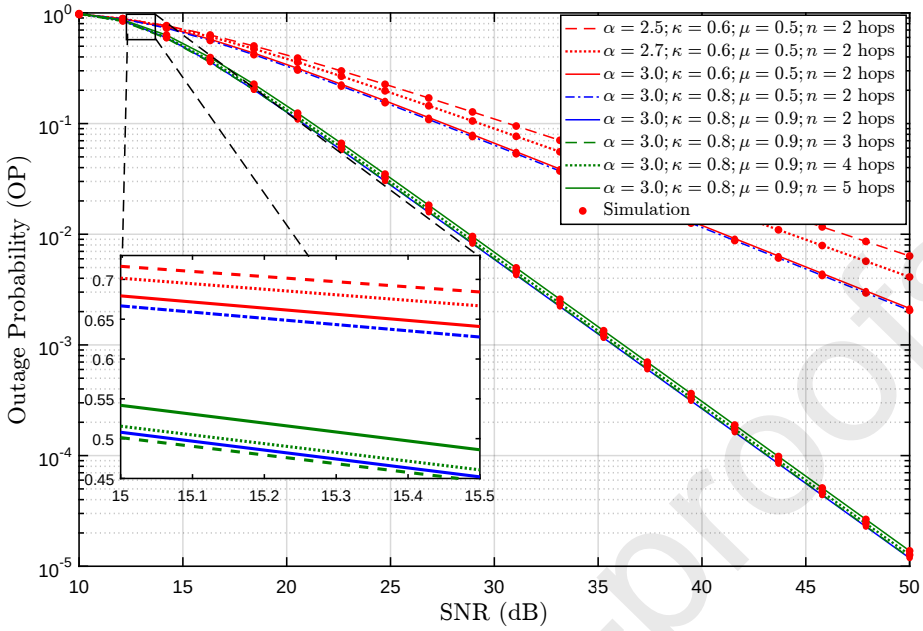


Fig. 3. OP curves for the $\alpha - \kappa - \mu$ case under severe fading showing the effects of increments in hop number and fading model parameter values.

at lower SNR levels which is a particularly desirable trait. The increase in performance is albeit of a diminishing type. This implies that there is an optimal number/range of hops that maximizes system performance. That is, relaying for the sole purpose of improving outage is only fruitful up to a certain number of relays upon which any additional relaying is futile. Another important result which was reported in [8] and has been noted to apply to the present model is that OP decreases with decrease in the spectrum efficiency $\mathcal{R}$. We avoid displaying the result here for the sake of brevity. Figure 3 shows the effect of severe fading on OP for Case I. Even though there is still a considerable system improvement for increments in parameter values, little is gained or there is actual performance degradation with the increase in number of hopes. This means that an engineer who is trying to increase coverage under severe fading conditions can afford to sacrifice a negligible loss in system performance with addition of a few hops, but will pay a steeper price in outage probability as the number of hops increases considerably.

Finally, Figure 4 shows the OP curves under Case II for severity fading scenario. It is clear

Journal Pre-proofs

that the system performs slightly better under $\alpha-\kappa-\mu$-Extreme fading model with less steeper penalties paid as a result of increase in the hop number than $\alpha-\kappa-\mu$ model. Another important result which was reported in [8] and has been noted to apply to the present model is that OP decreases with decrease in the spectrum efficiency. We avoid displaying the result here for the sake of brevity..

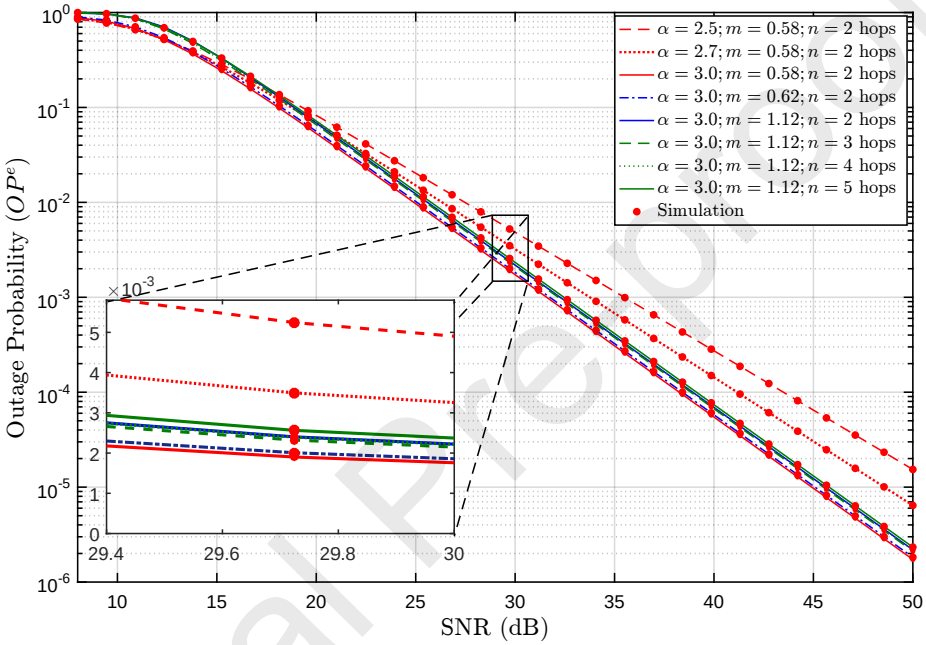


Fig. 4. OP curves for the $\alpha-\kappa-\mu$-Extreme case under severe fading conditions similar to Case I showing the effects of increments in hop number and fading model parameter values.

We now discuss simulation results of what is undoubtedly the most revealing of performance metrics, the BER. Figure 5 depicts the behaviour of BER under non-coherent DBPSK modulation scheme for **Case I** in moderate fading conditions. Consistent with observations from OP analyses, there is notable system improvement with increments in fading parameter values and a diminishing system performance improvement with the increase in hop number at low SNR levels. An additional system behavior is observed for high SNR level which shows that there is a non-null BER even as signal power increases indefinitely. This characteristic was observed by Rabelo et al. [13] for the linear $\kappa-\mu$-Extreme fading model where they showed that as SNR increases,

Journal Pre-proofs

BER under non-coherent DPSK modulation approaches the constant $(1/2)\exp(-2m)$. It is also worth pointing out that the application of multi-hops in the system is advantageous in the SNR window of 10 dB to 20 dB in the studied scenarios. Figure 6 presents BER curves under

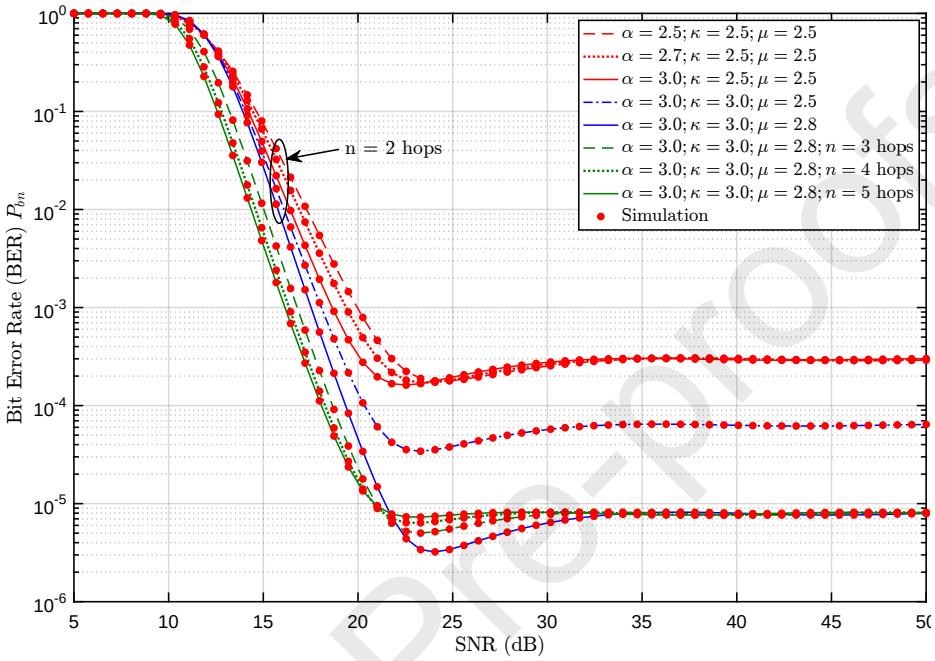


Fig. 5. BER curves for the $\alpha - \kappa - \mu$ case under non-coherent DBPSK modulation showing the effects of increasing model parameters and the number of hops and N in Poincare approximations for moderate fading conditions

non-coherent DBPSK modulation for **Case II**. We once more observe a consistent behaviour in terms of parameter and hop number variation. Furthermore, BER levels are considerably lower as compared to **Case I** even under moderate fading. This suggest that the $\alpha - \kappa - \mu$-Extreme fading model is much more robust and more flexible than its counterpart the, $\alpha - \kappa - \mu$ model. Figure 6 also shows that Poincare series-based asymptotic BER approximations perform quite well at low SNR regimes for the choice in parameter values, but the increase in N does not seem to improve convergence or guarantee improvement in accuracy of approximations. Moreover, an increase in hop number or μ results in a higher non-null BER level under **Case I** than for increases in m under Case II.

This is further supported by the system behaviour via BER curves under extremely severe fading conditions in Figure 7 and Figure 8 for **Case II** and **Case I**, respectively. From the two

Journal Pre-proofs

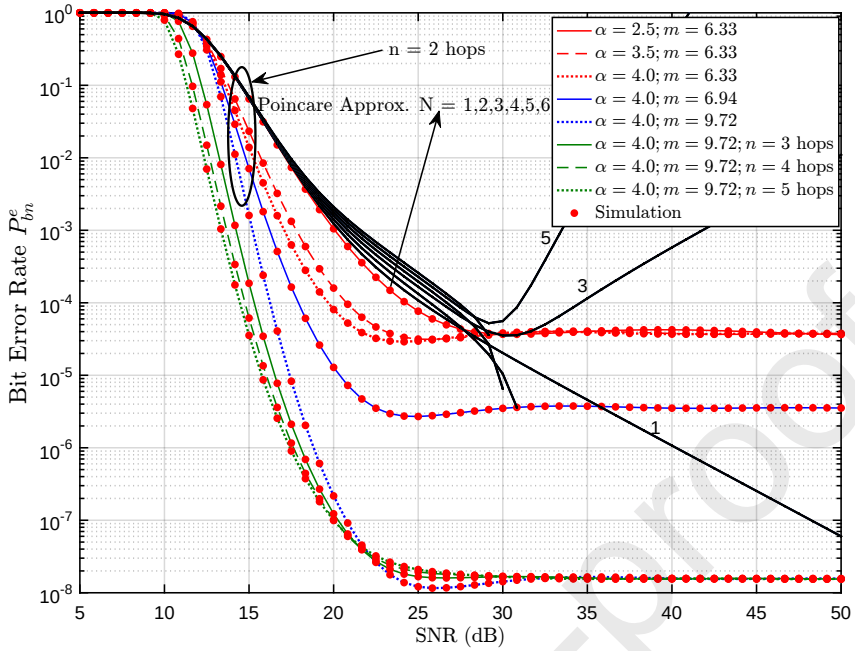


Fig. 6. BER curves for the $\alpha - \kappa - \mu$-Extreme case under non-coherent DBPSK modulation showing the effects of increasing model parameters, the number of hops and N in Poincare approximations for moderate fading conditions

figures, we note that while there appears to be some gain in system performance for increments in parameter values, system performance actually degrades as more relays are added to the system, with steeper penalties apparent for **Case I** than for **Case II**. Figure 9 depicts BER curves under Case II using coherent BPSK modulation scheme for moderate fading. From the figure, we note a consistent system behaviour as reported above with respect to parameter value and number of hop . Additionally, it is clear that one-term Poincare approximations improve with decrease in BER levels. It is also worth pointing out that the shapes of the curves for OP and BER are very similar to the shape of the Marcum Q-function curves. This can be attributed to the fact that both performance metrics are expressed as linear combinations and products of Marcum Q-functions. One of the interesting observation in all the BER plots is that irrespective of the hops number, the BER level is constant for initial values of SNR, as the SNR is further increased the benefit of using multi-hops system on the BER performance is quite visible and finally at higher SNR values all the plots become saturated at the same BER level. This gives a window of SNR values

Journal Pre-proofs

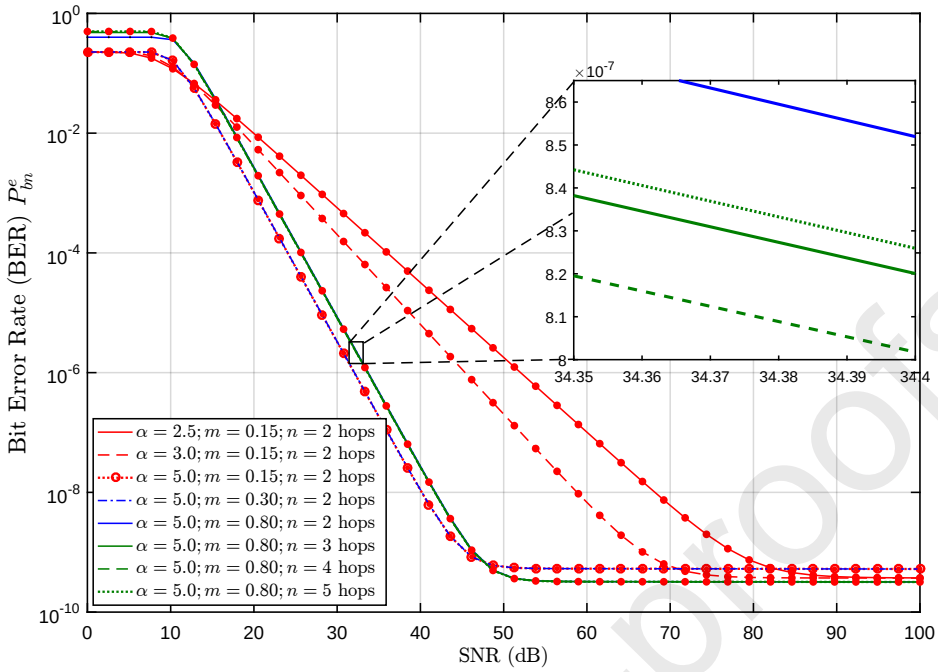


Fig. 7. BER curves for the $\alpha - \kappa - \mu$-Extreme case under non-coherent DBPSK modulation showing the effects of increasing model parameters and number of hops for extremely severe fading conditions

where the performance gains of multi-hop communication system can be explored.

VII. Conclusions

We have presented the closed-form expressions for the OP and BER for an end-to-end DF-based multi-hop relaying system over generalized $\alpha - \kappa - \mu$ and $\alpha - \kappa - \mu$-Extreme fading channels. The effects of the variation of the fading parameters and the number of hops on the system performance is demonstrated. As reported in similar literature, it was found that system performance improves with increase in any of the three parameters for both fading models, with increments in μ, the multi-path cluster parameter, resulting in steeper improvements as compared to increases in α or κ. Additionally, it was found that performance metrics based on the $\alpha - \kappa - \mu$-Extreme fading model appear to be better than those obtained for the $\alpha - \kappa - \mu$ fading model even under non-severe fading conditions. The results conclude that, DF multi-hop technique, which is used to extend the communication range of wireless systems also improves the end-to-end performance of the communication systems.

Journal Pre-proofs

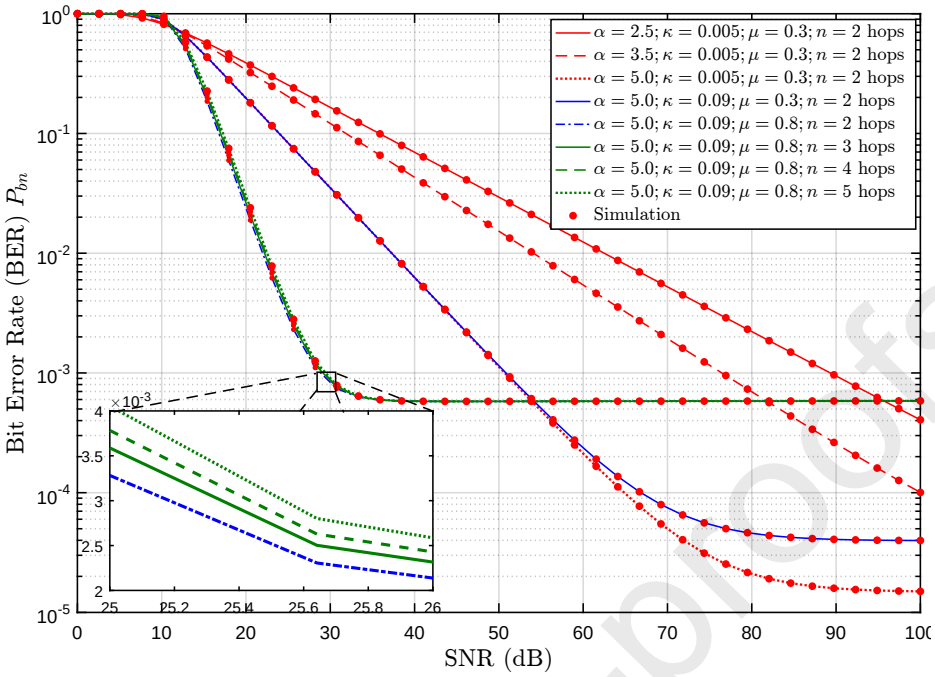


Fig. 8. BER curves for the $\alpha - \kappa - \mu$ case under non-coherent DBPSK modulation showing the effects of increasing model parameters and number of hops for extremely severe fading conditions

REFERENCES

[1] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3037–3063, 2005.

[2] S. N. Islam, P. Sadeghi, and S. Durrani, "Error performance analysis of decode-and-forward and amplify-and-forward multi-way relay networks with binary phase shift keying modulation," *IET Communications*, vol. 7, no. 15, pp. 1605–1616, 2013.

[3] D. T. J. Laneman and G. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.

[4] G. Farhadi and N. Beaulieu, "On the ergodic capacity of wireless relaying systems over rayleigh fading channels," *IEEE Transactions on Wireless Communications*, vol. 11, no. 7, pp. 4462–4467, 2008.

[5] M. R. Bhatnagar, "On the capacity of decode-and-forward relaying over rician fading channels," *IEEE Communications Letters*, vol. 17, no. 6, pp. 1100–1103, 2013.

[6] L. Yang, M. O. Hasna, and M.-S. Alouini, "Average outage duration of multihop communication systems with regenerative relays," *IEEE Transactions on Wireless Communications*, vol. 4, no. 4, pp. 1366–1371, 2005.

[7] D. Dixit and P. Sahu, "Performance of multihop communication systems with regenerative relays in η-μ fading channels," in *Vehicular Technology Conference (VTC Spring), 2014 IEEE 79th*. IEEE, 2014, pp. 1–5.

[8] F. H. Tha'er, H. B. Salameh, and T. Aldalgamouni, "Performance study of multi-hop communication systems with decode-and-forward relays over α- μ fading channels," *IET Communications*, vol. 11, no. 10, pp. 1641–1648, 2017.

Journal Pre-proofs

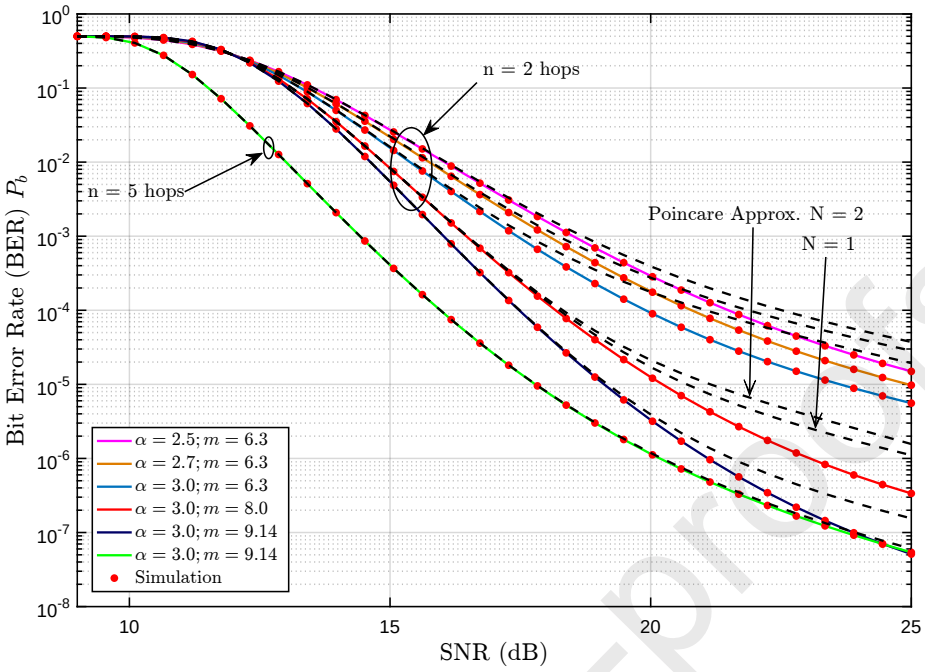


Fig. 9. BER performance for coherent BPSK modulation schemes over $\alpha - \kappa - \mu$-Extreme fading channel showing the variation in parameter values, number of hops and N in Poincare approximation

[9] D. Dixit and P. Sahu, "Exact closed-form aber for multi-hop regenerative relay systems over $\kappa - \mu$ fading," *IEEE Wireless Communications Letters*, vol. 6, no. 2, pp. 246–249, 2017.

[10] J. Cao, L.-L. Yang, and Z. Zhong, "Performance analysis of multihop wireless links over generalized-k fading channels," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 4, pp. 1590–1598, 2012.

[11] S. Kumar, "Performance of ED based spectrum sensing over α–η–μ fading channel," *Wireless Personal Communications*, vol. 100, no. 4, pp. 1845–1857, 2018.

[12] M. D. Yacoub, "The κ-μ distribution and the η-μ distribution," *IEEE Antennas and Propagation Magazine*, vol. 49, no. 1, pp. 68–81, 2007.

[13] G. S. Rabelo, U. S. Dias, and M. D. Yacoub, "The κ-μ extreme distribution: Characterizing severe fading conditions," in *Microwave and Optoelectronics Conference (IMOC), 2009 SBMO/IEEE MTT-S International*. IEEE, 2009, pp. 244–248.

[14] G. Fraidenraich and M. D. Yacoub, "The α-η-μ and α-κ-μ fading distributions," in *Spread Spectrum Techniques and Applications, 2006 IEEE Ninth International Symposium on*. IEEE, 2006, pp. 16–20.

[15] P. C. Sofotasios and S. Freear, "The α-κ-μ extreme distribution: Characterizing non-linear severe fading conditions," in *Telecommunication Networks and Applications Conference (ATNAC), 2011 Australasian*. IEEE, 2011, pp. 1–4.

[16] T. R. Rasethuntsa and S. Kumar, "An integrated performance evaluation of ed-based spectrum sensing over α-κ-μ and α-κ-μ-extreme fading channels," *Transactions on Emerging Telecommunications Technologies*, 2019.

[17] J. M. Moualeu, D. B. da Costa, W. Hamouda, U. S. Dias, and R. A. A. de Souza, "Performance analysis of digital

Journal Pre-proofs

communication systems over α - κ - μ fading channels," *IEEE Communications Letters*, vol. 23, no. 1, pp. 192–195, 2019.

[18] E. Salahat and A. Hakam, "Performance analysis of α - η - μ and α - κ - μ generalizedmobile fading channels," in *20th European Wireless Conference*, 2014, pp. 992–997.

[19] A. K. Papazafeiropoulos and S. A. Kotsopoulos, "Second-order statistics for the envelope of α - κ - μfading channels," *IEEE Communications Letters*, vol. 14, no. 4, pp. 291–293, 2010.

[20] B. Samudhyatha, S. Gurugopinath, and K. Saraswathi, "Maximum eigenvalue-based spectrum sensing over α - κ - μ and α - η - μ fading channels," in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2017, pp. 1–5.

[21] S. Kumar, M. Kaur, N. K. Singh, K. Singh, and P. S. Chauhan, "Energy detection based spectrum sensing for gamma shadowed α-η-μ and α-κ-μ fading channels," *AEU-International Journal of Electronics and Communications*, vol. 93, pp. 26–31, 2018.

[22] S. Kumar, P. Chauhan, P. Raghuwanshi, M. Kaur, and K. Singh, "Ed performance over α-η-μ/ig and α-κ-μ/ig generalized fading channels with diversity reception and cooperative sensing: A unified approach," *AEU-International Journal of Electronics and Communications*, vol. 97, pp. 273–279, 2018.

[23] N. C. Beaulieu and J. Hu, "A closed-form expression for the outage probability of decode-and-forward relaying in dissimilar rayleigh fading channels," *IEEE Communications Letters*, vol. 10, no. 12, 2006.

[24] M. K. Simon and M.-S. Alouini, *Digital communication over fading channels*. John Wiley & Sons, 2005, vol. 95.

[25] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series, and Products*. Elsevier Science, 2014.

[26] A. M. Mathai, R. K. Saxena, and H. J. Haubold, *The H-function: theory and applications*. Springer Science & Business Media, 2009.

[27] J. Proakis and S. M., *Digital Communications*. McGraw-Hill, 2005.

[28] B. Kumbhani and R. S. Kshetrimayum, *MIMO Wireless Communications over Generalized Fading Channels*. CRC Press, 2017.

[29] Y. A. Brychkov, O. Marichev, and A. Prudnikov, *Integrals and Series, vol 3: more special functions*. Gordon and Breach science publishers, 1990.

[30] A. M. Mathai and R. K. Saxena, *The H function with applications in statistics and other disciplines*. Wiley, 1978.

[31] K. Cao and X. Gao, "Solutions to generalized integrals involving the generalized marcum Q-function with application to energy detection," *IEEE Communications Letters*, vol. 20, no. 9, pp. 1780–1783, 2016.

[32] O. S. Badarneh and M. S. Aloqlah, "Performance analysis of digital communication systems over α-η-μ fading channels." *IEEE Trans. Vehicular Technology*, vol. 65, no. 10, pp. 7972–7981, 2016.

[33] A. Annamalai and E. Adebola, "Asymptotic analysis of digital modulations in $\kappa - \mu$, $\eta - \mu$ and $\alpha - \mu$ fading channels," *IET Communications*, vol. 8, no. 17, pp. 3081–3094, 2014.

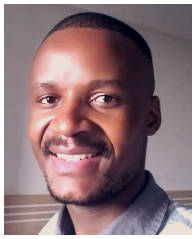
[34] R. Souza, A. Ribeiro, and D. A. Guimaraes, "On the efficient generation of α-κ-μ and α-η-μ white samples with applications," *International Journal of Antennas and Propagation*, vol. 2015, pp. 1–13, 2015.

[35] M. D. Yacoub, "The α-η-κ-μ fading model," *IEEE Transactions on Antennas and Propagation*, vol. 64, no. 8, pp. 3597–3610, 2016.

[36] H. Chergui, M. Benjillali, and M. Alouini, "Rician k-factor-based analysis of xlos service probability in 5g outdoor ultra-dense networks," *IEEE Wireless Commun. Lett.,*, vol. 8, no. 2, pp. 428–431, 2019.

Journal Pre-proofs

Author's Biographies



Tau Raphael Rasethuntsa received his BSc. in Mathematics and Statistics from the National University of Lesotho, Maseru, Lesotho in 2014 and MSc. in Mathematical Engineering from Istanbul Technical University, Istanbul, Turkey in 2017. His research interests include reliability, applied statistics as well as mathematical and statistical aspects of signal processing with special focus on performance evaluation of wireless communication systems.



Manpreet Kaur received her B. Tech. in Electronics and Communication Engineering from Punjab Technical University, Punjab, India in 2005 and Master of Engineering in Electronics and Communication from Thapar University, Patiala, India in 2007. She is pursuing her PhD. from Delhi Technological University, Delhi, India. She is currently working as Member (Senior Research Staff) at Central Research Laboratory, Ghaziabad, Bharat Electronics Limited. Her research interest includes the wireless communication and cognitive radios networks.



Sandeep Kumar received his B. Tech. in electronics and communication from Kurukshetra University, India in 2004 and Master of Engineering in Electronics and Communication from Thapar University, Patiala, India in 2007. He received his PhD. from Delhi Technological University, Delhi, India in 2018. He is currently working as Member (Senior Research Staff) at Central Research Laboratory, Bharat Electronics Limited Ghaziabad, India. His research interests include the study of wireless channels, performance modeling of fading channels and cognitive radio networks. He is also serving as a reviewer for several international journals of IEEE, Springer, Elsevier etc.

Journal Pre-proofs

Physical Communication 41 (2020) 101114

Contents lists available at ScienceDirect

Physical Communication

journal homepage: www.elsevier.com/locate/phycom

Full length article

Performance analysis of beamforming aided uncorrelated mm-Wave MIMO system for IEEE 802.15.3c standard

Anand Agrawal [a,*], Vimal Kumar Mishra [b], Ashish Gupta [b], Bajrang Bansal [b]

[a] Department of Electronics and Communication Engineering, Indian Institute of Information Technology, Kota, Rajasthan, 302017, India
[b] Department of Electronics and Communication Engineering, Jaypee Institute of Information Technology, Noida, U.P, 201301, India

ARTICLE INFO

Article history:
Received 25 October 2019
Received in revised form 17 February 2020
Accepted 1 May 2020
Available online 12 May 2020

Keywords:
Mm-wave communication
Analog beamforming
IEEE802.15.3c standard

ABSTRACT

Millimeter wave (mm-wave) at 60 GHz is a revolutionary technology for ultra-high data rate indoor communications. Unfortunately, due to unfavorable channel characteristics, communication at 60 GHz suffers from poor link budget, high penetration loss and severe signal attenuation. To improve the link budget and the quality of the received signal, the beamforming techniques must be incorporated. In this paper, we derive analytically tractable bit error rate (BER) for the analog beamforming aided mm-wave communication system over the modified IEEE 802.15.3c channel model. It also takes into account the impact of the Rake receivers, which are used to capture the signal energy carried by the multipath components. Monte Carlo simulations have also been carried out to validate the accuracy of the analytical framework. In addition, the impact of Rake's fingers, power decay factors and arrival times on the BER of the 60 GHz system is also highlighted.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Millimeter wave (mm-wave) communication at 60 GHz has shown enormous potential for the bandwidth-hungry applications, such as high definition video streaming and live gaming. It has also drawn great attention because of the large available bandwidth (7 GHz) [1,2]. In 2005, the Federal communications commission formally affirmed the 57–64 GHz frequency spectrum for unlicensed utilization of commercial applications. Specifically, at 60 GHz, this commission also formed a task group (TG) which has published several standards, targeting wireless personal and local area networks such as IEEE 802.15.3c, IEEE 802.11ad and ECMA 387 [3,4]. In 2007, TG (TG3c) carried out several measurements of the IEEE 802.15.3c standard and published their parametric specifications for different indoor environments [5]. Although there are quite a few merits of using a large continuous spectrum, 60 GHz spectrum faces various challenges. Some of them include unfavorable channel characteristics, large path loss (68 dB/m), attenuation due to rain, foliage and atmospheric oxygen [6,7]. This reduces the coverage area of the mm-wave technology and makes the spectrum not suited for long-range communication. Often, the required antenna spacing ($\lambda/2$) in mm-wave communication is of the order of millimeters which facilitates in integrating a large number of antennas array into a smaller area [8–11]. In mm-wave communication, these antenna arrays are combined at the transceivers which on using antenna beamforming technique can potentially overcome the poor propagation characteristics of the channel and can also improve the coverage area [12]. One of the major challenges in designing such transceivers is the implementation of the fully digital beamforming because it requires a separate radio frequency (RF) chain for each antenna element. This leads to a high hardware complexity and excessive power consumption. Therefore, people have suggested the analog and digital hybrid beamforming (HBF) architecture [13,14]. This architecture helps in directionalizing the signal transmission and reception by changing the amplitude and phase of the signal with the help of power variation and beam steering, respectively. The principle of HBF is to partition the conventional beamforming scheme into the digital as well as the analog beamforming. In digital beamforming, the amplitude or phase of the digital signal is varied at the transmit end. It also requires multiple RF chains which are proportional to the antenna elements in use. On the other hand, the analog beamforming is implemented by a set of variable phase shifters. Here, the amplitude/phase of the analog signal is varied at the transmit end. Unlike digital beamforming, it is restricted to one RF chain even when the large number of antenna arrays are being used. Hence, analog beamforming is best suited for coverage due to low power consumption & low cost characteristics. Moreover, it effectively addresses the hardware limitation of the fully-digital beamforming scheme [15].

* Corresponding author.
E-mail address: anand.ece@iiitkota.ac.in (A. Agrawal).
URL: http://www.iiitkota.ac.in/anand (A. Agrawal).

https://doi.org/10.1016/j.phycom.2020.101114
1874-4907/© 2020 Elsevier B.V. All rights reserved.

1.1. Related work

Several recent works have investigated the various performance measures of HBF aided mm-wave system. Hybrid precoding in mm-wave system has effectively reduced the number of RF chains but it limits the spatial multiplexing gain whereas, spatial modulation can convey an extra data stream via the index of the active antennas group. To further improve the achievable spectral efficiency (SE), the performance of spatial modulation aided mm-wave multiple-input-multiple-output (MIMO) system was studied in [16,17]. With this emphasis, in [18] authors have studied the spatially sparsed precoding based mm-wave MIMO system. Furthermore, the simulation and semi-analytical results of bit error rate (BER) for mm-wave system over narrowband fading channels were presented in [19–21]. The authors in [22–24] investigated the BER of mm-wave MIMO system over simplified IEEE 802.15.3c channel. The BER analyses discussed above have either considered the conventional narrow-band fading channels or simplified version of IEEE 802.15.3c channel models. It did not take into account the random nature of the total number of clusters, arrival times and the lognormal behavior of multipath gain coefficients [5]. This questions on the sustainability of the presented BER analyses, as these are very important aspects of wireless networks and can not be ignored while evaluating the performance of mm-wave MIMO systems at 60 GHz spectrum. Moreover, the original IEEE 802.15.3c channel model does not support beamforming technique because it fails to define the angle of departure (AOD) of the channel paths. Therefore, the modified version of the IEEE 802.15.3c channel model (define in the Section 2) was suggested to evaluate the performance of the 60 GHz system with beamforming architecture [25,26]. In [27], the authors calculated the average signal-to-noise ratio for an orthogonal frequency division multiplexing (OFDM) system over the modified IEEE 802.15.3c channel model.

1.2. Main contribution

To the best of the authors' knowledge, the BER analysis of mm-wave MIMO system with beamforming architecture has not been reported in the literature. The contribution of the paper can be summarized in the following bullet points:

- Theoretical analysis is carried out to drive a computable BER formula for analog beamforming aided mm-wave MIMO system over the modified version of the IEEE 802.15.3c channel model by using Gauss-Quadrature rule of numerical integration. This numerical rule facilitates in finding the computable solutions of mathematically intractable integrals [28].
- We have considered the impact of the Rake receiver which gives an insight of the number of Rake's fingers (L) required to capture the maximum signal energy carried by the multipath components (MPCs) in the given time window. Rake's fingers are nothing more than the set of Rake receiver, where each finger can independently tune a single multipath component.
- Characteristic function based probability density function (PDF) for approximate signal energy collected by the fingers of the Rake receiver in the given time window is also presented in this paper.
- We have also discussed the impact of various parameters of IEEE 802.15.3c channel on the BER. In addition, Monte Carlo simulations of the investigating system for IEEE 802.15.3c channel have also been carried out to validate the accuracy of the analysis.

1.3. Paper organization and notations

The rest of the paper is organized as follows: System and channel model are described in Section 2. Analytical BER expression is derived in Section 3. Section 4 presents the simulation results and discussions followed by the conclusion in Section 5.

Notations. Lowercase boldface letters represent the vectors. $\mathbb{C}^{m\times n}$ denotes an $m \times n$ matrix. $\psi_X(\cdot), f_X(\cdot)$ and $F_X(\cdot)$ respectively represent the characteristic function, PDF and cumulative distribution function of a random variable X. $\mathcal{N}(\mu, \sigma^2)$ represents Gaussian random variable with mean μ and variance σ^2, and $\mathrm{E}[\cdot]$ is the expectation operator. Furthermore, $|.|$, $\|.\|_2$ and $(.)^*$ are the absolute value, 2-norm and conjugate of a vector respectively; while $I_A(.)$ denotes an indicator function.

2. System and channel model

In this section, modified version IEEE 802.15.3c channel model and mm-wave MIMO system with antenna beamforming have been described.

2.1. Modified IEEE 802.15.3c channel model

IEEE 802.15.3c channel model is known as the modified Saleh-Valenzuela (SV) indoor channel model. It combines a line-of-sight (LOS) and non-LOS components of the S-V model using a two-path model [29] and gives the information about the spatial temporal phenomenon and angle-of-arrival (AOA), which have been widely used for describing the multi-path propagation behavior in the indoor environment. Besides, this model has been failed to define AOD of channel paths which is necessary for antenna beamforming [30–33]. Therefore, this analysis considers the modified version of the IEEE 802.15.3c channel, which has four parameters, including multipath gain coefficients, time of arrival (TOA), AOA and AOD. Here, AOD is assumed to have same statistics as that of the AOA. During propagation, multipath coefficients fall in the form of rays and clusters. If M is the total number of clusters and R is the number of rays falling in each cluster then the impulse response of the modified version of IEEE 802.15.3c channel is given by [34]

$$h(t,\theta,\phi)=\beta\delta(t)\,\delta(\theta)\,\delta(\phi)+\sum_{m=1}^{M}\sum_{r=0}^{R}\alpha_{r,m}\delta\left(t-T_m-\tau_{r,m}\right)\times\ \delta\left(\theta-\theta_m-\omega_{r,m}\right)\delta\left(\phi-\phi_m-\varpi_{r,m}\right) \quad (1)$$

where β represents the gain coefficient of the LOS path which depends on the antenna-height, antenna gain, Tx–Rx separation and path loss of the first impulse. The parameter $\alpha_{r,m}$ denotes the complex MPCs of the rth ray in the mth cluster and it is distributed as lognormal distribution. The parameters T_m, θ_m and ϕ_m are the TOA, AOA and AOD of the mth cluster respectively; while $\tau_{r,m}$, $\omega_{r,m}$ and $\varpi_{r,m}$ are the TOA, AOA and AOD for the rth ray of the mth cluster respectively. According to TG3c, the total number of clusters (M) is assumed to be random whose probability mass function (PMF) is given as

$$f_M(m)=\begin{cases}\left(1+\bar{L}\right)e^{-\bar{L}}, & m=1\\ \bar{L}^m e^{-\bar{L}}/m!, & m\geq 2\\ 0, & otherwise\end{cases} \quad (2)$$

where $\bar{L}=E[m]$. The propagation behavior of rays and clusters in the IEEE 802.15.3c channel is shown in Fig. 1. The figure clearly shows that the arrival times of the channel paths are random and independent to each other. The parameter Ω_0 is the average power of the first ray of the first cluster and Γ and γ are the

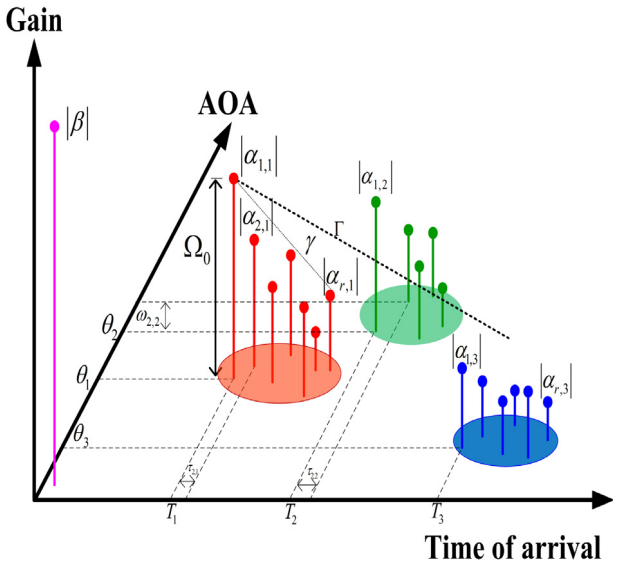


Fig. 1. Graphical representation of the multipath propagation in IEEE 802.15.3c channel.

inter-cluster and intra-cluster power decay factors respectively. According to TG3c channel report, the arrival times of the cluster and ray are modeled as Poisson processes. Therefore, the distribution of the cluster inter-arrival time and the ray inter-arrival time are modeled as exponential random variables, whose PDFs are given as

$$f\left(T_m/T_{m-1}\right) = \Lambda \exp\left[-\Lambda\left(T_m - T_{m-1}\right)\right], \quad m > 0 \tag{3}$$

and

$$f\left(\tau_{r,m}/\tau_{r-1,m}\right) = \lambda \exp\left[-\lambda\left(\tau_{r,m} - \tau_{r-1,m}\right)\right], \quad r > 0 \tag{4}$$

respectively, where Λ and λ denote the arrival rates of the cluster and ray respectively. In the angular domain, AOA of the LOS component, $\{\theta_{LOS}\}$, is assumed to be zero and the distribution of the cluster AOAs, $\{\theta_m\}$, is assumed to be independent and identically distributed (i.i.d.) over $[0,\ 2\pi]$. On the other hand, the distribution of the relative AOAs, $\{\omega_{r,m}\}$, is assumed to be independent Gaussian random variable with zero mean and standard deviation σ_ϕ, and its PDF is given as

$$f\left(\omega_{r,m}\right) = \frac{1}{\sqrt{2\pi}\sigma_\phi} \exp\left(-\omega_{r,m}^2/2\sigma_\phi^2\right) \tag{5}$$

The second-order moment of $\left(\alpha_{r,m}\right)$ can be calculated as (see Appendix for its proof)

$$\mathrm{E}\left[\left|\alpha_{r,m}\right|^2\right] = \Omega_0 \exp\left(-\frac{T_m}{\Gamma} - \frac{\tau_{r,m}}{\gamma} - \tilde{\kappa} I_{[\tilde{\kappa}\neq 0]}\right) \tag{6}$$

where $\tilde{\kappa}$ is the Rician factor.

On the basis of practical measurements, IEEE P802.15 wireless personal area network (WPAN) channel modeling subcommittee has studied 60 GHz network over various environments. These channel environments are further categorized according to the scenarios type (LOS/ NLOS) and the Half-Power Beam-Width (HPBW) of the transceiver antenna array. The detail description and the measured values of the channel's parameters for each environment are given in [5].

2.2. System model

An analog beamforming aided mm-wave MIMO system having N_t transmitting and N_r receiving antennas with a bank of Rake receivers is shown in Fig. 2. Initially, the mm-wave signal is split by the power diver and passed through the analog beamformer. The analog beamforming is implemented via phase shifters which have the provision to change the phase of the signals in each of the paths going towards the uniform linear antennas arrays (ULAs).

Note that the mm-wave technology at 60 GHz exclusively supports the rank one transmission, i.e., all transmitting antennas are fed with the same symbol at one time instant. The mathematical expression of binary phase shift keying (BPSK) modulated signal is given by [35]

$$s\left(t\right) = \sqrt{E_b} \sum_{i=-\infty}^{\infty} b_i\, p\left(t - iT_s\right) \tag{7}$$

where E_b denotes the signal energy per bit, $b_i \in \{\pm 1\}$ represents the ith transmitted symbol, $p\left(t\right)$ is the unit energy of the raised-cosine pulse of duration T_p, and T_s is the symbol duration. We denote $R_g\left(.\right)$ as the normalized autocorrelation function of $p(t)$, where $R_g\left(\delta\right) \approx 0$ for $|\delta| \geq T_p$. In the derivation, it is assumed that $T_s = T_p$ and the transmitting and receiving antennas are perfectly aligned to each other. Therefore, the effective mm-wave channel impulse response (CIR) after incorporating the ULAs at the transmitter and the receiver can be expressed as [18]

$$\boldsymbol{h}\left(t\right) = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \boldsymbol{h}\left(t,\theta,\phi\right) \boldsymbol{\vartheta}\left(\theta\right) \boldsymbol{\zeta}\left(\phi\right) d\theta d\phi \tag{8}$$

where $\boldsymbol{\vartheta}\left(\theta\right)$ and $\boldsymbol{\zeta}\left(\phi\right)$ represent the array response vector of ULAs at the transmitter and the receiver respectively. After satisfying $\|\boldsymbol{\vartheta}\left(\theta\right)\|_2 = 1$ and $\|\boldsymbol{\zeta}\left(\phi\right)\|_2 = 1$, these are respectively defined as

$$\boldsymbol{\vartheta}\left(\theta\right) = \frac{1}{\sqrt{N_t}}\left[1, e^{\left(\frac{j2\pi d \sin(\theta)}{\lambda_c}\right)}, \ldots, e^{\left(\frac{j2\pi (N_t-1) d \sin(\theta)}{\lambda_c}\right)}\right]^T, \tag{9}$$

and

$$\boldsymbol{\zeta}\left(\phi\right) = \frac{1}{\sqrt{N_r}}\left[1, e^{\left(\frac{j2\pi d \sin(\phi)}{\lambda_c}\right)}, \ldots, e^{\left(\frac{j2\pi (N_r-1) d \sin(\phi)}{\lambda_c}\right)}\right]^T, \tag{10}$$

where $d = \lambda_c/2$, λ_c is the wavelength of the carrier frequency and $\boldsymbol{h}\left(t\right) \in \mathbb{C}^{N_r \times N_t}$ is the effective CIR of mm-wave system. The received signal vector, $\boldsymbol{r}(t)$, can be obtained as

$$\boldsymbol{r}(t) = \boldsymbol{h}(t)\boldsymbol{s}(t) + \boldsymbol{\eta}(t) \tag{11}$$

where $\boldsymbol{r}(t) \in \mathbb{C}^{N_r}$, $\boldsymbol{s}(t)$ is a $(N_t \times 1)$ vector of the same transmitted signal $s(t)$, and $\boldsymbol{\eta}(t) \in \mathbb{C}^{N_r}$ is the Gaussian noise vector whose samples are $\mathcal{N}\left(0, \frac{N_t N_0}{N_r}\right)$. Since, all the transmitting antennas are independent and transmit one symbol at one time instant, the received signal at the vth antenna is given by

$$r_v\left(t\right) = \sqrt{\frac{E_b}{N_t N_r}} \sum_{i=-\infty}^{\infty} \sum_{u=1}^{N_t} b_i \left\{\beta_{u,v} p\left(t - iT_s\right) + \sum_{m=1}^{M}\sum_{r=0}^{R} \alpha_{r,m}^{u,v} p\left(t - \delta_{r,m}^{u,v} - iT_s\right) e^{\left(j\Phi_{r,m}^{u,v}\right)}\right\} + n_v\left(t\right) \tag{12}$$

where $\Theta_{r,m}^{u,v} = \frac{2\pi d}{\lambda_c}\left[(u-1)\sin\left(\theta_m^{u,v} + \omega_{r,m}^{u,v}\right) + (v-1)\sin\left(\phi_m^{u,v} + \varpi_{r,m}^{u,v}\right)\right]$ and $\delta_{r,m} = T_m + \tau_{r,m}$, $n_v\left(t\right) \sim \mathcal{N}\left(0, \frac{N_t N_0}{N_r}\right)$.

3. Proposed BER analysis

In this section, using the characteristic function based approach, we derive the BER of binary signals for analog beamforming aided mm-wave MIMO system with the bank of Rake

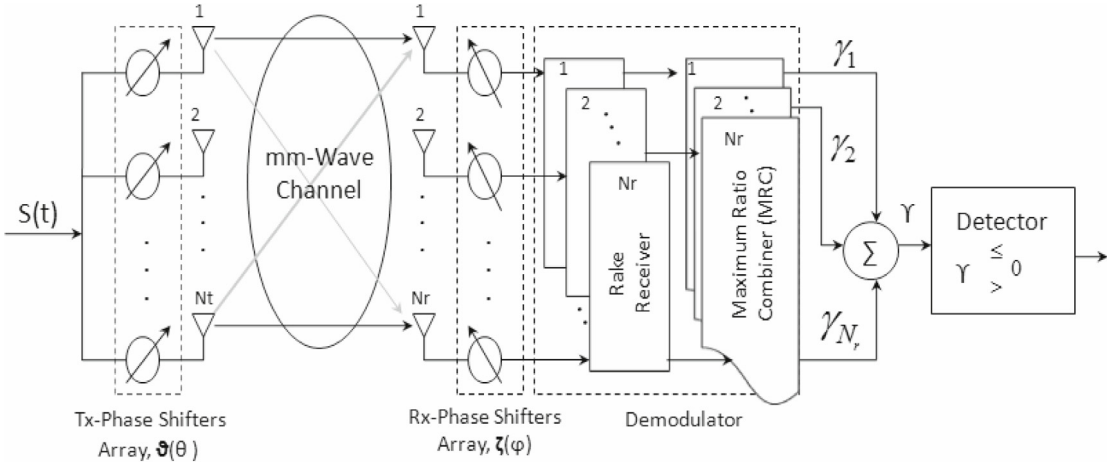


Fig. 2. Block diagram of phase shifters aided mm-wave MIMO system.

receivers. The average BER of binary signals with a coherent Rake receiver can be computed as [36]

$$p_e = \int_0^\infty Q\left(\sqrt{(1-\rho_r)\,\gamma}\right) f_\Upsilon(\gamma)\, d\gamma \tag{13}$$

where $\rho_r = -1$ and $\rho_r = 0$ for the antipodal signals and the orthogonal binary signals respectively, and $f_\Upsilon(\gamma)$ is the PDF of the received SNR (Υ).

3.1. Calculation of the total received instantaneous SNR (Υ)

According to [37], the template signal of rake receiver $\chi(t)$, under the assumption of perfect channel state information is available at the receiver and is given by

$$\chi(t) = \beta p\left(t - i'T_s\right) + \sum_{m'=1}^{M}\sum_{r'=0}^{R} \alpha_{r',m'} p\left(t - \delta_{r',m'} - i'T_s\right) e^{\left(j\Phi_{r',m'}\right)} \tag{14}$$

The output of the Rake receiver can be computed as $R_v = \int_{i'T_s}^{(i'+1)T_s} r_v(t)\,\chi_v^*(t)\,dt$, where $\chi_v^*(t)$ denotes the complex conjugate of the Rake receiver's template. Under the assumption of perfect synchronization, i.e., $(m = m')$ & $(r = r')$ and using the relation $R_g(\delta) = R_g(-\delta)$, the output of the correlator at the vth antenna can be evaluated as

$$R_v = \sqrt{\frac{E_b}{N_t N_r}} \sum_{i'=0}^{\infty}\sum_{u=1}^{N_t} b_{i'} \left(\beta_{u,v}^2 + \sum_{m=1}^{M}\sum_{r=0}^{R} 2\beta_{u,v}\,\alpha_{r,m}^{u,v} R_g\left(\delta_{r,m}^{u,v}\right) \times \cos\left(\Phi_{r,m}^{u,v}\right) + \sum_{m=1}^{M}\sum_{r=0}^{R}\left|\alpha_{r,m}^{u,v}\right|^2\right) + \tilde{n}_v(t) \tag{15}$$

where $\tilde{n}_v(t) = \int_{i'T_s}^{(i'+1)T_s} n_v(t)\,\chi^*(t)\,dt$ is the noise component. Assuming inter-pulse interference free transmission i.e. $R_g\left(\delta_{r,m}^{u,v}\right) \approx 0$, the above equation can be written as

$$R_v = \sqrt{\frac{E_b}{N_t N_r}} \sum_{i'=0}^{\infty}\sum_{u=1}^{N_t} b_{i'} \left(\beta_{u,v}^2 + \sum_{m=1}^{M}\sum_{r=0}^{R}\left|\alpha_{r,m}^{u,v}\right|^2\right) + \tilde{n}_v(t) \tag{16}$$

Considering all the receiving antennas to be sufficiently far away, the approximate received SNR (Υ) can now be viewed as the summation of the individual received SNRs at N_r antennas and is given by

$$\Upsilon \approx \frac{E_b}{N_t^2 N_0} \sum_{v=1}^{N_r}\sum_{u=1}^{N_t} \left(\beta_{u,v}^2 + \sum_{m=1}^{M}\sum_{r=0}^{R}\left|\alpha_{r,m}^{u,v}\right|^2\right) \tag{17}$$

3.2. PDF of the total received instantaneous SNR (Υ)

Let us define $\mathcal{E}_{u,v} = \sum_{m=1}^{M}\sum_{r=0}^{R}\left|\alpha_{r,m}^{u,v}\right|^2$ as the energy carried by all multipath rays excluding the LOS component in the $(u-v)$th link of modified IEEE 802.15.3c channel. According to [38], characteristic function, $\boldsymbol{\Psi}_\Upsilon(\nu)$, of Υ in i.i.d. mm-wave channel links can be computed by

$$\boldsymbol{\Psi}_\Upsilon(\nu) \approx \prod_{v=1}^{N_r}\prod_{u=1}^{N_t} \psi_{\mathcal{E}}^{u,v}\left(\frac{E_b}{N_t^2 N_0}\nu\right) \exp\left(\frac{jE_b\,\beta_{u,v}^2}{N_t^2 N_0}\nu\right) \tag{18}$$

where $\psi_{\mathcal{E}}^{u,v}(.)$ is the characteristic function of $\mathcal{E}_{u,v}$ which is defined in the next subsection. Using (16), PDF of the received SNR is calculated as $f_\Upsilon(\gamma) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \boldsymbol{\Psi}_\Upsilon(\nu)\, e^{(-j\nu\gamma)} d\nu$. This integration has no closed form solution and therefore, by using the Gauss–Hermite numerical quadrature rule, the approximate expression of $f_\Upsilon(\gamma)$ can be written as

$$f_\Upsilon(\gamma) = \frac{1}{2\pi}\sum_{k=1}^{N^H} \left\{w_k^H\right\} e^{\left(-j\gamma x_k^H + \left(x_k^H\right)^2\right)} \times \left(\prod_{v=1}^{N_r}\prod_{u=1}^{N_t} \psi_{\mathcal{E}}^{u,v}\left(\frac{E_b}{N_t^2 N_0} x_k^H\right) e^{\left(\frac{jE_b\beta_{u,v}^2}{N_t^2 N_0} x_k^H\right)}\right) \tag{19}$$

where x_k^H, w_k^H and N^H represent the kth root, kth weight and the number of nodes of the Gauss–Hermite quadrature rule respectively.

3.3. Calculation for $\psi_{\mathcal{E}}^{u,v}(\nu)$

For simplicity, we omit the antenna indexing (u, v) in the calculation of the characteristic function, $\mathcal{E}_{u,v}$, which is already defined in the above subsection. From the definition of parameter M which is given in (2), the average received energy $(\tilde{\mathcal{E}})$ with respect to parameter M can be computed as

$$\begin{aligned}\tilde{\mathcal{E}} &= \mathrm{E}_M[\mathcal{E}] \\ &= \sum_{m=1}^{\infty}\left(\sum_{\ell=1}^{m}\sum_{r=0}^{R}\left|\alpha_{r,\ell}\right|^2\right) f_M(m) \\ &\approx \left(1+\bar{L}\right)e^{-\bar{L}}\mathcal{E}_1 + \sum_{m\geq 2}\mathcal{E}_m\left(\frac{\bar{L}^m e^{-\bar{L}}}{m!}\right)\end{aligned} \tag{20}$$

where $f_M(m)$ is the PMF of M and parameters $\mathcal{E}_1 = \sum_{r=0}^{R}\left|\alpha_{r,1}\right|^2$ and $\mathcal{E}_m = \sum_{\ell=2}^{m}\sum_{r=0}^{R}\left|\alpha_{r,\ell}\right|^2$ are the energies carried by the

MPCs of the first cluster and the remaining clusters respectively. In this analysis, we conjecture that $\mathcal{E}_1$ is close to $\tilde{\mathcal{E}}_1$, the sum of the squared path gains of the first cluster fall within $[0, LT_c]$ time window [36], where L is the number of fingers of the Rake receiver and T_c is the chip duration between two consecutive fingers. Similarly, $\mathcal{E}_m$ is close to $\tilde{\mathcal{E}}_m$, and the sum of squared path gains of rays in the remaining clusters are defined as $\tilde{\mathcal{E}}_1 = \sum_{(r,1):0<(T_1+\tau_{r,1})\leq LT_c} |\alpha_{r,1}|^2$ and $\tilde{\mathcal{E}}_m = \sum_{(r,m):0<(T_m+\tau_{r,m})\leq LT_c} |\alpha_{r,m}|^2$. Note that $\tilde{\mathcal{E}}_1$ and $\tilde{\mathcal{E}}_m$ are assumed to be statistically independent to each other. Therefore, the characteristic function, $\psi_{\tilde{\mathcal{E}}}(\nu)$, of $\tilde{\mathcal{E}}$ can be evaluated as

$$\psi_{\tilde{\mathcal{E}}}(\nu) \approx \psi_{\tilde{\mathcal{E}}_1}(\zeta_1) \left\{ \prod_{m\geq 2} \psi_{\tilde{\mathcal{E}}_m}(\zeta_2) \right\} \tag{21}$$

where $\psi_{\tilde{\mathcal{E}}_1}(.)$ and $\psi_{\tilde{\mathcal{E}}_m}(.)$ are the characteristic functions of $\tilde{\mathcal{E}}_1$ and $\tilde{\mathcal{E}}_m$ respectively. The parameters $\zeta_1 = \left(1+\bar{L}\right) e^{-\bar{L}}\nu$ and $\zeta_2 = \frac{\bar{L}^m e^{-\bar{L}}\nu}{m!}$. From [39], $\psi_{\tilde{\mathcal{E}}_1}(\zeta_1) = \exp\left[-\lambda\tilde{\psi}_{\zeta_1}(0,L)\right]$ and $\psi_{\tilde{\mathcal{E}}_m}(\zeta_2) = \exp\left[-\Lambda\tilde{J}(\zeta_2, L)\right]$. According to [30, eq.(14)], function $\tilde{\psi}_{\zeta_1}(T, L)$ can be computed as

$$\tilde{\psi}_{\zeta_1}(T, L) = \begin{cases} \int_T^{LT_c} \left[1 - \mathcal{L}_{T,t}(\zeta_1)\right] dt, & \leq LT_c \\ 0, & > LT_c \end{cases} \tag{22}$$

where $\mathcal{L}_{T,t}(.)$ represents the characteristic function of the square of single path coefficient in the IEEE 802.15.3c channel, with ray arrival time $t = T + \tau$. Using the Gauss–Hermite quadrature, the Hermite solution of $\mathcal{L}_{T,t}(.)$ can be evaluated as

$$\mathcal{L}_{T,t}(\nu) \approx \frac{1}{\sqrt{\pi}} \sum_{k=1}^{N^H} w_k^H \exp\left(j\nu 10^{\left(\frac{\sqrt{2}\sigma x_k^H + \mu_{T,t}}{10}\right)} \right) \tag{23}$$

where $\mu_{T,t} = \frac{10}{\ln(10)}\left[\ln\Omega_0 - \frac{T}{\Gamma} - \frac{t-T}{\gamma} - \tilde{\kappa} I_{[\tilde{\kappa}\neq 0]} - \left(\frac{\ln 10}{10}\right)^2 \frac{\sigma^2}{2}\right]$ and $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$. The proof of $\mu_{T,t}$ is given in the Appendix. Now using Gauss–Legendre quadrature, the above characteristic function can be written as

$$\tilde{\psi}_{\zeta_1}(T, L) \approx \frac{LT_c - T}{2} \sum_{b=1}^{N^L} w_b^L \left[1 - \mathcal{L}_{T,t}(\zeta_1)\right]\Big|_{t=\frac{LT_c-T}{2}x_b^L + \frac{LT_c+T}{2}} \tag{24}$$

Similarly, function $\tilde{J}(\zeta_2, L)$ can be computed as

$$\tilde{J}(\zeta_2, L) \approx \frac{LT_c}{2} \sum_{b=1}^{N^L} w_b^L \left[1 - \mathcal{L}_{T,T}(\zeta_2)\, e^{-\lambda\tilde{\psi}_{\zeta_1}(T,L)}\right]\Big|_{T=\partial_b^L} \tag{25}$$

where $\partial_b^L = \frac{LT_c}{2}\left(x_b^L + 1\right)$ and $\left\{x_b^L, w_b^L, N^L\right\}$ are the bth root, bth weight and number of nodes of the Gauss–Legendre quadrature rule respectively. Combining (11), (17), and (19) and using Gauss Laguerre quadrature $\int_0^\infty e^{-x} f(x)\, dx = \sum_{a=1}^{M^L} w_a f\left(x_a^L\right)$, the derived BER expression is given by

$$\bar{p}_e \approx \frac{1}{2\pi} \sum_{i_1=1}^{M^L} \sum_{k=1}^{N^H} \left\{ w_{i_1}^L w_k^H e^{\left(x_k^H\right)^2} Q\left(\sqrt{\frac{(1-\rho_r)\, x_{i_1}^L}{j x_k^H}}\right) \times \prod_{v=1}^{N_r} \prod_{u=1}^{N_t} e^{\left[\frac{jE_b}{N_t^2 N_0}\beta_{u,v}^2 x_k^H - \lambda\tilde{\psi}_{\tilde{\kappa}_1}^{u,v}(0,L)\right]} \prod_{m\geq 2} e^{\left[-\Lambda\tilde{J}^{u,v}(\tilde{\kappa}_2, L)\right]} \right\} \tag{26}$$

where $\tilde{\kappa}_1 = \frac{E_b(1+\bar{L})e^{-\bar{L}}}{N_0 N_t^2} x_k^H$ and $\tilde{\kappa}_2 = \frac{E_b}{N_0 N_t^2}\frac{\bar{L}^m e^{-\bar{L}}}{m!} x_k^H$. The parameters $x_{i_1}^L$, $w_{i_1}^L$ and M^L represent the kth root, kth weight and the number of nodes of the Gauss–Legendre quadrature respectively.

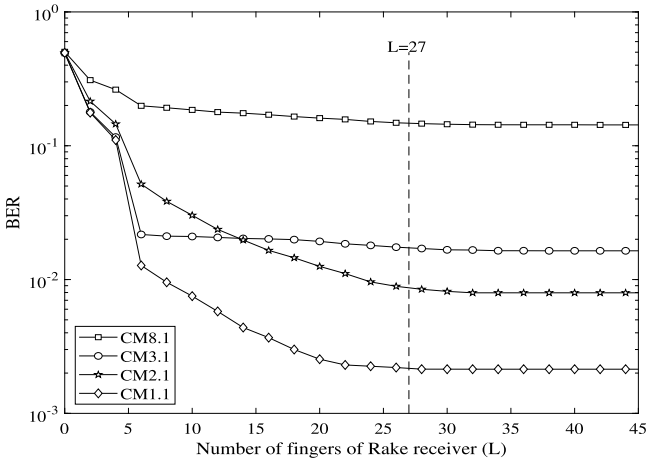


Fig. 3. Fingers of the Rake receiver vs BER of mm-wave SISO communication system at 5 dB SNR.

4. Simulation results and discussions

This section presents the simulated and analytical BERs of the mm-wave MIMO communication system for four different environments of the IEEE 802.15.3c channel model: Residential LOS (CM1), Residential NLOS (CM2), Office LOS (CM3) and Desktop NLOS (CM8). The parametric specifications of IEEE 802.15.3c model for these environments are given in the report published by the TG3c [5]. The simulation has been carried out for 10^6 samples of signal of interest and same has been averaged for 10^6 times. These signals are first transmitted by the highly directional antennas and then passed through the modified version of IEEE 802.15.3c faded channel. For analysis, we set $N^H = N^L = 60$, $T_c = 1$ ns, $M = 4$, $T_p = 0.5$ ns, $T_s = 0.5$ ns, sampling time $T_{samp} = 0.01$ ns and $\rho_r = -1$. The carrier frequency $f_c = 60$ GHz and the roll-off factor of the raised-cosine pulse is 0.8. The HPBW at the Tx is 360°, 360°, 30°, 30° while at the Rx it is 15°, 15°, 30°, 30° for CM1.1, CM2.1, CM3.1 and CM8.1 respectively. For simplicity, we have considered the statistics of AOD same as that of the AOA.

Fig. 3 shows the effect of L on the BER of SISO-mm-wave communication system, over various IEEE 802.15.3c channel models, at 5 dB SNR. It is observed that for different channel models, the BER becomes flat after certain values of L. These values are $L \geq 27$, $L \geq 27$, $L \geq 10$ and $L \geq 10$ for CM1.1, CM2.1, CM3.1 and CM 8.1 respectively. From this observation it is inferred that for CM1.1, the rays which carry most of the energy fall within first 27 ns, whereas for CM3.1, they fall within 10 ns. Therefore, for further simulations, we choose $L = 27$, for all four channel models so that the performance of all these four cases can be fairly compared.

Fig. 4 illustrates the simulated and analytical BER plots of the mm-wave SISO communication system for different environments of the IEEE 802.15.3c channel model. Clearly, the proposed BER (24) is in good agreement with the corresponding simulated plot, which indeed validates our analysis. It should be noted that there are some marginal differences between the analytical and the simulated BERs which are caused by: (a) the use of numerical integration rules like Gauss–Hermite and Gauss–Legendre quadrature and (b) while calculating the energy at the Rake receiver, the error due to the approximations used ($\varepsilon_1 \approx \tilde{\varepsilon}_1$ and $\varepsilon_m \approx \tilde{\varepsilon}_m$). From (24), it is clear that the aforementioned errors increase with increase in the size of the antenna array. For better understanding regarding the accuracy of the approximation, the approximate error plots (difference between the analytical and the simulated BERs) for different environments of IEEE 802.15.3c channel model are also shown in Fig. 5.

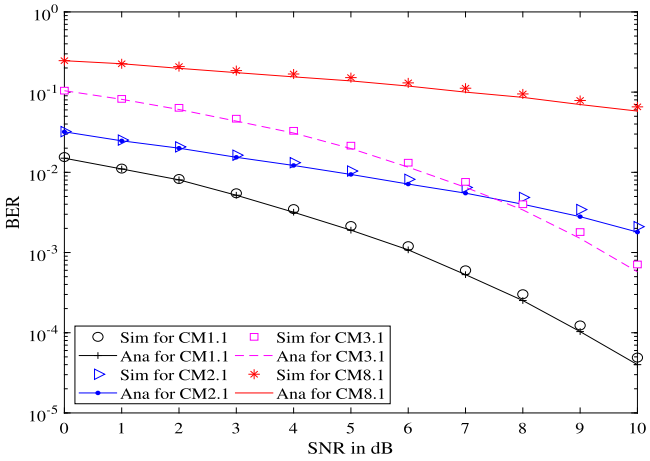


Fig. 4. BER (simulation and analytical) of mm-wave SISO communication system (Ana $--> $ Analytical and Sim $--> $ Simulation).

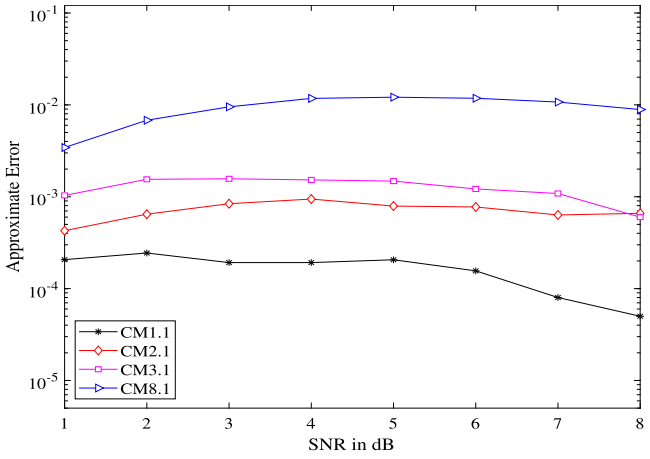


Fig. 5. Approximate error vs SNR: for mm-wave SISO system.

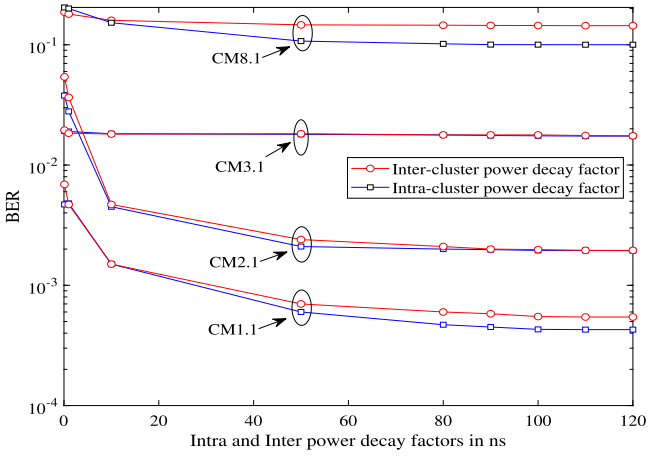


Fig. 6. BER vs $\{\Gamma$ & $\gamma\}$ in mm-wave SISO communication system at 5 dB SNR.

Fig. 6 shows the effect of power decaying factors on the BER of mm-wave SISO communication system, over various environments of the IEEE 802.15.3c channel model. It can be seen that initially the BER decreases with increase in Γ or γ, but after certain values of the power decay factor, the BER does not show any improvement. This phenomenon can be explained as follows:

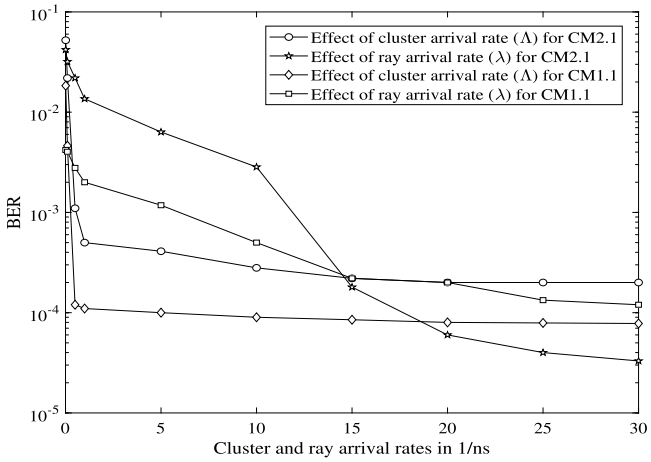


Fig. 7. BER vs $\{\Lambda$ & $\lambda\}$ in mm-wave SISO communication system at 5 dB SNR.

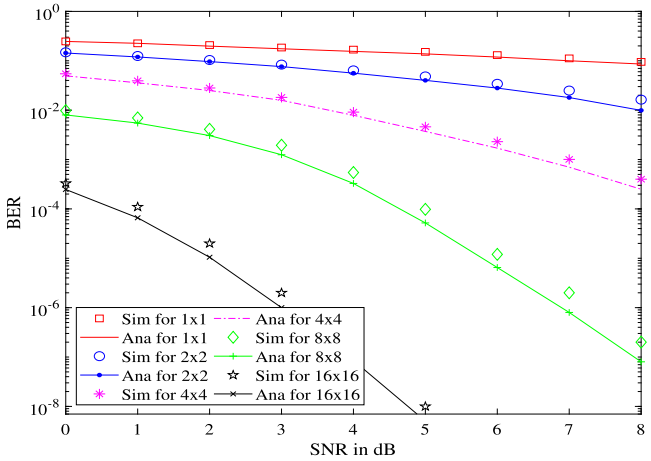


Fig. 8. BER (simulation and analytical) for $N_t \times N_r$ mm-wave MIMO system over IEEE 802.15.3c CM8.1.

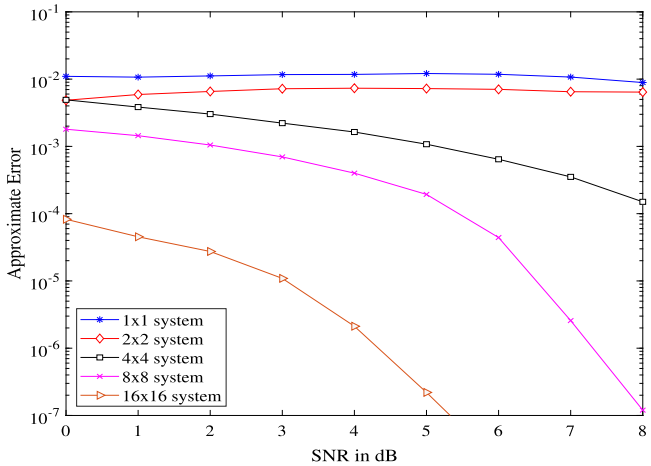


Fig. 9. Approximate error vs SNR: for $N_t \times N_r$ mm-wave MIMO system over IEEE 802.15.3c CM8.1.

From (6), it is clear that the energy carried by the MPCs increases with Γ or γ. As a result, the total energy captured by the Rake receiver (ε) also increases, which reduces the BER. However, after certain point of time, ε remains unchanged as most of the

energy carried by the MPCs has already been captured by the Rake receiver. Therefore, BER curves become flat.

The similar trends can be observed through Fig. 7, which shows the effect of the arrival rates on the BER. This is due to the fact that as Λ or λ increases, the arrival times of MPCs decrease and therefore more MPCs arrive in $[0,\ LT_c]$ time window which increases ε. But for large Λ or λ, most of the energy carrying MPCs arrive in $[0,\ LT_c]$ time interval, which causes the BER curves to get flattened. The results also indicate that the BER curves are less variant for a change in λ and reach the lower saturation value slower than the saturation value obtained while changing Λ. This is because the ray arrival rate influences the rays which are within the cluster whereas cluster arrival rate influences all the rays of the cluster.

For different antenna array configurations, Fig. 8 compares the simulated and the analytical BER for mm-wave MIMO system over CM8.1 channel model. It is noticed that as the size of the antenna array increases, the BER of the investigated system improves; however, the normalized difference between the simulated and the analytical BERs also increases with increase in the antenna array size. This is because the derived BER expression (24) is the function of the number of antennas used i.e. (N_r, N_t). As we increase the number of the transmitting and receiving antennas, the normalized error introduced between the simulated and the analytical BER is added up. Thus, an increase in the number of antennas, increases the error introduced by the analysis. Fig. 9 further gives an insight of this approximated error MIMO-mm-Wave system for CM8.1.

5. Conclusion

This paper derived a computable formula for the BER of binary signals incorporating analog beamforming aided mm-wave MIMO system using Gauss Quadratures rules of numerical integration. In the beamforming architecture, we have considered Tx-ULAs and Rx-ULAs to address the high path loss scenario experienced at 60 GHz. Apart from the analytical results of the BER, the simulations have also been carried out for different antenna arrays over the different environments of IEEE 802.15.3c channel model. From the results, we found that the analytical BER is in excellent agreement with the corresponding simulated results, which validates the suitability of the analysis. We have also considered the impact of the Rake receivers which are employed at the receiver, and suggested the sufficient number of Rake fingers required to capture the maximum energy of the mm-wave signal in various environments of the IEEE 802.15.3c channel model, at 60 GHz. In addition, we have also discussed the impact of various parameters of IEEE 80.15.3c channel on the performance of the proposed system. The presented analysis can be extended for Non-Orthogonal Multiple Access [40,41].

Acknowledgments

Authors would like to thank Prof. Rakhesh singh Kshetrimayum from Indian Institute of Technology Guwahti, India and Prof. Chi Chao-Chao from National Tsing Hua University in Hsinchu city, Taiwan for their valuable remarks and guidance.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

The complex multipath coefficient $(\alpha_{r,m})$ is written as

$$\alpha_{r,m} = \sqrt{\Omega_0}\, e^{\left(j\theta - \frac{\tilde{\kappa}}{2} I_{[\tilde{\kappa}\neq 0]}\right)}\, \xi_m\, \varrho_{r,m} \tag{A.1}$$

where θ is the uniform random variable over the interval $[0, 2\pi]$, parameters ξ_m, $\varrho_{r,m}$ are the amplitude coefficients of the mth cluster and the rth ray within the mth cluster, respectively and are distributed as independent lognormal random variables, i.e., $\xi_m \sim \ln\mathcal{N}\left(-\frac{T_m}{2\Gamma} - \sigma_1^2, \sigma_1^2\right)$ and $\varrho_{r,m} \sim \ln\mathcal{N}\left(-\frac{\tau_{r,m}}{2\gamma} - \sigma_2^2, \sigma_2^2\right)$. Parameters σ_1, σ_2 are the lognormal standard deviations of the cluster and ray respectively. Due to the random phase, the first order moment of complex MPCs is zero, i.e. $\mathrm{E}\left[\alpha_{r,m}\right] \approx 0$. The second order moment of complex MPCs can be computed as

$$\begin{aligned}\mathrm{E}\left[\left|\alpha_{r,m}\right|^2\right] &= \Omega_0 e^{-\tilde{\kappa} I_{[\tilde{\kappa}\neq 0]}} \mathrm{E}\left[\xi_m^2\right] \mathrm{E}\left[\varrho_{r,m}^2\right] \\ &= \Omega_0 e^{-\tilde{\kappa} I_{[\tilde{\kappa}\neq 0]}} e^{2\left[\left(\frac{-T_m}{2\Gamma} - \sigma_1^2\right) + \sigma_1^2\right]} e^{2\left[\left(\frac{-\tau_{r,m}}{2\gamma} - \sigma_2^2\right) + \sigma_2^2\right]} \\ &= \Omega_0 e^{\left(-\frac{T_m}{\Gamma} - \frac{\tau_{r,m}}{\gamma} - \tilde{\kappa} I_{[\tilde{\kappa}\neq 0]}\right)}\end{aligned} \tag{A.2}$$

If $X \sim \ln\mathcal{N}\left(\mu, \sigma^2\right)$, then the second order moment in the log scale is given as

$$\mathrm{E}\left[X^2\right] = e^{\left(2\left(\frac{\ln 10}{20}\right)\mu + 2\left(\frac{\ln 10}{20}\right)^2 \sigma^2\right)} \tag{A.3}$$

From (A.2) and (A.3), we have

$$\mu = \frac{10}{\ln 10}\left[\ln \Omega_0 - \frac{T_m}{\Gamma} - \frac{\tau_{r,m}}{\gamma} - \tilde{\kappa} I_{[\tilde{\kappa}\neq 0]} - \left(\frac{\ln 10}{10}\right)^2 \frac{\sigma^2}{2}\right] \tag{A.4}$$

Now, in order to fit in our formula, we just substitute T_m by T and $\tau_{r,m}$ by $(T - t)$ in (A.4), which yields

$$\mu = \frac{10}{\ln 10}\left[\ln \Omega_0 - \frac{T}{\Gamma} - \frac{T - t}{\gamma} - \tilde{\kappa} I_{[\tilde{\kappa}\neq 0]} - \left(\frac{\ln 10}{10}\right)^2 \frac{\sigma^2}{2}\right] \tag{A.5}$$

References

[1] P. Zhouyue, K. Farooq, An introduction to millimeter-wave mobile broadband systems, IEEE Commun. Mag. 49 (6) (2011) 101–107.
[2] B. Tuncer, S.C. Sean, L. Zhou, W. Junyi, R.M. Azizur, H. Hiroshi, K. Shuzo, IEEE 802.15. 3c: the first IEEE wireless standard for data rates over 1 Gb/s, IEEE Commun. Mag. 49 (7) (2011) 114–124.
[3] Y.S. Khiong, X. Pengfei, V.G. Alberto, 60 GHz Technology for Gbps WLAN and WPAN: From Theory to Practice, John Wiley & Sons, 2011.
[4] P. Cheolhee, R.S. Theordore, Short-range wireless communications for next-generation networks: UWB, 60 GHz millimeter-wave WPAN, and ZigBee, IEEE Wirel. Commun. 14 (4) (2007) 70–78.
[5] Y.S. Khiong, TG3c Channel Modeling Sub-Committee Final Report, IEEE802. 15-06-0195-07-003c, 2006.
[6] S. Harkirat, O. Jisung, K. Changyeul, Q. Xiangping, S.H. Rong, N. Chiu, A 60 GHz wireless network for enabling uncompressed video communication, IEEE Commun. Mag. 46 (12) (2008) 71–78.
[7] S.K. Mathew, R.S. Theodore, 3-D statistical channel model for millimeter-wave outdoor mobile broadband communications, in: Proc. IEEE Int. Conf. Commun., ICC, 2015, pp. 2430–2436.
[8] H. Sooyoung, K. Taejoon, L. David, K.V. James, A.T. Timothy, G. Amitava, Millimeter wave beamforming for wireless backhaul and access in small cell networks, IEEE Trans. Commun. 61 (10) (2013) 4391–4403.
[9] P. Liu, T. Jiang, Q. Zhang, M. Matthaiou, Pilot power allocation through user grouping in multi-cell massive MIMO systems, IEEE Trans. Commun. 65 (4) (2017) 1561–1574.
[10] H.Q. Ngo, L.-N. Tran, T. Duong, M. Matthaiou, E.G. Larsson, On the total energy efficiency of cell-free massive MIMO, IEEE Trans. Green Commun. Netw. 2 (1) (2018) 25–39.
[11] E. Bjornson, M. Matthaiou, M. Debbah, Massive MIMO with non-ideal arbitrary arrays: Hardware scaling laws and circuit-aware design, IEEE Trans. Wirel. Commun. 14 (8) (2015) 4353–4368.
[12] R. Wonil, S.J. Yun, P. Jeongho, L. Byunghwan, L. Jaekon, K. Yungsoo, C. Jaeweon, C. Kyungwhoon, A. Farshid, Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results, IEEE Commun. Mag. 52 (2) (2014) 106–113.

[13] R. Kilian, H. Pirzadeh, A.L. Swindlehurst, J.A. Nossek, A comparison of hybrid beamforming and digital beamforming with low-resolution ADCs for multiple users and imperfect CSI, IEEE J. Sel. Top. Sig. Process. 12 (3) (2018) 484–498.
[14] F. Sohrabi, W. Yu, Hybrid analog and digital beamforming for mmWave OFDM large-scale antenna arrays, IEEE J. Sel. Areas Commun. 35 (7) (2017) 1432–1443.
[15] R. Rakshith, H. Lajos, Hybrid beamforming in mm-wave MIMO systems having a finite input alphabet, IEEE Trans. Commun. 64 (8) (2016) 3337–3349.
[16] L. Peng, B. Jiri, P.N. Stefan, D.R. Marco, S. Andreas, Performance of generalized spatial modulation MIMO over measured 60 GHz indoor channels, IEEE Trans. Commun. 66 (1) (2018) 133–148.
[17] H. Longzhuang, W. Jintao, S. Jian, Spatial modulation for more spatial multiplexing: RF-chain-limited generalized spatial modulation aided MM-wave MIMO with hybrid precoding, IEEE Trans. Commun. 66 (3) (2018) 986–998.
[18] E.A. Omar, R. Sridhar, A.S. Shadi, P. Zhouyue, H.W. Robert, Spatially sparse precoding in millimeter wave MIMO systems, IEEE Trans. Wirel. Commun. 13 (3) (2014) 1499–1513.
[19] Y. Candy, S. Suresh, Empirical capacity of mmWave WLANs, IEEE J. Sel. Area Commun. 27 (8) (2009).
[20] X. Zhu, A. Doufexi, T. Kocak, On the performance of IEEE 802.15. 3c millimeter-wave WPANs: PHY and MAC, in: Proc. 6th IEEE Conf. on Wireless Advanced, WiAD, 2010, pp. 1–6.
[21] T.V. Nguyen, M. Elias, M.B. Laurence, Channel model and performance analysis of QAM multiple antenna systems at 60 GHz in the presence of human activity, in: Proc. IEEE Global Telecommun. Conf., GLOBECOM, 2011, pp. 1–6.
[22] Z. Liang, O. Yoji, Performance analysis of mmWave LOS-MIMO systems with uniform circular arrays, in: Proc. IEEE 81st Veh. Tech. Conf., VTC Spring, 2015, pp. 1–5.
[23] Y. Yavuz, S. Predrag, Reducing the LOS ray beamforming setup time for IEEE 802.11 ad and IEEE 802.15. 3c, in: Proc. IEEE Military Commun. Conf., MILCOM, 2016, pp. 448–453.
[24] N. Preeti, Analysis of Multiple Antenna Ultra-Wideband and Millimeter Wave Communication Systems (Ph.D. thesis), UC San Diego, 2007.
[25] C. Liu, S. Efstratios, P.S. Tony, E.J. Robin, Angle of arrival extended SV model for the 60 GHz wireless desktop channel, in: Proc. 17th IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun., 2006, pp. 1–6.
[26] G. Carl, H. Katsuyuki, W. Shurjeel, T. Fredrik, On mm-wave multipath clustering and channel modeling, IEEE Trans. Antennas and Propagation 62 (3) (2014) 1445–1455.
[27] H. Yueh, T. Hua, W.D. Wei, C.C. Chao, Performance analysis of OFDM systems over 60 GHz indoor Channels, in: Proc. IEEE Veh. Tech. Conf., VTC Fall, 2012, pp. 1–5.
[28] G.H. Gene, W.H. John, Calculation of Gauss quadrature rules, Math. Comp. 23 (106) (1969) 221–230.
[29] M. Matthaiou, P. de Kerret, G.K. Karagiannidis, J.A. Nossek, Mutual information staistics and beamforming analysis of optimized LOS MIMO systems, IEEE Trans. Commun. 58 (11) (2010) 3316–3329.
[30] A.M. Saleh, V. Reinaldo, A statistical model for indoor multipath propagation, IEEE J. Sel. Area Commun. 5 (2) (1987) 128–137.
[31] S.H. Quentin, J.D. Brian, J.A. Michael, S.A. Lee, Modeling the statistical time and angle of arrival characteristics of an indoor multipath channel, IEEE J. Sel. Area Commun. 18 (3) (2000) 347–360.
[32] S. Hirokazu, S. Yozo, C.C. Soon, Proposal of novel statistic channel model for millimeter wave WPAN, in: Proc. Asia-Pacific Microwave Conf., APMC, 2006, pp. 1855–1858.
[33] Gilb, P.K. James, IEEE Standards 802.15. 3c–Part 15.3: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for High Rate Wireless Personal Area Networks (WPANs) Amendment 2: Millimeter-Wave-Based Alternative Physical Layer Extension [S], IEEE Computer Society, New York, 2009.
[34] Y. Seokhyun, J. Taehyun, L. Wooyong, Hybrid beam-forming and beam-switching for OFDM based wireless personal area networks, IEEE J. Sel. Area Commun. 27 (8) (2009) 1425–1432.
[35] S.K. Marvin, A.S. Mohamed, Digital Communication Over Fading Channels, Vol. 95, John Wiley & Sons, 2005.
[36] W.L. Chun, L.W. Cheng, Bit error rate analysis in IEEE 802.15. 3a UWB channels, IEEE Trans. Wirel. commun. 9 (5) (2010) 1537–1542.
[37] B. Zhao, C. Yunfei, G.J. Roger, Average SINR analysis of DS-BPSK UWB systems with IPI, ICI, ISI, and MAI and its application, IEEE Trans. Veh. Technol. 58 (8) (2009) 4690–4696.
[38] A. Agrawal, R.S. Kshetrimayum, Analytical BER calculation of TAS/MRC-based two hop UWB communication system over IEEE 802.15.4a channel, IETE Tech. Rev. 35 (5) (2018) 467–475.
[39] H. Kei, G.A. John, The distribution of sums of path gains in the IEEE 802.15. 3a UWB channel model, IEEE Trans. Wirel. Commun. 6 (3) (2007) 811–816.
[40] X. Zhenyu, Z. Lipeng, C. Jinho, X. Pengfei, X. Xiang-Gen, Joint power allocation and beamforming for non-orthogonal multiple access (NOMA) in 5G millimeter wave communications, IEEE Trans. Wireless Commun. 17 (5) (2018) 2961–2974.
[41] Z. Lipeng, Z. Jun, X. Zhenyu, C. Xianbin, W.D. Oliver, X. Xiang-Gen, Millimeter-wave NOMA with user grouping, power allocation and hybrid beamforming, IEEE Trans. Wireless Commun. 18 (11) (2019) 5065–5079.



Dr. Anand Agrawal received his B.Tech degree from Samrat Ashok Technological Institute, Vidisha, India, in 2008, M.Tech degree from National Institute of Technology Bhopal, India, in 2010 and Ph.D from Indian Institute of Technology Guwahati, India in 2018. Since June 2019, he is working as an Assistant Professor in the department of Electronics and Communication Engineering at Indian Institute of Information Technology Kota, India. From Jan. 2018 to May 2019, he worked as an Assistant Professor in the department of Electronics and Communication Engineering at Jaypee Institute of Information Technology Noida , India. During PhD, he was a visiting research scholar at National Tsing Hua University (NTHU), Taiwan from Sept 2015 to Jan 2016. He has been involved in organizing several national/ international conferences in various capacities. He is a member of Journals Pub editorial board of, International Journal of Radio Frequency Design, International Journal of Broadband Cellular Communication and International Journal of Telecommunications & Emerging Technologies. He is also a reviewer of various reputed national and international journals including, Physical Communication, Wireless Personal Communication, Transaction of emerging Telecommunication technologies, IEEE communication letter and so on. His current area of research interest are UWB Communication, mmWave Communication at 60 GHz, Cooperative Communication, IEEE802.15 Standards, 5G Technology, Massive MIMO and D2D Communication. technology.



Dr. Vimal Kumar Mishra received his B.Tech. Degree in Electronics and Communication Engineering from Uttar Pradesh Technical University, Lucknow, India in 2009, M.Tech. Degree in Microelectronics and Embedded Technology from the Jaypee Institute of Information Technology, Noida, India in 2012, and Ph.D. in VLSI Design from Madan Mohan Malaviya University of Technology, Gorakhpur, India in 2017. Currently, he is an Assistant Professor at Jaypee Institute of Information Technology, Noida, India. His research interests include semiconductor physics and device modeling. He has filed 2 patents and published more than 25 papers in the reputed journals and conferences.



Dr. Ashish Gupta was born on 4th August 1986. Currently, he is an Assistant Professor in Jaypee Institute of Information Technology, Sector 128, Noida. He has done his Ph.D. in RF and Microwave Engineering from Department of Electronics Engineering, Indian Institute of Technology (Indian School of Mines) Dhanbad in 2017. He did his MTech from ShriGovindramSeksaria Institute of Technology and Science, Indore in 2010, where he got fellowship from MHRD Govt. of India. He won the best paper award in ICAICV-2010 at Coimbatore. He has done his B.Tech. from Rustumji Institute of Technology, B.S.F. Academy, Tekanpur in year 2008. From year 2010 to 2013, he was associated with different engineering institutions as an Assistant Professor, in Gujarat, India. He is the author of more than 15 articles in the relevant area. His current research interests involve Metamaterials and its applications in active and passive devices. He is also serving as reviewer in few peer-reviewed journals of repute.



Dr. Bajrang Bansal received his B.E. Degree in Electronics and Communication Engineering from Institute of Technology and Management, Gurgaon, India in 2005, M.Tech. Degree in VLSI Design and CAD from the Thapar University, Punjab, India in 2008, and Ph.D. in Wireless Communication from Delhi Technological University, Delhi, India in 2017. Currently, he is an Assistant Professor at Jaypee Institute of Information Technology, Noida, U.P, India. His research interests include wireless channel modeling and UWB.

International Journal of Progressive Research in Science and Engineering
Volume-1, Issue-3, June-2020
www.ijprse.com

Performance of Masonry Infill Framed Structures During Earthquake: A State-Of-The-Art Review

Priyanka[1], Alok Verma[2], Shobha Ram[3]

[1]*Student, Department of Civil Engineering, Gautam Buddha University, India.*

[2]*Professor, Department of Civil Engineering, Delhi Technological University, India.*

[3]*Professor, Department of Civil Engineering, Gautam Buddha University, India.*

Corresponding Author: priyankap780@gmail.com

Abstract: - Masonry infill walls are widely known to increase lateral stiffness of the structure and for this reason it is accepted all over the world. This paper presents a review work on performance of infill framed structures that were damaged during several earthquakes. Study of behavior of damaged building during different earthquakes in world has been carried out. The mentioned earthquakes substantially caused damaged to the RC buildings. The RC buildings were damaged primarily because of improper design and reinforcement detailing at the design phase and improper workmanship and quality control at the construction phase. The main objective of this paper is to describe and analyze the failure patterns observed in reinforced concrete frame buildings with masonry infill walls and without masonry infill walls all over the world.

Key Words: — ***Infill Walls, Earthquake, Non Structural Elements, Failure Mechanism.***

I. INTRODUCTION

The present paper is concerned with the performance of buildings all over the world primarily due to ground shaking. From the past researches and field investigations by various researches it can be noted that building with masonry infill walls (or unreinforced masonry infill walls) perform better in earthquake. As compared to bare RC frame masonry infill frames can carry more seismic effect. From study it can be concluded that masonry infill walls or panels play very crucial role with respect to severe ground motions. Many structural deficiencies are highlighted here which were caused by earthquake, which consists of soft and weak stories, strong-beam and weak column, etc.

This paper mentions five great earthquakes that struck the world. Among these, two are Bhuj (2001) and Sumatra (2006) earthquake which happened in India. Bhuj earthquake is the largest earthquake that struck India. The remaining earthquakes are being mentioned here, Wenchuan Earthquake (2008), L'Aquila (2009) and Lorca (2011) earthquake.

II. LITERATURE REVIEW

Humar J. M. et al. (2001) and Sudhir K. Jain et al. (2001) reported on the performance of building during Bhuj earthquake that were damage during the earthquake. They both individually investigated that the open storey were severely damage as compare to that of infill storey. Hemant B. Kaushik and Sudhir K. Jain et al. (2004) reported that the great Sumatra earthquake of 26 December 2004 caused damages to RC buildings at Port Blair in the Andaman Islands in India. Li, B., Wang, Z., Mosalam, K. M., & Xie, H. et al. (2008, October) carried out field reconnaissance after Wenchuan earthquake in Dujiangyan, Hanwang, and Yingxiu, to investigate the types of damage occurred in the infill walls and/or the bounding frames. Manfredi and masi reported (Manfredi, V., & Masi, A. et al., (2011, 2014)) on the key role of the non-structural element in building during L' Aquila 2009 earthquake. They put forward an infill model capable of accounting for the effects arising from in-plane (IP) damage on the out-of –plane (OOP) capacity of infill panels. Hermanns, L., Fraile, A., Alarcón, E., & Álvarez, R. et al., (2014) analyzed the failure patterns observed in reinforced concrete frame buildings with masonry infill walls.

III. AN OVERVIEW

A. Bhuj Earthquake:

It is seen that most of the modern RC buildings that were collapsed in the damaged area were open at ground storey for the purpose of parking. The buildings which had few or no infill walls that were damaged are shown in figure 1(a). Open storey led to insufficient strength and stiffness of the structure resulting in failure of structure as shown in figure 1(b). Whereas it was observed from the study that the buildings with full scale infill walls at ground storey withstand earthquake without collapse.

IJPRSE
Progressive Research
International Journal of Progressive Research in Science and Engineering
Volume-1, Issue-3, June-2020
www.ijprse.com

Earthquake Name	Time And Date	Location	Magnitude
Bhuj Earthquake	8:46 Am, January 26, 2001	Bhuj, Gujarat	M7.9
Sumatra Earthquake	6:28:53 Am, December 26, 2004	Port Blair, Andaman Islands	M9
Wenchuan Earthquake	2:28 Pm, May 12, 2008	Northwestern Sichuan, China	M8
L' Aquila Earthquake	3:32:52 (Local Time) , April 6, 2009	Abruzzo, Italy	M6.3
Lorca Earthquake	6:47pm(Local Time), May 11, 2011	Lorca, South-East Spain	M5.1



Fig.1(a). Precariously balance severely damaged RC frame buildings with open ground storey in Ahmadabad



Fig.1(b). Ground storey collapse of a four storey building with open ground storey at Bhuj

B. Sumatra Earthquake:

In the great Sumatra earthquake many damaged pattern were observed due to infill panel among these the common damages were open storey damage, OOP failure of masonry infill walls.

The buildings that were constructed without engineering supervision and were located in silts were collapsed partially or completely. In a collapsed 3-storey building at Naya Gaon, the upper stories suffered only nominal damages in the RC frame and masonry infill (Fig. 2a). Some recently constructed buildings on stilts at Bamboo Flat collapsed completely (Fig. 2b). It was also noticed that a 2-storey Police-Barrack Government owned building, which had the open first storey for storage purpose collapsed at Haddo wharf (Fig. 2c), whereas the adjacent infill buildings in the first storey survived the ground motions. The passenger terminal building at Haddo Wharf (Fig. 2d), was constructed under the supervision of a construction firm gets partially collapsed as it has open masonry infill panels for the purpose of ventilation and in the inner periphery infill walls were not present for the purpose of open space. Due to poor masonry quality and loose frame joints OOP failure was observed.

Figure 3 shows OOP failure of masonry infills in buildings at (a) Mohanpura (school building), (b) Bamboo flat, and (c) Passenger Terminal Building at Haddo wharf.

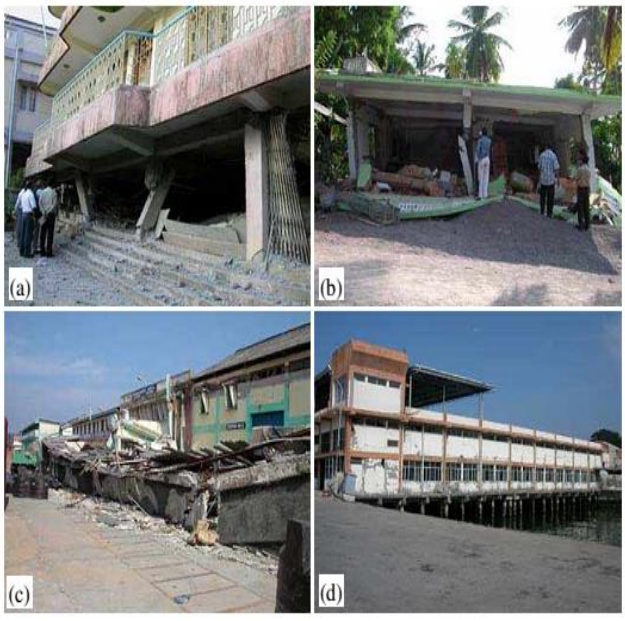


Fig.2. Collapse of RC buildings at (a) Naya Gaon, (b) Bamboo Flat, and (c) Haddo wharf, and (d) partially collapsed Passenger Terminal Building at Haddo wharf.

IJPRSE
Progressive Research
International Journal of Progressive Research in Science and Engineering
Volume-1, Issue-3, June-2020
www.ijprse.com



Fig.3. shows OOP failure of masonry infills in buildings at (a) Mohanpura (school building), (b) Bamboo flat, and (c) Passenger Terminal Building at Haddo wharf.

C. Wenchuan Earthquake:

It was observed that many residential and commercial buildings in the epicenter suffered severe damage due to soft storey effect of infill walls. Many buildings had open storey which serve as the purpose of parking. Many buildings had less or no infill walls. Many buildings were made using hollow clay tiles and gas concrete masonry infill walls. It was observed that damage to masonry infill walls was concentrated in the lower stories of buildings where less or no infill walls were provided. It can be noticed from the results that hollow infill walls contribute to stiffness but less strength. Arrangement of infill panels in orthogonal direction to create residential space resulted in creating stiffness discontinuities. Figure 4 shows a six-storey building where the first storey was available for the purpose of parking and had less infill walls while the upper stories are residential and has many infill walls.



Fig.4. Formation of weak storey: leaned 6-story building in Dujiangyan

D. L' Aquila Earthquake:

During past earthquakes in Italy these non-structural elements have contributed significantly in resisting lateral loads. Figure 5 shows example of masonry panel expulsions from the structural frame, resulting from a poor connection.



Fig.5. Examples of out-of-plane collapse of infills poorly connected to the RC frame

E. Lorca Earthquake:

Figure 6 shows one of the parapet walls that collapsed during lorca earthquake which shows the importance of studying the behaviour of non-structural components. It can be noted that acceleration at roof level is significantly higher than that at ground level. Figure 7 shows soft storey mechanisms due to infill panels with lower stiffness at ground storey.



Figure.6. Building with a partially collapsed parapet wall



Fig.7(a). Building with soft storey



Fig.7(b). Severely damaged column

IV. CONCLUSION

From the above study it can be contributed that masonry infill walls contributes significantly in increasing the lateral stiffness of the structure and increasing the strength of the structure. Also it can be noticed that connection between frames and infill walls play important role in failure pattern. No infill panel or less infill panel leads to soft storey mechanism in which flexibility of the structure is reduced. Hence from the observation it can be mentioned that it is important to provide good quality masonry infill panels under engineer's supervision.

REFERENCES

[1]. Li, B., Wang, Z., Mosalam, K. M., & Xie, H. (2008, October). Wenchuan earthquake field reconnaissance on reinforced concrete framed buildings with and without masonry infill walls. In The 14th World Conference on Earthquake Engineering. Beijing, China.

[2]. Braga, F., Manfredi, V., Masi, A., Salvatori, A., & Vona, M. (2011). Performance of non-structural elements in RC buildings during the L'Aquila, 2009 earthquake. Bulletin of Earthquake Engineering, 9(1), 307-324.

[3]. Hermanns, L., Fraile, A., Alarcón, E., & Álvarez, R. (2014). Performance of buildings with masonry infill walls during the 2011 Lorca earthquake. Bulletin of Earthquake Engineering, 12(5), 1977-1997.

[4]. Murty, C. V. R., Dayal, U., Arlekar, J. N., Chaubey, S. K., & Jain, S. K. A National Technological Crisis: Structural and Geotechnical damages sustained during M7. 9 Bhuj earthquake force. Department of Civil Engineering, IIT Kanpur.

[5]. Kaushik, H. B., & Jain, S. K. Performance of Buildings In Port Blair (India) During The Great Sumatra Earthquake Of 26 December 2004.

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

PMS-Sharing : Framework for Automatically Authenticating users in a Group to Allow Sharing Storage

Nitesh Suthar
System R&D – Convergence Group
Samsung Electronics
Noida, India
s.nitesh@samsung.com,
suthar.nitesh@gmail.com

Rajesh Kr. Yadav
Dept. of Computer Science & Engineering
Delhi Technological University
Delhi, India
rkyadav@dtu.ac.in

***Abstract*—Cloud storage is a very famous buzz word nowadays. There are many cloud storage service providers in the market and millions of mobile users are accessing cloud storage to store and access various personal data such as photos, videos and documents. In this era of smart mobile devices, every user is having devices with various storage capacity i.e. 8, 16, 32, 64, 128 GB and even more storage is going to be embedded in next generation of devices. These storage are widely used for storing and sharing data among multiple people. In this paper, a model and framework is proposed to automatically authenticate users within a group to allow sharing data while connected to same wireless network where initial group is formed by user (group owner) to allow set of users to access shared contents thereafter users in a group are automatically authenticated. The proposed solution is called personal mobile storage sharing (PMS-Sharing). The application and framework is developed and tested on android mobile phones supporting Android version 4.4 and above.**

Keywords—shared storage; network; mobile computing; cloud storage; IoT

I. INTRODUCTION

In this era of smart mobile devices, millions of users are using smartphones and tablets; these smartphones are equipped with high-performance processors, RAM for computing, gigabytes of inbuilt storage memory and additional SD card slot to extend the storage capacity. Substantial growth of smartphone devices also increases the need of Cloud Storage, with the roll-out of 4G and 5G technology, it is assumed that there will be no issue of connectivity and smartphone market is also increasing day by day. Many companies are engaged in developing new smartphones with new and improved capabilities with reasonable price. This motivates smartphone users to upgrade their smartphone to the advanced one.

PMS-Sharing provides a shared space for a group of users to share files (documents, images, videos etc.) among each other while connected to the same network. The Personal Mobile Storage Sharing (PMS-Sharing) allows user to form a local group of users and allows them to access shared storage directory. Currently if user needs to share any file, he has to do it manually by transferring file via Bluetooth or Wi-Fi even though all users are connected to the same local network (WLAN Network). PMS-Sharing overcomes this problem of manual transferring and authenticating. In PMS-Sharing, a framework is introduced for automatically authenticating users in a group to allow sharing storage. RSA encryption algorithm is systematically used along with randomly generated pair key. Pair key is unique for each user, helps Group Owner device to identify and ensures that an authenticate user is accessing the group's data. Public key cryptography based on Rivest–Shamir–Adleman (RSA), is used for encrypting messages to securely exchange pair key with designated user. It is using two level of encryption one at group level and another at user level.

In PMS-Sharing, the user can share the directory among all other users in a group and other user can download and view the files of their choice easily by clicking on file name at displayed in list of files. User can do multiple functions like add, delete and download view or play files. PMS-Sharing will introduce a new framework to utilize the unused mobile phones as storage nodes to build big storage space which can be deployed as a shared storage on distributed network. The application is developed and tested on Android mobile phones using existing infrastructure. In section II, Related work –cloud storage types, service providers, existing cloud storage, key challenges faced and limitations imposed by cloud storage system are discussed. In section III, Proposed Solution – the detailed architecture and framework of PMS-Sharing application are discussed, the algorithm for identification and authentication of user who belong to a particular group are also discussed. In section IV, Conclusion and Future Scope – the current challenges and future scope of this proposed work are discussed.

II. RELATED WORK

A. Cloud Storage System

Cloud Storage means storing data in a remote location which is accessible on any device. Cloud Storage improves efficiency and productivity in terms of backing up and securing

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:02:08 UTC from IEEE Xplore. Restrictions apply.

the data. It is beneficial and Business can pay only for the amount of storage they require [1]. There are different kinds of cloud storage system like a private, public, hybrid, many cloud storage providers provide service to store user data, and they charge pricing for these services.

Cloud Storage is categorized, based on how data is managed on servers- 1. File Storage, 2. Block Storage, and 3. Object storage. There are many types of File Storage with different characteristics like structure, access method, speed, security, and size. File storage is applied on local as well as network storage devices. Some prominent network file storage is Network File System (NFS), Server Message Block(SMB)/ Common Internet File System (CIFS) and Plan 9 File system protocol (9P)[2], [3]. Block Storage provides a way to store data in an evenly-sized memory blocks to host the database, which supports random read and write operations [3], [4]. Amazon Elastic Block Storage (EBS) is developed for use with Amazon Elastic Compute Cloud (EC2) to deploy a broad category of data loads such as databases, containerized applications, big data analytics engine, file system, media data, and enterprise application data [5]. Object Storage provides an easy way to store unstructured data in the form of Objects. An object is a simple repository with the unique ID number, metadata, and actual data [6]. Amazon Simple Storage Service (Amazon S3) is capable of managing a significant amount of unstructured data. Table I represents the basic differences between file storage, block storage, and object storage system.

B. Existing Cloud Storage Providers and Concept

IDrive [7], pCloud [8], Mega [9], OneDrive [10], iCloud [11], Google [12], Box [13], NextCloud [14], and SpiderOak [15] are some of the cloud storage service providers which provide few GB's of storage free of cost. Cloud storage demand is increasing in proportion to the number of mobile device users are increasing each day.

GrandStore [16] proposed a system where user can store the credential of all cloud storage accounts to expand the free storage capacity, they provide a way to manage their all free storage accounts using GrandStore application.

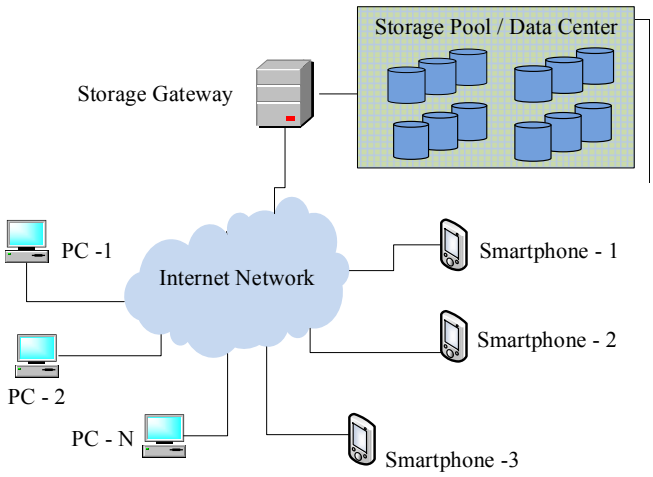


Fig. 1: Exsting Cloud Storage concept

TABLE I. DIFFERENCE BETWEEN FILE, BLOCK AND OBJECT STORAGE SYSTEM

Attribute	Storage System		
	File	***Block***	***Object***
Use Cases	File Sharing, Local Archiving, Data protection	Databases, Email Servers, RAID, Virtual Machines	Big Data, Web Applications, Backup Archives
Access	Full path should be known , Directory, sub directory , folder and file name	Provides Direct Access to the block	No direct access to object. An Object storage system places a file-system bridge [Gateways] in-between the two.
Metadata	No metadata, full path should be known.	No additional information, provides direct access to block	expandable and variable amount of metadata
Ease of Use	Simple to use and setup	Block Storage provide ability to incrementally edit any part of the file	Object storage is powerful and customizable because of variable amount of metadata of any type
Client Side Visibility	Visibility of files and folders is same on client and server	supports formatting of file systems like NTFS or SMB, NFS, VMFS (VMware) as need	Expose object storage as Network File System via Storage Gateways.
Recommendation	Recommended for shared storage	Recommended for dedicated storage.	Recommended for unstructured data storage

Chetan et al [17] proposed a framework using OpenStack [18] software to deploy storage as a service private cloud model that secure data using credential and Advanced Encryption Standard (AES) algorithm [19].

Adiseshu et al [20] proposed a solution to build the personal cloud using unused computers as nodes. Virtual Machine is setup on each unused computer node to build a datacenter-less, distributed cloud.

Alginahi et al [21] explained the various methods of authenticating to guarantee the trusted flow of data, they described the various challenges of authentication methods and usage of cryptographic algorithm in Cloud and IoT.

"Fig. 1" draws the basic representation of existing cloud network, where a smartphone participates as a client like other peer nodes such as personal computers, Tablets, and Laptops. Internet Network represent the Internet Service Provider and Backbone Network. There exists one Storage Gateway Server which basically redirects the user's data GET and PUT request to data center, then data center retrieve and stores the requested data.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:02:08 UTC from IEEE Xplore. Restrictions apply.

C. *Key Challanges with Cloud Storage*

Even though cloud is emerging very fast, there are various challenges related to security, ownership of data, capability, efficiency and cost. Table II describes these major challenges in using Cloud Storage. Our proposed solution provides solution to some of these challenges by empowering user to control their own local cloud and sharing.

TABLE II. CLOUD STROAGE CHALLANGES

Key Challanges	Brief Description
Compliance and Security	• Threats of data leakage • Unique Crendential • Security Threats during Data upload and download • No global policy to protect users cloud storage rights.
Ownership of data	• Only Cloud Storage service provider has full control on data. • No idea about how many copies of data are created by service provider for backup and restore.
Capability	• Rehousing of data is slow • Complex process to change service provider.
Efficency	• Downtime of service provider • Dependency on Network Speed.
Cost	• No pure private or personal cloud storage unless user has dedicated storage in his own premises which leads to high cost.

III. PROPOSED WORK

Our proposed system is called Personal Mobile Storage Sharing (PMS-Sharing). This systemremoves dependencies on cloud storage service providers to store the data on servers in order to send/share files/documents to any other user. In our proposed solution data is stored on the devices belonging to the known members of group. In our proposed way user's data are safe and reachable physically in case of network failure/unavailability.

In this era of wireless data transmission, Wi-Fi data throughput will increased far better with the deployment of 5G and 6G Wi-Fi routers. Wi-Fi Alliance [22] also introduced WiFi-6, next generation of highly efficient Wi-Fi network with theoretical speed of 11 Giga bits per second. These make proposed solution feasible to deploy on private as well as on personal local network. Unlike traditional cloud storage "Fig. 1", in proposed solution smartphones can participate in both ways as peer as well as hosting device as storage provider. "Fig. 2" draws the high level concept of the proposed framework.

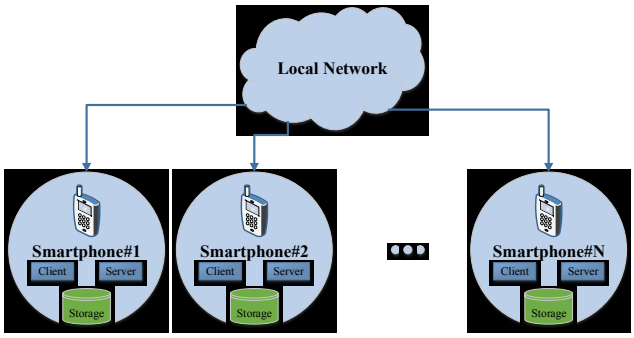


Fig. 2: PMS-Sharing Concept

Objective of PMS-Sharing is to provide a shared storage. PMS-Sharing achieves its objective by operating as both client and server mode. In case of unused smartphones the full storage can be utilized by PMS-Sharing and in case of smartphones devices which are in use, user can limit the size of storage to use as shared storage.

A. *Architecture Overview*

"Fig. 3" represents the high level architecture diagram of the PMS Sharing application. PMS Architecture has following main modules

1) *Login Activity*

Login Activity control the user interaction with Login Screen, Sign up Screen and Reset Password Screen. Login Activity is responsible for authenticating user to enter into the PMS Sharing application and use it. Login Activity is dependent on Account Manager.

2) *Dashboard Activity*

Dashboard Activity control the user interaction with Dashboard Screen, Create Group Screen, and Edit Group Screen. Dashboard Activity is mainly dependent on Group Manager.

3) *File Explorer Activity*

File Explorer Activity controls the user interaction with File Explorer Screens (Remote and Local Directory View). File Explorer is responsible for providing rich experience to end user while using, sharing and accessing the files which are made available by the File Repository

4) *Account Manager*

Account Manager provides services to the Login Activity to authenticate the user and granting Login Activity and the necessary permissions to enter into the PMS Sharing

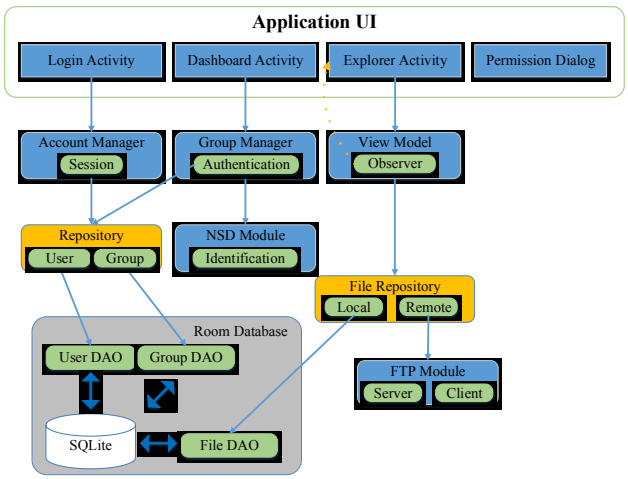


Fig. 3: Architecture PMS-Sharing Application

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:02:08 UTC from IEEE Xplore. Restrictions apply.

Application. Account Manager also manages the user session. Once the user is authenticated, he is granted the access for 12 hours or until the user logout manually.

5) Group Manager

Group Manager is at the core of PMS Sharing. Group Manager provides all the services that enable all the function supported by the Dashboard Activity. Group Manager controls function of Discovering Client, Creating Group, Editing the Group and showing group details.

6) NSD Module

Android provides the API to use Network Service Discovery (NSD). Using NSD PMS-Sharing application can easily identify and detect the users which are using PMS-Sharing services on the local Network. NSD Module help Group Manager to populate all the users available on the same Network and using PMS-Sharing services. User device will start broadcasting PMS-Sharing service on local network after registering the PMS-Sharing services, User device can also discover similar services broadcasted by other user on the same network. On discovering any service, NSD resolves the discovered services to get IP and Port of the peer discovered device. Then NSD can stop discovery any time by un-registering services.

7) File Repository

File repository provides a protection wall between actual storage media and the shared folder thus the remote user is capable to access the files and folder added into the shared storage directory path. File repository also keeps track of file added by any remote user.

id(primary_key)	user_id(string)	group_id(long)	file_name(String)	file_path(String)

Here user_id is the id of user who added the file/directory in the group folder, group_id is the id of the group in which this file is added, file_name is the name of file or directory. file_path is the complete local path of the directory in which file is kept. In case it is directory, it is the complete local path of the directory in which this directory is kept. File owner information helps to authenticate and validate the actual source of the user and in case of shared file it may store multiple user's information to provide access to multiple users.

B. Communication Interface

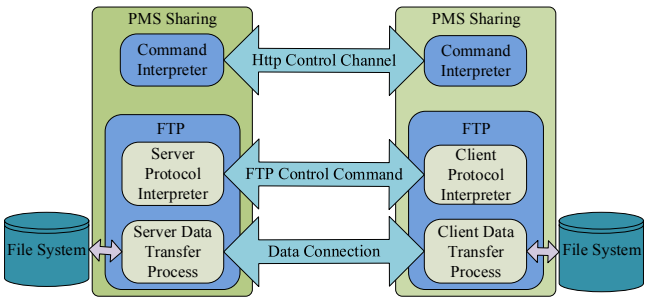


Fig. 4: High level representation of Communication Interface

PMS Sharing application uses FTP for Data Transfer Channel and HTTP for Control channel. "Fig. 4" depicts the communication interface between two PMS-Sharing applications, installed on two different devices belonging to different user. Command Interpreter is used to send and receive PMS-Sharing specific commands and control messages.

C. User, Group and File Repository

User repository is developed to manage the group management and user management. It also incorporated three entity tables:-

1) Entity Name - "NetworkGroupDetails"

In this entity each record will store detail about group id, group name, and total number of users in the group and user id of master user.

id(primary_key)	group_id(long)	group_name(String)	total_user(int)	master_user(string)

2) Entity Name - "NetworkGroupUserDetails"

In this entity each record will store information about each user of the group and its associated pair key to authenticate the user to access the group.

id(primary_key)	group_id(long)	user_id(string)	ip(string)	port(int)	pair_key(long)

3) Entity Name - "User"

In this entity each record will store information about the application's user.

id(primary_key)	user_name(string)	email(string)	password(string)	sec_q(string)	sec_a(string)

Where sec_q is security question and sec_a is security answer. Password, Security Question and Security Answer are encrypted before storing in database. It has used apaches salted password encryption.

D. Authentication Protocol

Authentication method is composed of multiple steps as Device Registration for service discovery, Identification of devices registered for same service. Complete process is discussed in below section.

1) Device Registration

Each device registering PMS-Service, will be visible to all other devices registering same PMS-Service. When any device gets connected to the network, then all other user will automatically detect the device using NSD service discovery. Default service name for registering in NSD Module is "pms-service".

Private static final String DEFAULT_SERVICE_NAME = "pms-service";

2) *Device Identification*

Using "*pms-service*" service name tag, that will be able to identify all the device in the network which are registered with same service name in NSD Module. Now next challenge is to identify the role of the device like which device is operating in Group Owner mode and which device is just participating as a

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:02:08 UTC from IEEE Xplore. Restrictions apply.

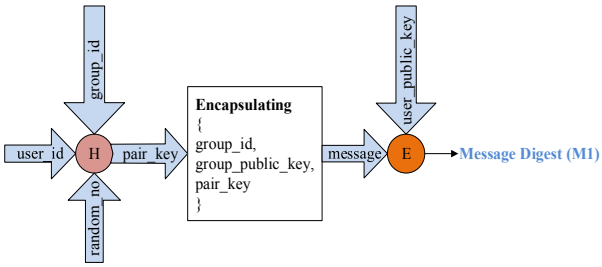


Fig. 5: Generation of Message Digest (M1)

user. Further, it extended the service registered to identify server and client. The device registering with server role is Group Owner and all other devices are normal users as shown in below code snippet.

```
Public String mServiceName = DEFAULT_SERVICE_NAME;

Public static enum TYPE {
   CLIENT,
   SERVER
}

Public void registerService(..., TYPE type) {
   ...
  If (type == TYPE.CLIENT)
      mServiceName = mServiceName + ".client";
  else if (type == TYPE.SERVER)
      mServiceName = mServiceName + ".server";
  ...
}
```

3) *User Identification and Authentication*

The next part is finding the ways to identify users who belong to a particular group and if there are multiple Group Owner in the network then how to identify the Group Owner in which particular user is added because user's IP and Port can change after each connect and disconnect from the Network. It consists of following steps:-

- At the time of group creation when Group Owner selects the users and click on create group, then Group Owner device starts user registration where he first connect with user 'X' device and notify about new group. On receiving notification for new group, User 'X' device will send

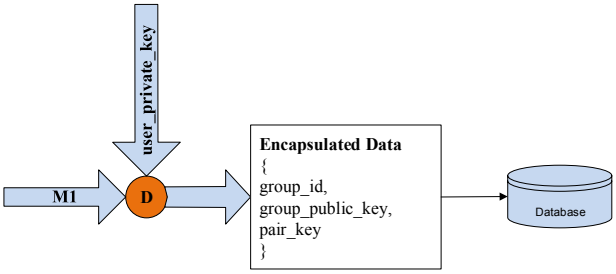


Fig. 6: Decryption of Message Digest (M1)

request to group owner for getting pair key with its public key and user id. Generation of pair_key and message digest is shown on "Fig. 5". Group Owner device will generate a random number, and pass this user_id, group_id and random number to Hash function (H) to generate a 32 bytes secured pair-key. Now it will encapsulate group_id, group_public_key and pair_key into single message and encrypt the message using user's public key to generate message digest (M1) and send to user 'X'.

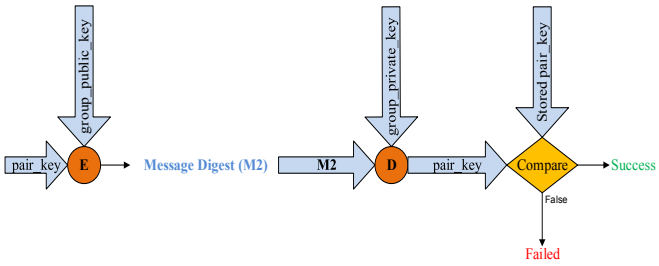


Fig. 7: User Authentication

- As shown in "Fig. 6", on receiving message digest (M1), User 'X' will decrypt the message digest using user 'X' private key to get encapsulated message {group id, group public key, pair key} and store this information in database.
- Now as shown in "Fig. 7", whenever user 'X' wants to access Group. User 'X' device will encrypt this pair_key using previously shared group's public key to generate message digest (M2) and send this to Group Owner.
- Group Owner decrypt this message (M2) using group's private key to get the pair_key and compare this pair key with the stored pair_key, if both are same means user 'X' is authenticated successfully.

IV. Result Analysis

PMS-Sharing application provides following major function:-

- User can create login account with valid email id.
- User can login to the application with valid id.
- User can create group of peer user whom with he wants to share the files.
- Dashboard should list out all the Groups created by user and Groups in which user is added.
- User can modify the group details (Name, and participant user).
- User can select the group to enter in shared directory of that group.
- User can add, remove and view the files in shared directory.
- User who creates the Group will be Group Owner by default for that particular Group.
- The Group Owner can turn ON and turn OFF sharing at any time from Quick Menu Button.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:02:08 UTC from IEEE Xplore. Restrictions apply.

- User can logout from the application any time from any screen

Let's take an example to compare the PMS-Sharing with other Sharing. Suppose there is a family with five members named as Lajjesh, Nitesh, Gunjan, Bhumika and Urav. Suppose Lajjesh wants to share some photo graphs with all other family members. Lajjesh will perform following steps to share files normally:-

- Lajjesh will choose files.
- Lajjesh will share via Wi-Fi Direct/Bluetooth and select user Nitesh, Device pairing request will be sent to Nitesh's device.
- Nitesh will accept device pairing request and both devices are paired.
- After pairing to Lajjesh's Device start transferring files to Nitesh's Device.
- Repeat above steps to share file with Gunjan, Bhumika and Urav also.

So it has been noted that in total 16 steps (4 Members * 4 Steps each) are needed to choose files, need to do manual device pairing, which incorporates active user involvement and file sending can fail also if devices don't have enough space. While in PMS-Sharing Lajjesh will perform following steps:-

- Lajjesh will login in PMS-Sharing application
- Select Family Group
- Add files which he wants to share with Nitesh, Gunjan, Bhumika and Urav.

So in PMS-Sharing only 3 steps in total are needed, whole process of transfer and authentication is automatic once the initial group is created. PMS-Sharing helps in saving overall time and user involvement for sharing. In PMS-Sharing only sender need to add files in shared folder and other user can sync/download the file whenever they want to download unlikely to the normal file transfer where file is transferred immediately even though another user don't want to do so. Similarly if it is compared with Cloud sharing where first user need internet access to use services, user can share uploaded contents URL to any other member, there is no control on data once uploaded to Cloud Storage, All the information is already exposed to the cloud service provider, any person who has account on cloud storage application and has the valid URL of shared files then he can easily access the files. While in case of PMS-Sharing user can use the services on local network without internet and the content is accessible only to the members of the group.

V. CONCLUSION AND FUTURE SCOPE

PMS-Sharing authentication model adds extra security in term of authentic access to the data shared by any other user. Unlikely to the cloud storage, in PMS Sharing user has full control on data. Based on PMS-Sharing concept, personalized cloud storage can be built with marginal cost and more research can be done to mitigate data recovery and security issues. PMS-Sharing is very useful to deploy Private Cloud for a group of know people say a Community and for a Military Services. During the development, various challenges and limitations are observed on which the future work can be done. First, Smartphones which are operating in server mode need more processing and battery power to serve the request from n-number of users. There is threat of losing connectivity with PMS-Sharing when device power down. Second, there is need to have secure and resource effecient method for Data recovery. Third, PMS-Sharing is capable of handling n number of users and groups on local networks but access points allows only limited number of users to connect at a time.

REFERENCES

[1] Yuanyuan Guo, Jianjun Hao, Yijun Guo, Tao Luo, " Research on Data Block Storage Strategy in Cloud Storage System", Proceedings of 2017 3rd IEEE International Conference on Computer and Communications, 2017, pp. 2394-2397.

[2] https://en.wikipedia.org/wiki/File_system, last accessed on: 15/09/2019.

[3] Yuanyuan Guo, Jianjun Hao, Yijun Guo, Tao Luo, " Research on Data Block Storage Strategy in Cloud Storage System", Proceedings of 2017 3rd IEEE International Conference on Computer and Communications, 2017, pp. 2394-2397.

[4] https://www.ibm.com/cloud/learn/block-storage, last accessed on: 15/09/2019.

[5] https://aws.amazon.com/ebs, last accessed on: 16-09-2019.

[6] https://www.ibm.com/cloud/learn/object-storage, last accessed on: 15/09/2019.

[7] https://www.idrive.com/, last accessed on : 21-09-2019

[8] https://www.pcloud.com/, last accessed on : 21-09-2019

[9] https://mega.nz/, last accessed on : 21-09-2019

[10] https://onedrive.live.com/about/en-gb/, last accessed on : 21-09-2019

[11] https://www.icloud.com/, last accessed on : 21-09-2019

[12] https://cloud.google.com/, last accessed on : 21-09-2019

[13] https://www.box.com/, last accessed on : 21-09-2019

[14] https://nextcloud.com/, last accessed on : 21-09-2019

[15] https://spideroak.com/, last accessed on : 21-09-2019

[16] Li Zhang, Bing Tang, "GrandStore: Towards Large-Scale Free Personal Cloud Storage", Proceedings of 2017 International Conference on Computer Network,2017, pp 118-123

[17] Chetan Gaikwad, Bhoomika Churi, Kanad Patil, Tatwadarshi P. N., "Providing Storage as as Service on Cloud using OpenStack", Proceedings of 2017 International Conference on Innovations in information Embedded and Communication Systems, 2017, pp 1-4.

[18] https://www.openstack.org/, last accessed on : 24-09-2019.

[19] B. Varsha and P. Suryateja, "Using Advanced Encryption Standard for Secure and Scalable Sharing of Personal Health Records in Cloud", International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 5 (6) , 2014.

[20] Adiseshu Hari, Ramesh Viswanathan, T. V. Lakshman, Y. J. Chang, "The personal cloud: design, architecture and matchmaking algorithms for resource management", Proceedings of the 2nd USENIX conference on Hot Topics in Management of Internet, Cloud, and Enterprise Networks and Services, 2012, pp 1-6.

[21] Alginahi, Yasser M., and Muhammad Nomani Kabir, eds. Authentication Technologies for Cloud Computing, IoT and Big Data. IET, 2019, chapter 3-4.

[22] https://www.wi-fi.org, last accessd on : 28-09-2019

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 06:02:08 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020)
IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1

Predicting Link Sign in Online Social Networks based on Social Psychology Theory and PageRank

Sanjay Kumar
Department of Computer Science and Engineering
Delhi Technological University
New Delhi, India.
sanjay.kumar@dtu.ac.in

Sudhanshu Shekhar Singh
Department of Computer Science and Engineering
Delhi Technological University
New Delhi, India.
singhsudhanshu204@gmail.com

Vipul Gupta
Department of Computer Science and Engineering
Delhi Technological University
New Delhi, India.
gvipul294@gmail.com

***Abstract*— Recently, the social networks are evolved as the primary platform for internet users to share their ideas and views. Social media interactions can be modelled as links between two users, where the link weight is different for different types of relationships and interactions. Social networks are dynamic in nature where the links are formed and deformed with time. Broadly, two types of connections, i.e. Positive and Negative can exist in most social networks. Positive connections signify trust and friendship, whereas negative connections show distrust and enmity. Reliable prediction of edge signs can greatly influence in enhancing user experience. Prediction of edge signs has been explored previously also. In this paper, the primary aim is to predict the sign of edges based on extracted features of nodes constructed upon theories of social psychology that includes classical balance theory, the status theory and emotional information theory. The ranking of nodes is used to extract features of nodes that quantify the popularity and trust of nodes on the network. These features are combined with the features extracted from socio psychological theories and try to improve the performance of our sign prediction model. Our results show that the proposed methodology achieves a reasonable performance when implemented on two real-life datasets namely Slashdot and Epinions.**

Keywords—Emotional information theory; Online Social Networks; Signed Network; Social Psychology; Node Ranking;

I. INTRODUCTION

Social networks have become a part of our daily life. They have gained enormous popularity in the past few years and play an important role in various aspects of people's lives. Social networks are not only the prime medium of communication between users, but also the primary source of information for most of the people. Social network analysis is not a formal theory but rather an investigation strategy for exploring social structures [1]. These days, product advertisement has become a huge industry. Effective advertising strategies include promoting the product to targeted audiences on online social networks. The recommendation systems are also playing an important role in deciding what a user buys on a website and also improving user experience by suggesting the relevant products and services. There are various other applications of social networks that rely on reliable and robust sign prediction and link prediction algorithms. A lot of effort in the research field of social networks is directed towards understanding the relationships between social entities [2].

Numerous applications of social network analysis are available in various fields such as social content filtering, recommendation systems, behavior analysis, network modeling, etc. Social network analysis has even demonstrated that it is effective in online collaborative learning also [3]. Some of the popular research fields in social network analysis include community detection, partitioning of social networks based on sign, ranking of nodes in social networks, sign prediction, etc.

Link prediction is one of the major research interests in the field of social network analysis. Link prediction has been at the core of many applications like bioinformatics, e-commerce and information retrieval. Link prediction can be used to predict the likelihood of formation of a link between the nodes in the future which currently have no link between them [4]. Sign prediction has also evolved as a growing research interest in social network analysis. Signed networks provide better understanding of social entities and their behavior than the unsigned networks. Various social networks can be represented as a large scale, enormous signed network. The weights and sign of links between users can be derived from the relationship and interaction of users on social networks [5].

In this paper, social psychology is used to analyze signed social networks. The socio-psychological theories used here are, the social status theory, social balance theory and emotional information theory. It has been further tried to use the modified version of the popular pagerank algorithm to determine the popularity and trust of a node. It is then analyzes on how the introduction of these features impacts the prediction accuracy of the model. In this paper, the features are extracted from these theories that helps to characterize the type of relationship between two nodes of a graph based on the social behavior and relationship with other nodes in the graph. Previous studies have been conducted to predict the signs of nodes in the network. So far, the sign of the network has been predicted based on Balance theory and Social status theory [7]. The pagerank algorithm has also been used many times in the field of sign prediction [20]. In this research paper, it has been tried to make use of both socio-psychological theories and the ranking of nodes to predict signs of nodes in a network. The

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:58:59 UTC from IEEE Xplore. Restrictions apply.

first class of features is based on the socio-psychological theories that include the social balance theory [8], the social status theory and the emotional information theory [9]. The second class of features has been extracted on the basis of the Troll-Trust model [20].

Our study provides insight into the nature of the relationships and also provides a large scale evaluation of socio-psychological theories [10] on real-life datasets namely Slashdot and Epinions.

The background of signed social network theories is divided into the following three subsections.

A. *Balance Theory*

According to social psychology, the signed graphs are incorporated to model the social situations where friendship can be inferred from the edges with the positive sign while the negative edges characterize enmity between the respective nodes. Figure1 (A) illustrates the balanced triads while Figure1 (B) illustrates the unbalanced triads. Balance theory suggests that the positive triads (i.e. triads with all three positive edges or two positive edges and one negative edge) are expected to be in stable social situations and thus, are balanced. On the other hand, negative triads are unbalanced and exist in rare unstable social circumstances. Structural balance can be used to express how the feelings, attitudes, and beliefs, which the social actors have towards each other, promote the formation of stable social groups [11].

B. *Status Theory*

The fundamental principle of status theory implies that the recipient of the positive directed link is supposed to have higher status than the originator node while in the case of the negative directed link, the recipient is considered to have lower status.

C. *Emotional Information Theory*

Emotional Information about the nodes can be derived from the nature of relationship that a node forms. It has also been tried to find out the optimism and pessimism of every node in the network. The node with high value optimism is more optimistic about the relationships in the network and is more likely to form links with other nodes and vice versa. High value of pessimism is attributed to skepticism about other nodes and the nodes with high value of pessimism are less inclined towards formation of relationships. It has also been experimented to exploit the information and figure out the values of optimism and pessimism for each node and use those values for prediction of sign of links in the network [9].

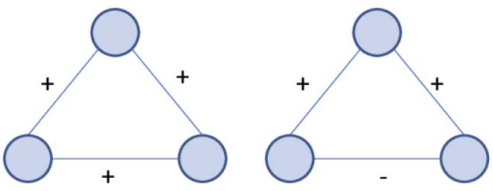


Figure 1 (A) represents a Balanced Triad

is m

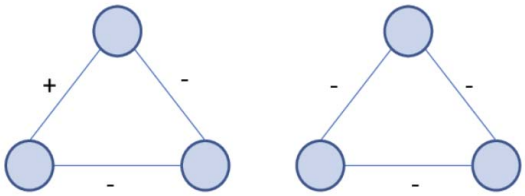


Figure 1 (B) represents an Unbalanced Triad

D. *PageRank algorithm*

The traditional PageRank algorithm is used to rank web pages on the search engine. PageRank counts the number and quality of links to a page and ranks the pages on the basis of popularity [13, 14]. The popular pages are likely to have links from various pages across the internet. The traditional pagerank calculates the probability that a person randomly clicking on links arrives at a particular page. Pagerank iterates over the graph multiple times to adjust the pagerank values of the nodes in the network.

Pagerank values are uniformly distributed in the network at the beginning of computation. Pagerank of every node is adjusted in every iteration using the following formula.

$$\boldsymbol{PR(u) = 1 - d + d \sum_{v \epsilon B^u} \frac{PR(v)}{L(v)}}$$

In the above equation PR (u) is the pagerank of the node u in consideration. B^u is the set of all the nodes that have an edge directed towards node u. The parameter d accounts for damping factor. Damping factor is the probability that the random surfer will continue surfing after the current step. L(v) is the total number of outgoing links from the node v [22].

This explanation provides intuition about the use of pagerank in unsigned networks. The nodes with more number of direct and indirect links directed towards it are given a higher rank, making them more likely to be liked by other nodes. Pagerank effectively captures various relations amongst vertices of graph [21]. Ranking of nodes in unsigned networks is a well known concept [15, 16]. However, use of pagerank for ranking of nodes in a signed network has been explored recently [17, 18]. The traditional pagerank is designed only for unsigned networks. However slight modifications to the existing pagerank algorithm to accommodate negative weights have shown impressive results in prediction of sign of the edges [19].

The rest of the paper is organized as follows: Section II discusses the datasets present. Section III elaborates the proposed methodology and implementation details followed by Section IV which elaborates and analyses the results obtained. Finally Section V concludes our research paper.

II. DATASET DESCRIPTION

In this work, two real-life data sets are considered, they are Slashdot and Epinions, which is publicly available at www.snap.stanford.edu/data. Table 1 summarizes the data sets.

The description of datasets is as follows:

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:58:59 UTC from IEEE Xplore. Restrictions apply.

A. Slashdot

Slashdot is a technology related news website. This website consists of current technology related news. Slashdot allows users to tag each other as friends or enemies.

B. Epinions

Epinions is a general product review website where users are allowed to rate the product. The value of rating indicates how much the product is liked by the user. Higher ratings indicate positive feedback whereas lower ratings indicate negative feedback.

Table 1. Dataset Description

	Epinions	**Slashdot**
Nodes	119,217	82,144
edges	841,200	549,202
+edges	85.0%	77.4%
-edges	15.0%	22.6%

III. PROPOSED METHODOLOGY

The method has a connected graph, which have nodes with both positive and negative links. Initially, it starts to extract the features that model the local relation of the node with the rest of the nodes of the network. It has two types of features i.e. features based on socio-psychological theories and features based on ranking of nodes.

The work has been conducted in following steps:

1. First of all, the features are extracted based on socio-psychological theory. These features can be further classified into three sets of features. These three sets of features are derived from three different theories.

   The features based on socio-psychological theories are enlisted below:

   a. These features include $D^{+}_{in}(v)$, $D^{-}_{in}(v)$, $D^{+}_{out}(u)$ and $D^{-}_{out}(u)$. $D^{+}_{in}(v)$ and $D^{-}_{in}(v)$ represent inward positive and inward negative edges of v while $D^{+}_{out}(u)$ and $D^{-}_{out}(u)$ represent outward positive and outward negative edges on U. And the feature C(u,v) is the total number of common neighbors of u and v, i.e. total number of nodes w that have edges connecting to both u and v, no matter whatever the type of edge is. Other two features include total outdegree of u i.e $D^{-}_{out}(u) + D^{+}_{out}$ as well as total in degree of $D^{+}_{in}(v) + D^{-}_{in}(v)$ [8].

   b. For this feature set, every triad (u, v, w) is considered, such that w has an edge either to or from u and also an edge either to or from v. Since the edges can be of any sign and any direction, there are 2*2*2*2=16 possibilities. Hence, there are sixteen types of triads possible where each triad can provide evidence of different signs [8]. Our sixteen-second class features include the number of occurrences of the sixteen types of features.

   c. For the third class of features, the features are derived by the emotional intelligence theory. Emotional intelligence theory shows the correlation between emotional information and positive/negative links. The measure of pessimism and optimism has been performed for all the nodes.

      A set of OL_i for each node u_i is considered, such that the elements of set OL are the nodes which have positive emotions from nodes u_i but have higher negative emotions received than the average negative emotions received by nodes in the graph [2].

      $$OL_i = \{u_j \mid P_{ij} != 0 \;\&\&\; N_j > N_{avg}\}$$

      It further defines OH such that OH is a subset of OL such that elements present in OH receive lower positive emotions than the average of the entire graph.

      $$OH_i = \{u_k \mid u_k \in OL_i \wedge P_k < P_{avg}\}$$

      The optimism is measured as the ratio of the length of OH to OL.

      $$O_i = \frac{|OH_i|}{|OL_i|}$$

      The high optimism feature shows that the node with high optimism has positive emotions towards the nodes with low positive and high negative emotions across the graph and hence is more likely to form positive relations with other nodes too.

      Pessimism is calculated and defined as a set PH_i for all the edges u_i such that PH_i contains all the nodes that have negative emotions from the node u_i but have higher value of positive emotions from other nodes in the graph than the average positive emotions.

      $$PH_i = \{u_j \mid N_{ij} != 0 \;\&\; P_j > P_{avg}\}$$

      It further defines PL_i for u_i such that PL_i is a subset of PH_i and contains elements of PL_i which

      Have more negative emotions than the average of the entire graph.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:58:59 UTC from IEEE Xplore. Restrictions apply.

$$PL_i = \{u_k \mid u_k \in PH_i \ \& \ N_k < N_{avg}\}$$

Pessimism is given by:

$$P_i = \frac{|PL_i|}{|PH_i|}$$

A high value of pessimism shows that the given node is more likely to form a negative relationship with undiscovered nodes because it already has negative relations with the nodes having positive emotions from other nodes across the graph. The optimism and pessimism of both nodes are considered for participating in the edge as the features for the learning algorithm.

The above three classes of features constitute the features based on socio-psychological theory. In the above feature set, which made use of one or the other socio-psychological theory to extract the features.

2. After this, the features are extracted based on ranking of nodes. Later, the graph is divided into two sub graphs, i.e. A^+ and A^-. A^+ is the positive sub graph containing all positive edges while A^- is a negative sub graph containing the negative edges only.

   The pagerank algorithm is used to calculate ranking scores ($\square_i$) of each node in both the subgraphs. It adopt two measures, reputation, and optimism, as features of nodes for classification purposes. The feature reputation quantifies the popularity of the node in the network, whereas Trust quantifies the amount of trust, the node has in other nodes of the network and the pattern in which the node makes votes.

$$\text{Reputation}_i = \frac{\Sigma\square_{j:(j,i)\epsilon A^+} - \Sigma\square_{j:(j,i)\epsilon A^-}}{\Sigma\square_{j:(j,i)\epsilon A^+} + \Sigma\square_{j:(j,i)\epsilon A^-}}$$

$$\text{Trust}_i = \frac{\Sigma\square_{j:(i,j)\epsilon A^+} - \Sigma\square_{j:(i,j)\epsilon A^-}}{\Sigma\square_{j:(i,j)\epsilon A^+} + \Sigma\square_{j:(i,j)\epsilon A^-}}$$

R_i and T_i represent the reputation and trust of the node (i).
The more popular the node is, the more edges will be directed towards it. Hence it will have a higher reputation. Whereas higher optimism of a node shows that it has many outward edges directed towards other nodes from itself. It use trust and reputation of both nodes in the edge as edge features [20].

It merges both feature sets defined above and train the machine learning model on a combined large dataset.

It chooses as a logistic regression classifier to predict sign prediction as it experiences a good performance of this classifier in social network analysis [12]. This has trained the model on the features extracted from the graphical data. This has extracted three classes of features (i.e. first, second and third class of features) from the data to train the model on each one of them separately.

IV. RESULT

In the following Table 2, values signify the minimum value of embeddedness allowed in the graph and the prediction accuracy of model corresponding to each minimum threshold embeddedness level (e.g., Embeddedness 25 means that only those links with embeddedness greater than or equal to 25 have been considered for prediction)

Table 2: Prediction accuracy of the model when the features based on social psychological theories were used for training

	Embedded-ness(0)	**Embedded-ness(10)**	**Embedded-ness(25)**
Slashdot	87.2	88.4	89
Epinions	93.3	95.9	96.6

The following Figure 2(A) shows the prediction accuracy of the model on Epinions dataset when trained on the features based on social psychological theories whereas the Figure 2(B) shows the prediction accuracy of the model on Slashdot dataset when trained on the features based on social psychological theories.

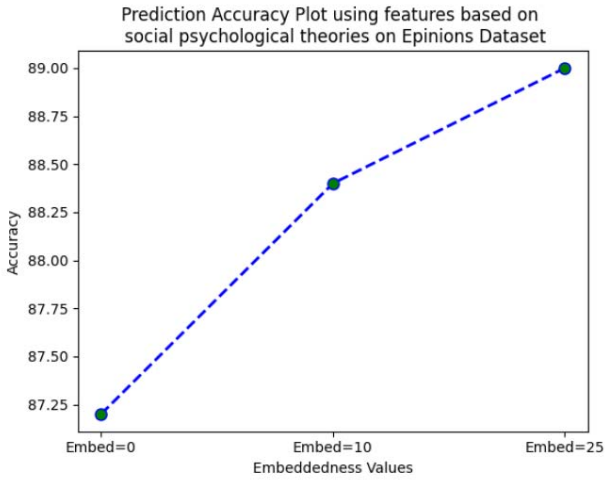


Figure 2(A)Prediction accuracy for Epinions Dataset when the model is trained on the features based on social psychological theories.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:58:59 UTC from IEEE Xplore. Restrictions apply.

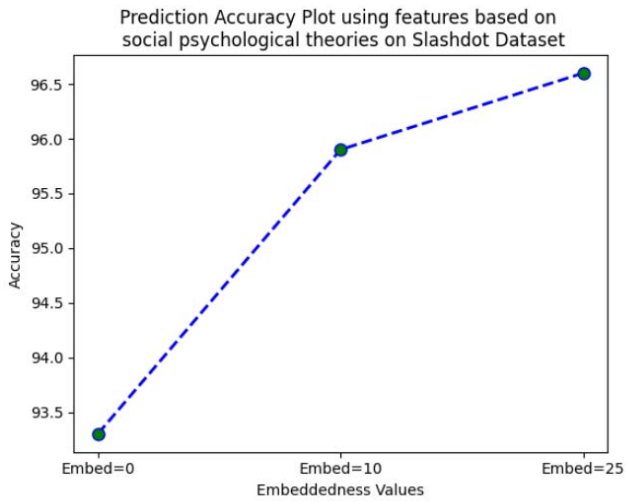


Figure 2(B)Prediction accuracy for Slashdot Dataset when the model is trained on the features based on social psychological theories.

The following table 3 and Figure 3 depict the accuracy values of the model based on the node ranking algorithms on both Epinions and Slashdot datasets.

Table 3: Prediction accuracy of the model when the features based on node ranking algorithms were used for training

	Accuracy
Slashdot	85.7
Epinions	94.4

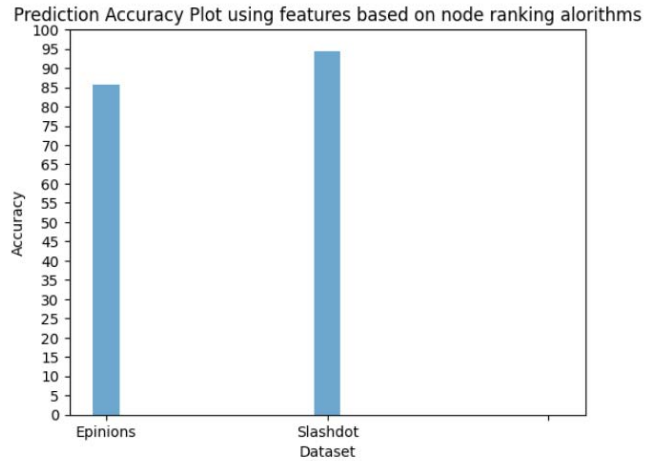


Figure 3 Prediction accuracy for Epinions and Slashdot Dataset when the model is trained on the features based on the node ranking algorithms.

The proposed methods has further used logistic regression as well as artificial neural networks on the feature set obtained by integrating the social psychological theories feature set and the features based on node ranking algorithms on both Epinions and Slashdot datasets. The prediction accuracy of logistic regression on the new integrated feature set is found to be higher than the prediction accuracy of both the first and second set of features when considered separately.

The following Table 4 and Figure 5 show the accuracy comparison of the model before and after the introduction of features based on node ranking algorithms on Epinions dataset.

Table 4: Accuracy Comparison Table of the model before and after the introduction of features based on node ranking algorithms on Epinions dataset.

Algorithm Architecture	**Improved Accuracy**	**Accuracy Earlier**
Embeddedness-0 Logistic Regression	95.5%	93.3%
Embeddedness-0 Neural Network	92.5%	92.4%
Embeddedness-10 Logistic Regression	97.1 %	95.9%
Embeddedness-10 Neural Network	92.5 %	92.5 %
Embeddedness-25 Logistic Regression	97.7 %	96.6%
Embeddedness-25 Neural Network	92.6 %	92.6 %

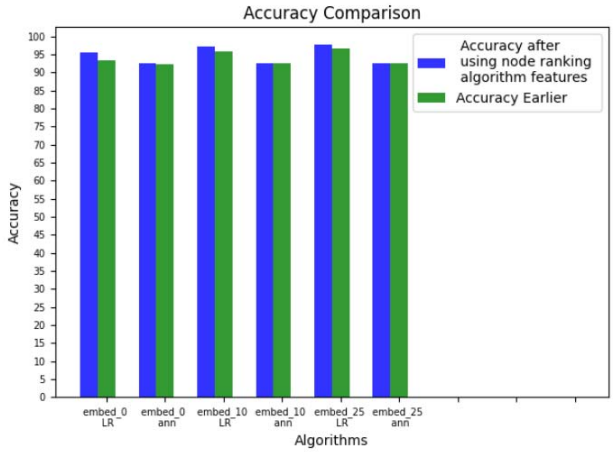


Figure 5: Accuracy comparison depiction for Table 4.

The following Table 5 and Figure 6 shows the accuracy comparison of the model before and after the introduction of features based on node ranking algorithms on Slashdot dataset.

Table 5: Accuracy Comparison Table of the model before and after the introduction of features based on node ranking algorithms on Slashdot dataset.

Algorithm Architecture	**Improved Accuracy**	**Accuracy Earlier**
Embeddedness-0 Logistic Regression	88.5%	87.2%
Embeddedness-0 Neural Network	92.5%	92.3%

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:58:59 UTC from IEEE Xplore. Restrictions apply.

Embeddedness-10 Logistic Regression	89.4%	88.4%
Embeddedness-10 Neural Network	91.5 %	91.6 %
Embeddedness-25 Logistic Regression	89.9 %	89%
Embeddedness-25 Neural Network	91.5 %	91.7 %

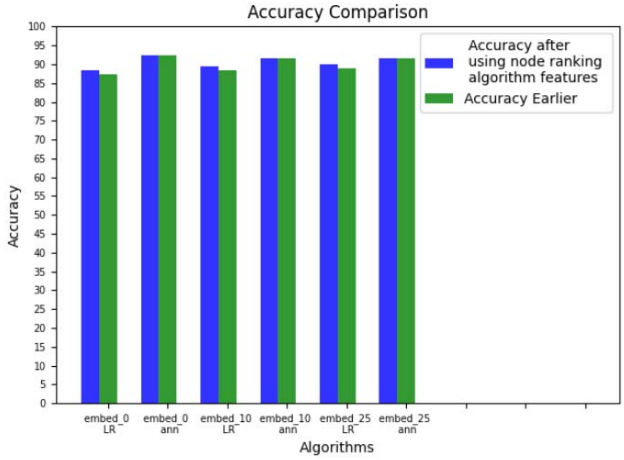


Figure 6: Accuracy comparison depiction for Table 5.

These have analyzed the results and draw insights from the results shown. The analysis of the result is divided into the following two subsections:

A. *Improvement of Accuracy with embeddedness:*

Embeddedness of a particular edge represents the number of common neighbors of the nodes participating in the link. The threshold embeddedness values are used for analyzing the result. From Figure 2(A), Figure 2(B) and Table 2, illustrates that there is a significant improvement in accuracy as the minimum embeddedness threshold is increased for a graph. As it increase the threshold embeddedness of links in the proposed network, where it only select nodes that have more information about their common neighbors. Our result shows that the accuracy of prediction increases with increase in the threshold embeddedness of the graph. As it increases the minimum embeddedness threshold of the graph, it selects only those nodes that have more behavioral information present in the network, which is encoded in the relationships between the nodes of the edge under consideration and their neighbors. This helps our model to make more informed predictions than that on the original graph. The trend of increasing predictive accuracy holds good even with the introduction of the set of features based on ranking of nodes.

B. *Improved accuracy:*

The observations in Table 3 depict the predictive accuracy of training the model on features based on node ranking algorithms. Table 4 and Table 5 show the comparison of prediction accuracies of models before and after introduction of node ranking based features. The prediction accuracy has improved with the introduction of features based on node ranking algorithms. This observation signifies that combining the first and second set of features for prediction increases the accuracy and makes the model more efficient.

The predictive accuracy of the model when trained only in the third class of features using Logistic Regression algorithm is given in Table 2. The features based on node ranking method are integrated with the features based on social psychological theories, and the model is trained on these features. The "Improved Accuracy" column of Table 4 and Table 5 shows the predictive accuracy of the model when trained on the big feature set which contains both the feature sets. There is a significant increase in the predictive accuracy of the model.

V. CONCLUSION

From the above analysis, it has been concluded that the accuracy of the learning model improves with the introduction of features based on node ranking. It is also observed that the accuracy increases by increasing embeddedness of the graph. The above analysis shows that the features based on ranking algorithms are very relevant to signed social networks. Training model on both, features based node ranking and features based on social theories improve the accuracy of prediction of the model. It is also concluded that the reputation of node is a measure of popularity of the node in a signed network. A node with high positive reputation value is likely to receive more positive edges towards it, whereas a node with high negative value of reputation is likely to receive more negative edges towards it. Therefore, nodes are more likely to form a positive relationship with a node with higher positive reputation, whereas other nodes in a network are more likely to dislike a node with high value of negative reputation. Similarly, the trust value of nodes depicts how optimistic a node is about other nodes in the network. A node with high value of positive trust is more likely to form positive relationships with other nodes of the network, whereas a node with high value of negative trust is more likely to form negative links with other nodes in its neighborhood. Introduction of the features, trust and reputation, provides vital information about the behavior of nodes in the network. Adding these features to our feature set enhances the performance of our model. The result also shows that the improvement in prediction accuracy of logistic regression model after introduction of features based on node ranking is more significant than the improvement in artificial neural network model.

The results show that the accuracy of prediction model has reasonably increased after augmenting the features trust and reputation that have been derived from the pageranks of the nodes.

REFERENCES

[1] E. Otte, and R. Rousseau. "Social network analysis: a powerful strategy, also for the information sciences." Journal of information Science 28.6 (2002): 441-453.
[2] S. Wasserman, and K. Faust. Social network analysis: Methods and applications. Vol. 8. Cambridge university press, (1994).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:58:59 UTC from IEEE Xplore. Restrictions apply.

[3] M. Saqr, U. Fors, and J. Nouri. "Using social network analysis to understand online Problem-Based Learning and predict performance." PloS one 13.9 (2018): e0203590.
[4] M. Al Hasan, and M. J. Zaki. "A survey of link prediction in social networks." In Social network data analytics, pp. 243-275. Springer, Boston, MA, 2011.
[5] S. Maniu,B. Cautis, T Abdessalem. " Building a signed network from interactions in Wikipedia." In Databases and Social Networks ,pp. 19-24. ACM, 2011.
[6] M Stojmenovic., G Lindgaard ."Benefits and Limitations of the Social Network Analysis when explaining instances of ineffective communication in two Chemical, Biological, Radiological, Nuclear, and Explosives simulations." In Cloud and Green Computing (CGC), 2013. Third International Conference on (pp. 327-334). IEEE, 2013.
[7] J. Leskovec, D. Huttenlocher, and J. Kleinberg. "Signed networks in social media." Proceedings of the SIGCHI conference on human factors in computing systems. ACM, 2010.
[8] J. Leskovec, D. Huttenlocher, and J. Kleinberg. "Predicting positive and negative links in online social networks." Proceedings of the 19th international conference on World wide web. ACM, 2010.
[9] B. Ghazaleh, J. Tang, and H. Liu. "Signed Link Analysis in Social Media Networks." ICWSM. 2016.
[10] A. Clifton , and G. D. Webster. "An introduction to social network analysis for personality and social psychologists." Social Psychological and Personality Science 8.4 (2017): 442-453.
[11] N. P. Hummon, and P. Doreian. "Some dynamics of social balance processes: bringing Heider back into balance theory." Social Networks 25.1 (2003): 17-49.
[12] J. L.Matthews, P. DeWan, and E. Y. Rula. "Methods for inferring health-related social networks among coworkers from online communication patterns." PloS one 8.2 (2013): e55234.
[13]Chen P, Xie H, Maslov S, Redner S. "Finding scientific gems with Google's PageRank algorithm." Journal of Informetrics. 2007 Jan 1;1(1):8-15.
[14]Franceschet M. "PageRank: Standing on the shoulders of giants." Communications of the ACM. 2011 Jun 1;54(6):92-101.
[15]Chiang KY, Hsieh CJ, Natarajan N, Dhillon IS, Tewari A. "Prediction and clustering in signed networks: a local to global perspective." The Journal of Machine Learning Research. 2014 Jan 1;15(1):1177-213.
[16]Kleinberg JM. Authoritative sources in a hyperlinked environment. Journal of the ACM (JACM). 1999 Sep 1;46(5):604-32.
[17]Kamvar SD, Schlosser MT, Garcia-Molina H. "The eigentrust algorithm for reputation management in p2p networks." InProceedings of the 12th international conference on World Wide Web 2003 May 20 (pp. 640-651).
[18]Kerchove CD, Dooren PV. "The pagetrust algorithm: How to rank web pages when negative links are allowed?." InProceedings of the 2008 SIAM International Conference on Data Mining 2008 Apr 24 (pp. 346-352). Society for Industrial and Applied Mathematics.
[19]Tang J, Chang S, Aggarwal C, Liu H. "Negative link prediction in social media." InProceedings of the eighth ACM international conference on web search and data mining 2015 Feb 2 (pp. 87-96).
[20]Wu Z, Aggarwal CC, Sun J. "The troll-trust model for ranking in signed networks.". InProceedings of the Ninth ACM international conference on Web Search and Data Mining 2016 Feb 8 (pp. 447-456).
[21] Chung F, Zhao W. "PageRank and random walks on graphs." InFete of combinatorics and computer science 2010 (pp. 43-62). Springer, Berlin, Heidelberg.
[22] Brin S, Page L. "The anatomy of a large-scale hypertextual web search engine."

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 16,2020 at 05:58:59 UTC from IEEE Xplore. Restrictions apply.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/341238308

Real Time Experimental control and Design of FOPI λ and L–PID Controllers tuned by Invasiveweed optimization for Trajectory Control of TRAS

Conference Paper · February 2020

DOI: 10.1109/PARC49193.2020.236577

CITATIONS
0

READS
13

2 authors:



Hem Prabha
Delhi Technological University
2 PUBLICATIONS **0** CITATIONS
SEE PROFILE



Rajul Kumar
Delhi Technological University
6 PUBLICATIONS **0** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Fuzzy controller View project

All content following this page was uploaded by Rajul Kumar on 07 July 2020.

The user has requested enhancement of the downloaded file.

2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC)
GLA University, Mathura, UP, India. Feb 28-29, 2020

2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC) 978-1-7281-6575-2/20/$31.00 ©2020 IEEE 10.1109/PARC49193.2020.236577

Real Time Experimental control and Design of FOPI^{λ} and L-PID Controllers tuned by Invasive-weed optimization for Trajectory Control of TRAS

Hem Prabha
Dept. of Electrical Engineering
Delhi Technological University
Delhi-110042, India
prabhasingh1686@gmail.com,

Rajul Kumar
Dept. of Electrical Engineering
Delhi Technological University
Delhi-110042, India
rajul_bt2k17@dtu.ac.in

***Abstract*— This article presents formational designing and real time experimental validation of Fractional order-Proportional integral (FOPI^{λ}) and Linear-Proportional Integral Derivative (L-PID) controllers for trajectory stabilization of Twin Rotor Aerial System (TRAS) which is conceded as helicopter model having pitch and yaw actuators for hovering of TRMS in vertical and horizontal planes respectively. Both the controller parameters are tuned using robust metamorphic algorithm called Invasive Weed Optimization. The control performances of both the designed controllers are examined for trajectory tracking and time taken to stabilize after the given manual disturbances. It is experimentally validated that FOPI^{λ} shows superior and robust control behavior then L-PID for benchmark TRAS plant.**

Keywords—IWO, FOPI$^{\lambda}$, L-PID, TRAS, Stabilization Control

I. INTRODUCTION

In contemporary year's automatic control of aerial vehicles are in curious interest because of their non-linear, complex system dynamics and wide range of applications in sky vigilance, detecting forest fire, natural calamity relief assistance[1]-[4]. Twin Rotor Ariel System (TRAS) plant mimics 2-DOF helicopter model having a two 44 watt DC motor propellers for pitch and yaw angle control and for achieving desired set-point hovering in vertical and horizontal planes without significant cross coupling between the two planes. The pitch and yaw DC motors receives voltage control signals from PC via PCI-1711 card and vertical beam which is mounted over electrical base plate. These control signals decisive the needed pneumatic torque by altering the rpm of pitch and yaw rotors resulting change in pitch and yaw hovering angles[5], [6].

TRAS is recognized as most non linear and complex behavior because of three moments i.e. gravity, friction and gyroscopic acting on pitch and yaw propellers [7], [8]. The stabilization control of such complex plants cannot be fulfilled by adoption traditional controllers like L-PID [9], [10].

To realize these complex system uncertainties two Fractional order-PI$^{\lambda}$ controllers are designed and experimentally validated in real time for pitch and yaw angle stabilization control. The designed FOPI$^{\lambda}$ controller have capability to stabilize close loop dynamics from dynamic environment through optimized fractional power of integral scheme [11], [12].

For optimal utilization and tuning of designed controllers a ecologically inspired catechistic algorithm [13], [14] called invasive weed optimization is used. Invasive weed optimization (IWO)[15]-[17] is influenced from robust pioneering of invasive weeds in plant cultivation. Invasive weed Colonize very vigorously and shows very adaptive nature to environmental diversification. Therefore the robust acclimating behavior of invasive weeds provides an edge for dealing with optimization of complex plants in control engineering practices. After tuning with IWO trajectory control performance both FOPI$^{\lambda}$ and L-PID controller are validated and compared mathematically and experimentally by giving real time manual disturbances for a regular time interval, these manual disturbances provided through counterbalance resembles aerial, wind and pneumatic disturbances on helicopter.

The paper is fragmented into five sections, Introduction then section II describing TRAS plant dynamical Modelling followed by Section III comprises designing of FOPI$^{\lambda}$ and L-PID controllers both for pitch and yaw propellers. Section IV intimates the experimental comparisons and results followed by conclusion and references in Section IV and V.

II. DYNAMICAL MODELLING OF TWIN ROTOR ARIEL SYSTEM (TRAS)

A. Dynamical Momentum Equations

The experimental laboratory setup for TRAS is given in fig.1 with pitch and yaw two angular rotors.

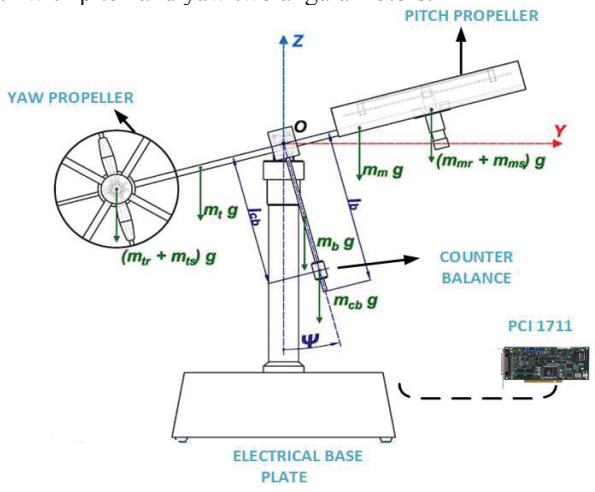


Fig. 1. TRAS Experimental system

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 18,2020 at 16:42:54 UTC from IEEE Xplore. Restrictions apply.

The mathematical statements of pitch angle (Ψ) propeller are provided in equation (1) - (5).

$$I_1\ddot{\Psi} = M_1 - M_{FG} - M_{B\Psi} - M_G \quad (1)$$

$$M_1 = a_1\tau_1^2 + b_1\tau_1 \quad (2)$$

$$M_{FG} = M_g sin\Psi \quad (3)$$

$$M_{B\Psi} = B_{1\Psi}\dot{\Psi} + B_{2\Psi}sign\dot{\Psi} \quad (4)$$

$$M_G = K_{gy}M_1\dot{\varphi}cos\Psi \quad (5)$$

The mathematical statements of yaw angle (φ) propeller are provided in equation (6) - (8).

$$I_2\ddot{\varphi} = M_2 - M_{B\varphi} - M_R \quad (6)$$

$$M_2 = a_2\tau_2^2 + b_2\tau_2 \quad (7)$$

$$M_{B\varphi} = B_{1\varphi}\dot{\Psi} + B_{2\varphi}sign\dot{\varphi} \quad (8)$$

Dynamical systemization of pitch and yaw rotors consist of gyroscopic, frictional and non-linear, gravity characteristics momentums that influence TRAS hovering non-linearities and panorama. The numerical values of system parameters including static, frictional and cross-coupling constants are given in Table.1.

TABLE I. TRAS PARAMETRIC VALUES

TRMS modeling Constants	Values
$\boldsymbol{I_1}$- M.I of Pitch Propeller	$\mathbf{7.9 \times 10^{-2} Kgm^2}$
$\boldsymbol{I_2}$- M.I of Yaw Propeller	$\mathbf{4 \times 10^{-3} Kgm^2}$
$\boldsymbol{a_1}$- Pitch Propeller Static Constant	**0.4134**
$\boldsymbol{b_1}$- Pitch Propeller Static Constant	**0.655**
$\boldsymbol{a_2}$- Yaw Propeller Static Constant	**0.67**
$\boldsymbol{b_2}$- Yaw Propeller Static Constant	**0.74**
$\boldsymbol{M_g}$- Gravity Moment	**0.399**
$\boldsymbol{B_{1\psi}}$- Frictional Moment of Pitch Propeller	$\mathbf{5 \times 10^{-0.6}}$
$\boldsymbol{B_{2\psi}}$- Frictional Moment of Pitch Propeller	$\mathbf{0.7 \times 10^{-0.6}}$
$\boldsymbol{B_{1\varphi}}$- Friction Moment of Yaw Propeller	$\mathbf{2 \times 10^{-0.5}}$
$\boldsymbol{B_{2\varphi}}$- Friction moment of Yaw Propeller	$\mathbf{3.5 \times 10^{-0.5}}$
$\boldsymbol{K_{gy}}$- Gyroscopic Moment	**0.053**
$\mathbf{k_1}$- DC Gain of Pitch Propeller	**1.45**
$\boldsymbol{k_2}$- DC Gain of yaw Propeller	**0.79**
$\boldsymbol{T_{11}}$- Moment Parameter of Pitch Propeller	**1.01**
$\boldsymbol{T_{10}}$- Moment Parameter of Pitch Propeller	**1**
$\boldsymbol{T_{21}}$- Moment Parameter of yaw Propeller	**1**
$\boldsymbol{T_{20}}$- Moment Parameter of yaw Propeller	**1**
$\boldsymbol{T_p}$- Cross-coupling value	**2.1**
$\boldsymbol{T_0}$- Cross-coupling value	**3.85**
$\boldsymbol{k_c}$- Gain of Cross-coupling	**−0.34**

After Solving and substituting TRAS system constants in non-linear pitch and yaw rotor equations gives transfer function in Laplace state-space form in equation (9) and (10).

$$T_P = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \quad (9)$$

$$T_{p(s)} = \begin{bmatrix} \frac{2.346}{(s+0.934)(s^2+0.0765s+3.689)} & 0 \\ \frac{2.567(s+0.365)}{s(s+6)(s+0.944)(s+0.95)} & \frac{4.8}{s(s+9)(s+3)} \end{bmatrix} \quad (10)$$

III. FOPI^λ AND L-PID CONTROLLER DESIGNING

Two FOPI^λ and L-PID controllers are designed for pitch and yaw propellers stabilization. The general control scheme is given in equation (11) consisting proportional, integral and derivative control arrangements where as equation (12) shows error function as difference of desired propeller angle and experimental propeller angle.

$$c(t) = k_p e(t) + k_i \int e(t)dt + k_d \frac{de(t)}{dt} \quad (11)$$

$$e(t) = z_{desired\ angles} - z_{actual\ angles} \quad (12)$$

A. L-PID CONTROL SCHEME

The control structural of L-PID control arrangement with control equation (13) in mathematical form is provided in Fig.2.

$$C(s) = k_p + \frac{k_i}{s} + k_d s \quad (13)$$

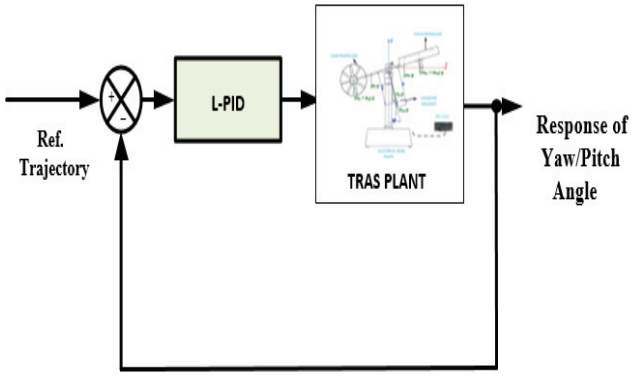


Fig. 2. L-PID control arrangement

B. FOPI^λ CONTROL SCHEME

According to equation (8), the modified transfer function of IO-PID controller structure is written in (10)

$$C(s) = k_p + \frac{k_i}{s^\lambda} \quad (10)$$

The design of IO-PID controller with TRMS plant is shown in Fig.3.

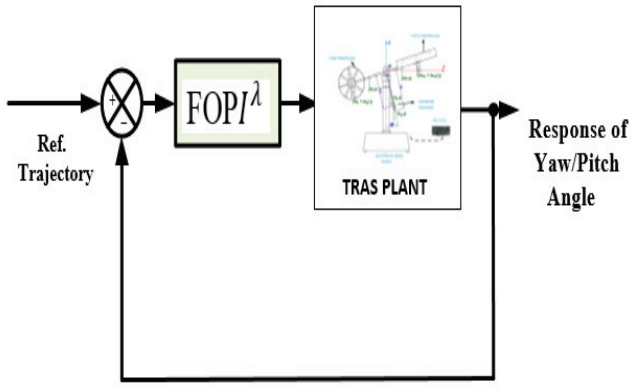


Fig. 3. FOPI$^\lambda$ control arrangement

IV. INVASIVE WEED OPTIMIZATION (IWO) OF DESIGNED FOPI^λ AND L-PID CONTROLLERS

In recent years scientists and industries are fascinated towards solving non-definitive or discontinuous control engineering problems by nature influenced non-gradient

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 18,2020 at 16:42:54 UTC from IEEE Xplore. Restrictions apply.

stochastic algorithms. Invasive weed optimization is one of the ecological influenced metamorphic optimization technique. In scientific language Weeds can be elaborated as any undesirable shrubs, trees, saplings in our cultivation or optimal solution search space. Invasive traits of weeds poses critical hazard to cultivation, so many attempts like tillage, defoliation and machine weeding are rendered to get rid of these nasty and stubborn plants.

They can reproduce by seeds or sex cells, when ancestor weed forms a seed by fertilization of egg through pollen. They flourish and reproduce at the last cycle of their life, although the way of living their entire life cycle depends upon two selections called r and k force criterion. The r criteria includes more abundance, little troop and low value to permanent colonization for winning in competition this cooperates them to succeed in insecure and unpredictable ecological conditions, whereas k criteria consist of bigger troops, slow fecundity and give more emphasis on permanent colonization. After reproduction process seeds are haphazardly dispersed in solution search space according to equations (11)-(13), through random values such that they domicile near their parent weeds shown in Fig.4, Fig.5, Fig.6.



Fig. 4. Distributing of initial spore



Fig. 5. Dispersing of spores in neighborhood orbits



Fig. 6. Rolling down of spores into most distant orbits

Weeds are very adaptive and flexible to their surroundings and environmental changes, they overrun and seize the search space by permanent colonization and also getting improved by having best fitness among plant community through interacting with peer weeds based upon their capability traits. Weeds are allowed to clone based upon the linear fitness enhancement of whole community as shown in Fig 7.

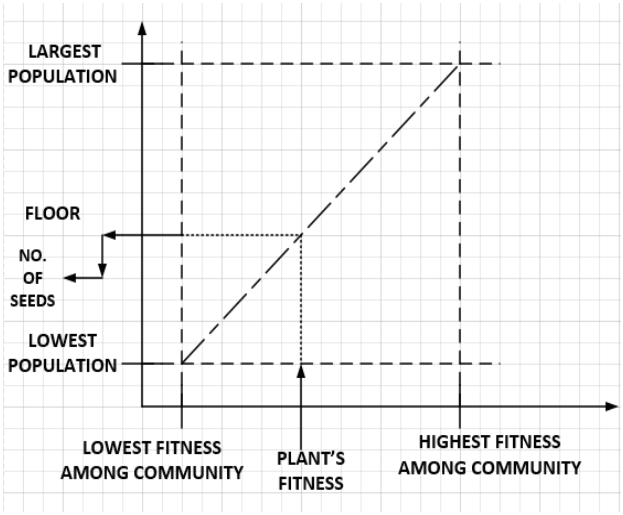


Fig. 7. Linear increase in reproduction of seeds in weeds community

$$Mean = 0 \tag{11}$$

$$Variance = varied\ over\ search\ space \tag{12}$$

Standard deviation is given by σ_{loops} ,

$$\sigma_{loops} = \frac{(loop_{max} - loop)^a}{(loop_{max})^a}(\sigma_{start} - \sigma_{end}) + \sigma_{end} \tag{13}$$

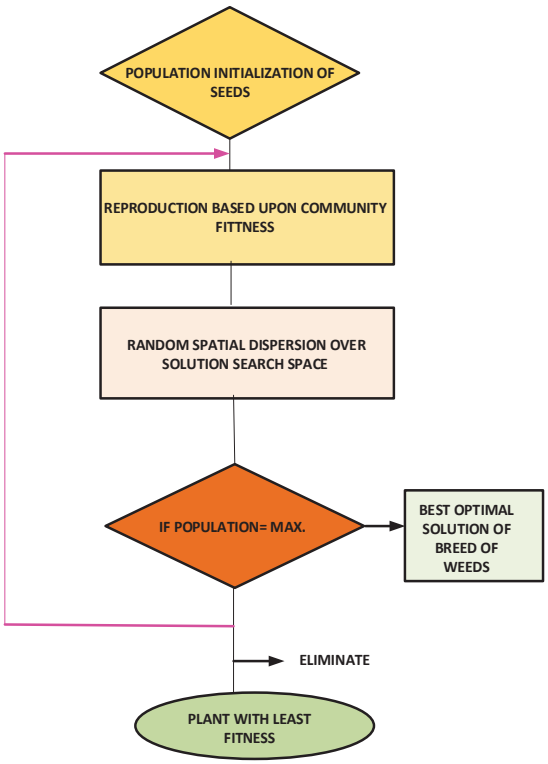


Fig. 8. IWO flowchart

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 18,2020 at 16:42:54 UTC from IEEE Xplore. Restrictions apply.

The loops of dispersion of seeds in search space works according to equitation (13) ensuring less chances to disperse in larger radial orbit in each successive iterations. The process includes elimination of less fit weeds, and revolution of their way of living from r to k force criterion. When maximum population is reached in community, weeds with highest fitness are segregated together and one with lower fitness are disqualified and not allowed to breed further. The outflow of IWO is elaborated in form of flow chart provided in Fig.8.

The tuned values of both the controller parameters are given in table.2.

TABLE II. PITCH AND YAW PROPELLERS OPTIMIZED VALUES FOR L-PID CONTROLLER

Rotor	Optimization parameter	Tuned Values
Pitch Controller	k_{p1}	6.95
	k_{i1}	8.923
	k_{d1}	7.784
Yaw Controller	k_{p2}	2.798
	k_{i2}	4.554
	k_{d2}	3.843

TABLE III. PITCH AND YAW PROPELLERS OPTIMIZED VALUES FOR FOPI^{λ} CONTROLLER

Rotor	Optimization Parameter	Tuned Values
Pitch Controller	k_{p1}	7.213
	k_{i1}	18.101
	λ_1	0.634
Yaw Controller	k_{p2}	4.051
	k_{i2}	19.544
	λ_2	0.863

These IWO tuned values of controller parameters are tested and validated on real time and stabilization graphs are obtained mentioned in section 5.

V. STABILIZATION RESULTS AND EXPERIMENTAL COMPARISION OF CONTROLLERS

The L-PID and FOPI^{λ} controller response on pitch and yaw propellers along with control signal are given in Fig.9, Fig.10, Fig.11, Fig.12

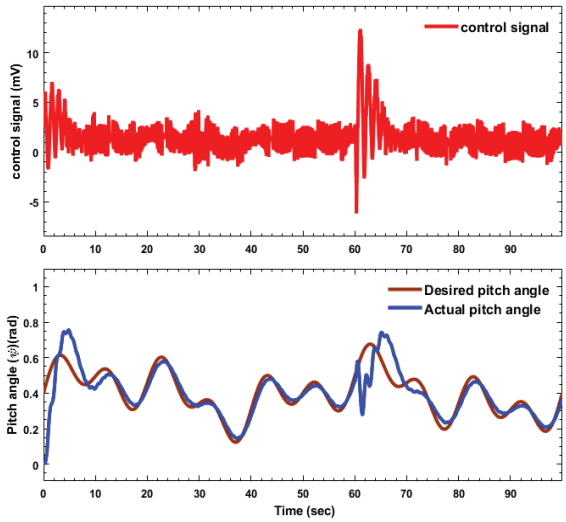


Fig. 9. L-PID controller Pitch angle response

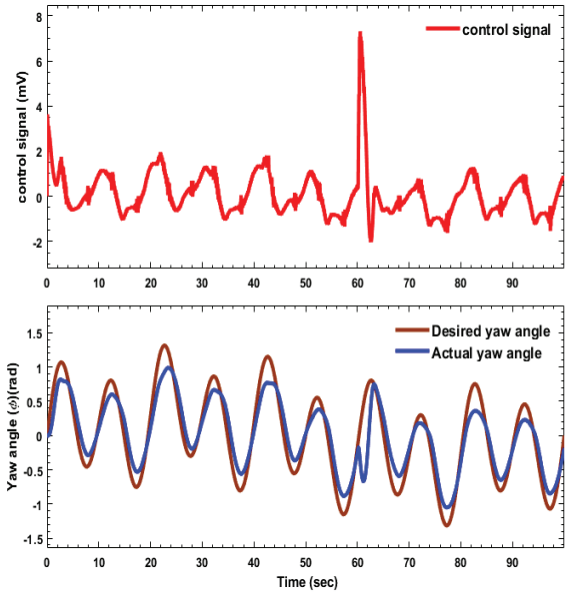


Fig. 10. L-PID Yaw angle Response

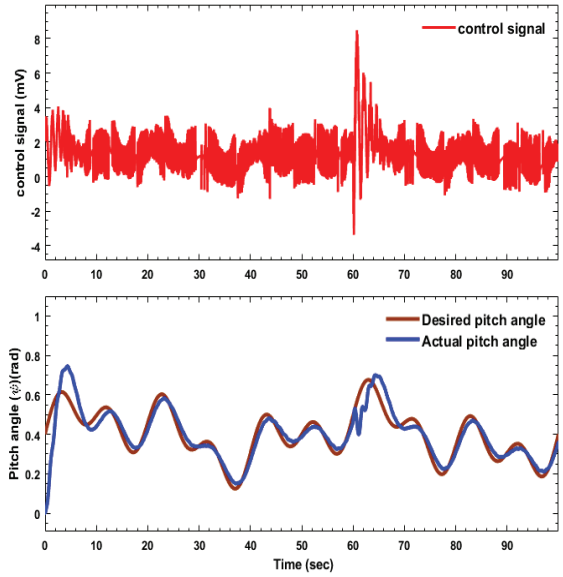


Fig. 11. FOPI^{λ} controller pitch response

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 18,2020 at 16:42:54 UTC from IEEE Xplore. Restrictions apply.

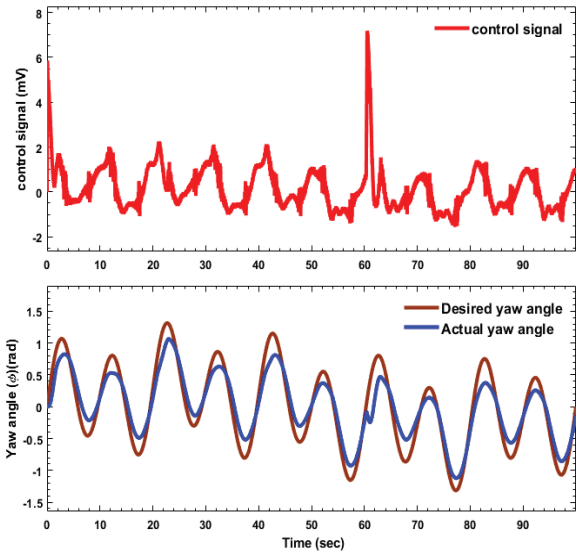


Fig. 12. FOPI^{λ} Controller yaw response

A comparison table describing time taken to stabilize after the given manual disturbance are given in table.4 and table.5 for L-PID and FOPI^{λ} controllers respectively.

TABLE IV. L-PID CONTROLLER RESPONSES

Rotor	**L-PID**		
Pitch	Distrubence Initiated (Sec)	Stabilization Instant (Sec)	Stabilization time (Sec)
	60	74	14
Yaw	60	65	5

TABLE V. FOPI^{λ} CONTROLLER RESPONSES

Rotor	***FOPI*$^{\lambda}$**		
Pitch	Distrubence Initiated (Sec)	Stabilization Instant (Sec)	Stabilization time (Sec)
	60	66	6
Yaw	60	62	2

It is observed from Table IV and Table V, that FOPI^{λ} controller takes less time to stabilize pitch and yaw angles of TRAS after the given manual disturbances and shows robust control signal then L-PID controller

VI. CONCLUSION

In this paper, optimal control of Twin rotor aerial system (TRAS) has been achieved by designing and real time experimental validation of L-PID and FOPI^{λ} controllers. These controller parameters are optimized between constraints through analytical metamorphic algorithm inspired from robust and stubborn colonization traits of invasive weeds. After experimental verification it is found that FOPI^{λ} shows better control behavior than L-PID controller for stabilization of pitch and yaw rotors of TRAS aerial plant, this signifies that it is also capable of receding wind and atmospheric disturbances on real helicopter and UAVs.

ACKNOWLEDGMENT

We would like express our warm gratitude to Indian government and DTU to provide the real time experimental plant of aerial system.

REFERENCES

[1] H. Shakhatreh *et al*., "Unmanned Aerial Vehicles (UAVs): A Survey on Civil Applications and Key Research Challenges," in *IEEE Access*, vol. 7, pp. 48572-48634, 2019.

[2] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, "A tutorial on UAVs for wireless networks: Applications, challenges, and open problems," arXiv preprint arXiv:1803.00680, 2018..

[3] Y. Sun, D. Wing Kwan Ng, D. Xu, L. Dai, and R. Schober, "Resource allocation for solar powered uav communication systems," in 2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), June 2018, pp. 1–5.

[4] G. Chmaj and H. Selvaraj, "Distributed processing applications for UAV/Drones: a survey," in Progress in Systems Engineering. Springer, 2015, pp. 449–454

[5] Ahmad, S. M., Chipperfield, A. J., & Tokhi, M. O. (2000). Dynamic modelling and optimal control of a twin rotor MIMO system. In Proceedings of the 2000 IEEE international conference on national aerospace and electronics (pp. 391–398)..

[6] Ahmad, S. M., Shaheed, M. H., Chipperfield, A. J., & Tokhi, M. O. (2000). Nonlinear modelling of a twin rotor MIMO system using radial basis function networks. In Proceedings of the 2000 IEEE international conference on national aerospace and electronics (pp. 313–320).

[7] Aldebrez, F. M., Mat-Darus, I. Z., & Tokhi, M. O. (2004). Dynamic modelling of a twin rotor system in hovering position. In Proceedings of the 2004 international symposium on control, communications and signal processing (pp. 823–826).

[8] Avila-Vilchis, J. C., Brogliato, B., Dzul, R., & Lozano, R. (2003). Nonlinear modelling and control of helicopters. Automatica, 39(9), 1583–1596.

[9] S. K. Pandey and V. Laxmi, "Control of twin rotor MIMO system using PID controller with derivative filter coefficient," *2014 IEEE Students' Conference on Electrical, Electronics and Computer Science*, Bhopal, 2014, pp. 1-6.

[10] V. S. Rao, V. I. George, S. Kamath and C. Shreesha, "Reliable robust PID controller design for TRMS," *2017 11th Asian Control Conference (ASCC)*, Gold Coast, QLD, 2017, pp. 565-569.

[11] S. K. Mishra and S. Purwar, "To design optimally tuned FOPID controller for twin rotor MIMO system," *2014 Students Conference on Engineering and Systems*, Allahabad, 2014, pp. 1-6.

[12] S. Chaudhary and A. Kumar, "Control of Twin Rotor MIMO System Using 1-Degree-of-Freedom PID, 2-Degree-of-Freedom PID and Fractional order PID Controller," *2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, 2019, pp. 746-751.

[13] H. N. Al-Duwaish and M. Bettayeb, "Control of nonlinear dynamical systems using genetic algorithms," *Proceedings of 12th IEEE International Symposium on Intelligent Control*, Istanbul, Turkey, 1997, pp. 421-424.

[14] P. Biswas, R. Maiti, A. Kolay, K. Das Sharma and G. Sarkar, "PSO based PID controller design for twin rotor MIMO system," *Proceedings of The 2014 International Conference on Control, Instrumentation, Energy and Communication (CIEC)*, Calcutta, 2014, pp. 56-60.

[15] Prithwish Chakraborty, Gourab Ghosh Roy, Swagatam Das and B. K. Panigrahi, "On population variance and explorative power of invasive weed optimization algorithm," *2009 World Congress on Nature & Biologically Inspired Computing (NaBIC)*, Coimbatore, 2009, pp. 227-232.

[16] D. Vyas and G. Parmar, "Reduced order modelling of ltisystems using invasive weed optimization," *2017 International Conference on Information, Communication, Instrumentation and Control (ICICIC)*, Indore, 2017, pp. 1-4.

[17] N. C. Patel, M. K. Debnath, B. K. Sahu, S. S. Dash and R. Bayindir, "Multi-Staged PID Controller Tuned by Invasive Weed optimization Algorithm for LFC Issues," *2018 7th International Conference on Renewable Energy Research and Applications (ICRERA)*, Paris, 2018, pp. 1358-1362

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on June 18,2020 at 16:42:54 UTC from IEEE Xplore. Restrictions apply.

View publication stats

Proceedings of the Fourth International Conference on Trends in Electronics and Informatics (ICOEI 2020)
IEEE Xplore Part Number: CFP20J32-ART; ISBN: 978-1-7281-5518-0

Realization of Regula-Falsi Iteration based Double Precision Floating Point Division

Riya Jain
Department of Electronics and Communication
Delhi Technological University
Delhi, India
riyajain_bt2k16@dtu.ac.in

Neeta Pandey
Department of Electronics and Communication
Delhi Technological University
Delhi, India
neetapandey@dce.ac.in

***Abstract*— One of the fundamental elements of modern microprocessor architectures is that they provide hardware support for arithmetic operations in floating point. In spite of the fact that division is a moderately rare operation in computer architecture, they are noteworthy and becoming progressively imperative in numerous modern applications. This paper presents design and implementation of Regula-Falsi (False Position) method for double precision floating point division. Floating point numbers are represented using IEEE 754 standard format. The presented algorithm iteratively computes reciprocal of 64-bit floating point divisor where two initial approximations are computed using ROM(s). The algorithm is implemented in Verilog Hardware Description Language (VHDL) and Zybo Zynq – 7000 Development FPGA Board is used to verify the same. For fairer comparison, the proposed method is compared with Newton Raphson counterpart, which is an existing algorithm that performs division operation iteratively. It has been proved that the proposed algorithm shows improvement in terms of number of iterations (computation cycles) and the accuracy of output.**

Keywords*— *Regula-Falsi, Newton-Raphson, division, multiplication, addition, floating-point, double precision, ROM

I. INTRODUCTION

In modern applications, wide range of numbers is processed, from small to large. Thus, for such applications, fixed point numbers and their arithmetic operations are unsuitable. This is where floating point numbers come into play and provide wider range of numbers. Although floating point numbers are just approximated real values but they can facilitate computation of extremely large values (astronomically large numbers) to extremely small values (atomically small numbers). Cost and performance of any system is affected by the type of number representation. Thus, representation of a number is an imperative thing in computer arithmetic. Depending upon the application and required performance, one can make a decision of which representation to be followed. Since floating point arithmetic [1] plays important role in wide range of applications, enhancing them makes a recognizable positive effect on the execution of the entire system. Among all basic arithmetic operations, division takes maximum clock cycles to compute the quotient for a given dividend and divisor. Floating point division algorithms [2] have changed over the last few decades. Most of them can be classified as subtractive (e.g. SRT division [3]-[8] and variable latency division [9]) or multiplicative (e.g. Newton-Raphson division [10], [11]) division algorithms. In subtractive floating point division, number of clock cycles is proportional to the word length of dividend. On the other hand, multiplicative algorithms perform operation by making initial approximations using look-up table [12]. Such algorithms are iterative in nature and their speed of convergence towards final result depends upon initial approximation.

In this paper, a new algorithm for floating point division using Regula-Falsi method is proposed. It computes the quotient by multiplying dividend with reciprocal of divisor. Reciprocal unit of the divider uses multiplication and addition/subtraction operations to compute the reciprocal of divisor. This method makes two initial approximations that are read from two 2^{10} x 10 ROMs and this value is updated in every iteration to compute the final result. Newton-Raphson is an existing iterative division algorithm which acts as a counterpart for the proposed algorithm. Newton-Raphson also makes an initial approximation, whose values are determined as given in [13], using one 2^{10} x 10 ROM. Both the algorithms are implemented in parallel for comparison.

The remaining of this paper is organized as follows: Section 2 comprises of mathematical analysis of the algorithm followed by its hardware implementation. This section also highlights the procedure of initial approximations. Section 3 makes comparison between the proposed algorithm and Newton-Raphson algorithm in terms of required computation cycles (iterations) and accuracy. Simulation results for an example are also placed in this section. Finally, conclusions are given in Section 4.

II. REGULA-FALSI ITERATIVE DIVISION USING LOOK-UP TABLE

Regula-Falsi, also called false position method is one of the equations solving technique used to find roots of any function in one variable. The method is simply a trial and error technique of using test ("false") values for the variable and then adjusting the test value according to the outcome.

A. Mathematical Analysis

Consider a function $f(x)$, shown in Fig. 1, that cuts the x-axis at point O. Two initial approximations x_0 and x_1 are made such that $f(x_0)$ is negative and $f(x_1)$ is positive.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:26:09 UTC from IEEE Xplore. Restrictions apply.

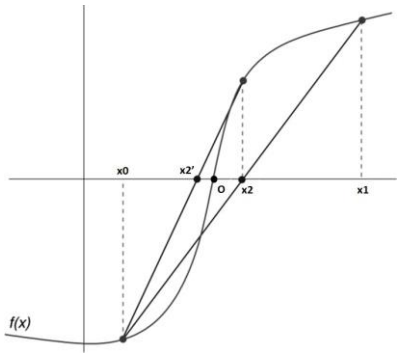


Fig. 1: Geometrical representation of the Regula-Falsi method

Construct the line through the points $(x0, f(x0))$ and $(x1, f(x1))$. This line is a chord of the graph of the function f and is represented in point-slope form (1).

$$y - f(x1) = \frac{f(x1) - f(x0)}{x1 - x0}(x - x1) \quad (1)$$

Considering $x2$ as x-intercept point of this line, i. e., y = 0 for $x = x2$, (1) may be rewritten as

$$f(x1) + \frac{f(x1) - f(x0)}{x1 - x0}(x2 - x1) = 0 \quad (2)$$

After simplifying,

$$x2 = \frac{x0\, f(x1) - x1\, f(x0)}{f(x1) - f(x0)} \quad (3)$$

Values of $x0$ and $x1$ are now updated according to sign of $f(x2)$. Thus there are two possible cases:

1. f(x2) > 0 ; x1 is replaced by x2 such that x0 remains the same.
2. f(x2) <0 ; x0 is replaced by x2 such that x1 remains the same.

After getting new values of $x0$ and $x1$, above process is repeated till two approximations get close together, and coincide (or almost coincide) at point O. To perform floating point division, both dividend (A) and divisor (B) are represented in IEEE 754 double precision binary floating point format. The quotient Q is obtained by multiplying A with reciprocal of B, given as:

$$Q = \frac{A}{B} \quad (4)$$

Or

$$Q = A\frac{1}{B} \quad (5)$$

Now, to compute the value of reciprocal $1/B$, let

$$f(x) = \frac{1}{x} - B \quad (6)$$

such that root of the function is at $x = 1/B$. With $x0$ and $x1$ as two initial approximations, substitute values of $f(x0)$ and $f(x1)$ in (3) to get the value of $x2$ as:

$$x2 = \frac{x0^2\,(1 - B\,x1) - x1^2\,(1 - B\,x0)}{x0 - x1} \quad (7)$$

$$x2 = \frac{x0^2 - x1^2 - B\,x1\,x2\,(x0 - x1)}{x0 - x1} \quad (8)$$

$$x2 = x0 + x1 - B\,x0\,x1 \quad (9)$$

In (9), two multiplications and two addition/ subtraction are used to compute $x2$. To check the sign of $f(x2)$, there are two cases:

1. If $Bx2 < 1$, then $f(x2) > 0$
2. If $Bx2 > 1$, then $f(x2) < 0$

When $x2$ coincides (or almost coincide) with point O, $x2$ $1/B$. Final result is obtained by substituting value of $1/B$ in (5).

B. Block Diagram

The block diagrammatic representation of the method is depicted in Fig. 2. The dotted block is the Reciprocal Unit that computes $1/B$ (1/divisor). The output of this block is multiplied with A (dividend) to get the final result.

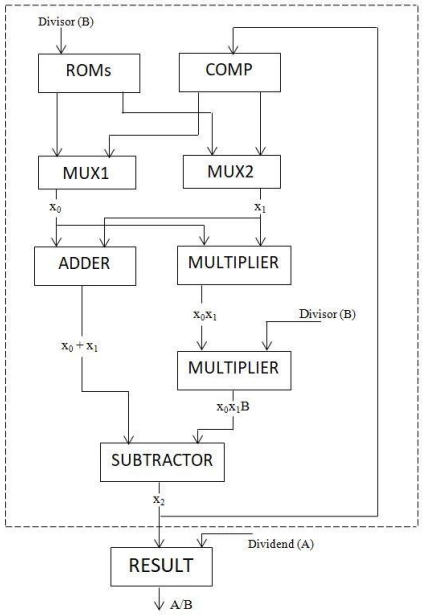


Fig. 2. Regula-Falsi Divider block diagram

In first clock cycle initial approximations $x0$ and $x1$, each of 10 bits, are obtained from two 2_{10} x 10 ROMs. Most significant 10 bits of the divisor (B) are truncated, and are

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:26:09 UTC from IEEE Xplore. Restrictions apply.

given as input to two ROMs that give 64-bit approximations of x_0 and x_1 at their output (detailed explanation of how these approximations are taken is in next section). In second clock cycle, initial values of x_0 and x_1, output of ROMs, are selected by MUX1 and MUX2. In next clock cycle, addition/subtraction and multiplication operations are performed to compute x_2 as in (9). The result obtained for x_2, is then sent as an input to COMP block to check the sign of $f(x_2)$ by computing Bx_2 and comparing the value with 1, as explained in previous section. Values of x_0 and x_1 are then updated and new values of x_0 and x_1 are selected by MUX1 and MUX2 in subsequent clock cycle. The process is repeated till reciprocal of divisor (B) approaches it's true value. Finally, in last clock cycle, end result is obtained at output by multiplying final value of x_2 with dividend (A). Similar approach is followed in Newton-Raphson algorithm where reciprocal of divisor (B) is computed and then multiplied with dividend (A) by taking an initial approximation of 10 bits. Even though they follow similar approach, it will be shown further that Regula-Falsi converges faster than Newton-Raphson and computes result with higher accuracy.

C. Look-Up Table Values

Look-up tables used in this method are of size 2_{10} x 10. Most significant 10-bits of the divisor (B) are truncated and used to find initial values of x_0 and x_1. They are given at input as address of two ROMs. 10-bits output of ROM1 and ROM2, corresponding to the address, determine the values of x_0 and x_1 respectively. 10-bits values stored in ROM can be determined as follows:

$$B = 1.\underbrace{x_1x_2x_3x_4x_5x_6x_7x_8x_9x_{10}}\ldots.x_{52} \quad (10)$$

$$B' = 1.x_1x_2x_3x_4x_5x_6x_7x_8x_9x_{10} \quad (11)$$

$$B' = 1 + x_1 2^{-1} + x_2 2^{-2} + x_3 2^{-3} + \ldots + x_{10} 2^{-10} \quad (12)$$

$$X' = \frac{1}{B'} \quad (13)$$

$$= 1.\underbrace{x'_1x'_2x'_3x'_4x'_5x'_6x'_7x'_8x'_9x'_{10}}\ldots.x'_{52}$$

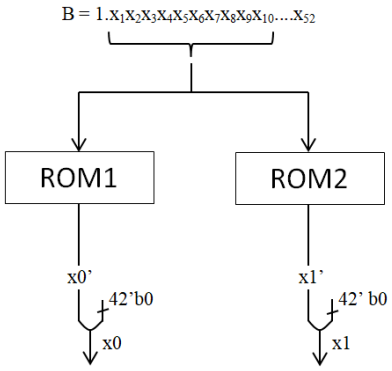


Fig. 3: Initial approximation using ROM

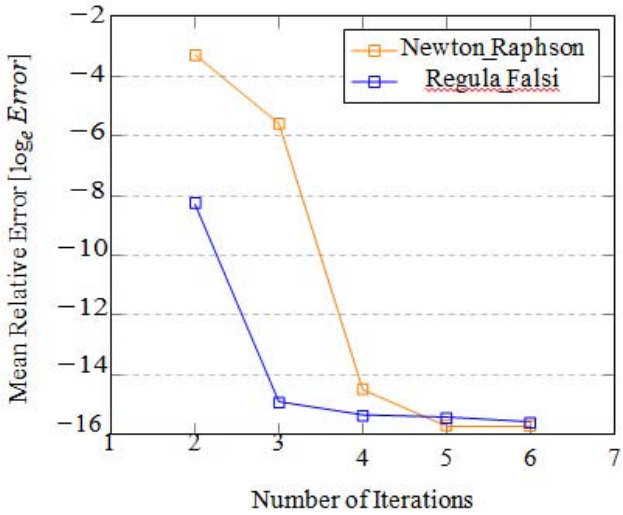


Fig. 4: Mean Relative Error

Values of ROM1:

$$x0' = x'_1x'_2x'_3x'_4x'_5x'_6x'_7x'_8x'_9x'_{10} - 0000000001$$

Values of ROM2:

$$x1' = x'_1x'_2x'_3x'_4x'_5x'_6x'_7x'_8x'_9x'_{10} + 0000000001$$

Finally, as shown in Fig. 3, x0 and x1 are obtained by appending 42 zeros to LSB of the output of ROM1 and ROM2 respectively to make it 52-bit significand of the floating point numbers.

III. Results and Comparison

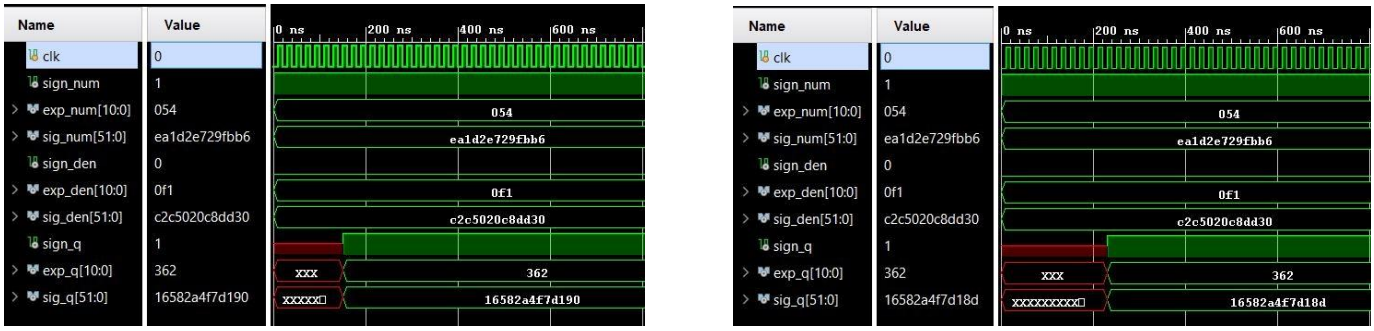
The proposed algorithm is implemented in Verilog HDL and simulated using ISIM-Xilinx. The FPFA board used is Zybo Zynq– 7000. Two examples are stated in Table 1 and Table 2 where N and D represent dividend and divisor respectively, Q_Actual denotes expected value of quotient for given dividend and divisor, Q_Regula denotes the quotient value obtained from Regula-Falsi method in 8 clock cycles, and Q_Newton_m denotes the quotient value obtained from Newton-Raphson method in 'm' clock cycles. In the first example, Newton-Raphson and Regula-Falsi method computes the result with $2.8e^{-16}\%$ error in 12 clock cycles and 8 clock cycles respectively while error increases to $-2.3e^{-10}\%$ for Newton-Raphson method when result is computed in 10 clock cycles. In second example, Newton-Raphson and Regula-Falsi method computes the result in 14 and 8 clock cycles respectively with 0% error while error increases to $-7.5e^{-11}\%$ for Newton-Raphson method when clock cycles are reduced to 12. For ready reference and comparison, simulation results of example 2 for the proposed method and Newton-Raphson method are placed in Fig. 4 and Fig. 5 respectively. It may be seen that in case of Regula-Falsi method, final result is obtained in 8 clock cycles: 7 clock cycles to compute reciprocal and 1 clock cycle to calculate the final result and in case of Newton Raphson method, final result is obtained in 14 clock cycles, which is 50% more than the proposed algorithm. For fairer comparison, simulations are performed for random 15 data sets and it is found that both the methods give similar approximated results

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:26:09 UTC from IEEE Xplore. Restrictions apply.

TABLE I: Example for floating point division by Regula-Falsi and Newton-Raphson method

Iterations	Regula-Falsi		Newton-Raphson	
	Output Quotient	MRE	Output Quotient	MRE
2	$-5.951592619252755e-48$	$5.6969893e-10$	$-5.950389005329300e-48$	$2.0223449e-04$
3	$-5.951592622643376e-48$	$8.1687894e-16$	$-5.951592379230435e-48$	$4.0898789e-08$
4	$-5.951592622643374e-48$	$4.0843947e-16$	$-5.951592622643361e-48$	$1.6337578e-15$
5	$-5.951592622643371e-48$	0	$-5.951592622643372e-48$	$2.0421973e-16$
6	$-5.951592622643371e-48$	0	$-5.951592622643372e-48$	$2.0421973e-16$

*Dividend(A) = $-4.119942981079925e^{-283}$, Divisor(B) = $6.922421009471027e^{-236}$, True value of quotient(Q) = $-5.951592622643371e^{-48}$



*sign xx: sign bit of xx, exp xx: exponent of xx, sig xx: significand of xx for xx = num (dividend), den (divisor), and q (quotient)

Fig. 5: Simulation results for both methods

but the proposed algorithm is better in terms of number of clock cycles and iterations required, that is, proposed method requires 4-6 clock cycles less than that required in Newton-Raphson method.

One of the samples, randomly selected from all samples, is stated as an example for ready reference in Table 1. For given value of dividend(A) and divisor(B), results of both methods for 2, 3, 4, 5, and 6 iterations are recorded in the table. Mean Relative Error (MRE) for these values with respect to true value of quotient(Q) are also mentioned. It is clear from the given example that it took 3 iterations (8 clock cycles) for Regula-Falsi method to nearly converge to the result while Newton-Raphson method took 5-6 iterations (12- 14 clock cycles). Simulation results of the same example (in hexadecimal representation) for Regula-Falsi (3 iterations) and Newton-Rapshson (5 iterations) are placed in Fig. 5 (a) and (b) respectively.

Possible explanation for this faster convergence in case of Regula-Falsi method is utilization of 2 initial approximation. Moreover, there is an advantage of using Regula-Falsi over Newton-Raphson method that it will always converge since two approximations are taken such that one lies on left side of the root and other on the right side. In Newton-Raphson method, the result of a function may diverge if the initial approximation is too far away from the root.

IV. Conclusion

The floating point division using Regula-Falsi iteration algorithm is presented in the paper. The proposed design computes the reciprocal by utilizing two look-up tables to obtain initial approximations. The algorithm takes eight cycles to compute the quotient for given double precision floating- point dividend and divisor with an MRE in a range of $e^{-}16$. A precision of 52-bit is obtained in the conducted experiment, but it is not guaranteed the correctly rounded least significant bit for all cases. The proposed algorithm is compared with Newton-Raphson algorithm of division and is verified to be faster with more accurate results. Number of clock cycles and iterations required in Regula-Falsi method is much less than that in Newton-Raphson method.

References

[1] B. Parhami, “Computer Arithmetic Algorithms And Hardware Designs,” 2nd ed., New York, USA: Oxford University Press, 2010.
[2] S.F. Oberman and M.J. Flynn, “An Analysis of Division Algorithms and Implementations,” Computer Systems Lab., Dept. of Elect. Eng. and Comp. Sci., Stanford Univ., Tech. Rep. CSL-TR-95-675, July 1995.
[3] P. Kornerup, “Digit Selection for SRT Division and Square Root”, IEEE Transactions on Computers, Vol. 54, No. 3, March 2005, pp. 294-303.
[4] L.S. Niwal, S. P. Hajare, “Design of Radix 4 Divider Circuit Using SRT Algorithm”, in ICCSP, Melmaruvathur, India, 2015, pp. 1107-1110.
[5] David M. Russinoff, “Computation and Formal verification of SRT Quotient and Square Root Digit Selection Tables”, in IEEE Transactions on Computers, Vol. 62, No. 5, 2013, pp. 900-913.
[6] D.L. Harris, S.F. Oberman, and M.A. Horowitz, “SRT Division Architectures and Implementations,” in Proceeding of 13th IEEE Sympsoium on Computer Arithmetic, Asilomar, USA, 1997, pp. 18-

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:26:09 UTC from IEEE Xplore. Restrictions apply.

25.
[7] A. Amaricai and O. Boncalo, "SRT Radix-2 Dividers with (5,4) Redundant Representation of Partial Remainder," in NORCHIP, Vilnius, Lithuania, 2013, pp. 1-5.
[8] P. Kornerup, "Revisiting SRT Quotient Digit Selection," in proceeding of 16th IEEE Symposium on Computer Arithmetic, 2003, pp. 38-45.
[9] P.M. Seidel, "Variable Latency Division," in MWSCAS, Windsor, Canada, 2018, pp. 944-947.
[10] N. Singh and T. N. Sasamal, "Design and Synthesis of Single Precision Floating Point Division based on Newton-Raphson Algorithm on FPGA", in MATEC Web of Conferences, ICAET, 2016.
[11] U. Kucukkabak and A. Akkas, "Design and Implementation of Reciprocal Unit Using Table Look-up and Newton-Raphson Iteration", in Euromicro Symposium on Digital System Design, Rennes, France, 2004.
[12] M. Ito, N. Takagi and S. Yajima, "Efficient Initial Approximation for Multiplicative Division and Square Root by a Multiplication with Operand Modification," in IEEE Transactions on Computers, Vol. 46, No. 4, 1997, pp. 495-498.
[13] D. Das Sarma and D.W. Matula, "Measuring The Accuracy Of ROM Reciprocal Tables," in Proceedings of IEEE 11th Symposium on Computer Arithmetic, Windsor, Canada, 2002, pp. 95-102.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 24,2020 at 09:26:09 UTC from IEEE Xplore. Restrictions apply.

Real-Time Implementation of the Optimal Predictor and Optimal Filter: Accuracy Versus Latency

SYED ASEEM UL ISLAM, ANKIT GOEL, and DENNIS S. BERNSTEIN

Summary

The Kalman filter uses noisy sensor measurements to estimate the states of a linear system that is subject to disturbances. An elegant implementation of this algorithm is the optimal two-step filter (OTSF), which involves a forecast step based on the dynamics and a data assimilation step based on sensor data. However, what happens if the order of the steps between state estimates is reversed? This variation of the OTSF, which is the optimal two-step predictor (OTSP), provides slightly earlier state estimates than the OTSF based on slightly older data. The goal of this article is to compare the accuracy of the OTSF and the OTSP to assess whether or not the delayed estimates of the OTSF are more accurate than the faster estimates of the OTSP.

Although the Kalman filter is often presented within a continuous-time context [1]–[4], the original derivation was provided in discrete time [5]. In practice, controllers and observers are invariably implemented digitally; thus discrete-time algorithms deserve special attention. This article focuses on the real-time implementation of the discrete-time Kalman filter and its relation to the discrete-time Kalman predictor. These algorithms differ in subtle ways, as highlighted in this article. Equations for the discrete-time Kalman filter and predictor are given in [6]. Unfortunately, the covariance matrix becomes indefinite after a few steps, which shows that those equations are erroneous. The present article corrects and extends the results from [6].

This derivation is based on necessary conditions and thus does not prove that the optimal one-step predictor (OOSP) is the globally optimal one. However, this derivation is succinct and uses minimal mathematics, making it efficient for classroom presentation. As shown in [2, p. 46], in the absence of Gaussian processes, this derivation yields the optimal linear predictor. Next, the OOSP is recast as an optimal two-step predictor (OTSP). The steps of the OTSP are then reversed to obtain the optimal two-step filter (OTSF), as described in "Summary." Finally, the steps of the OTSF are combined to obtain the optimal one-step filter (OOSF).

The OTSP and OTSF consist of assimilation and forecast updates. The OTSP begins with an assimilation update followed by a forecast update, whereas the OTSF begins with a forecast update followed by an assimilation update. An additional distinction between these two algorithms concerns their real-time implementation. Within the context of constraints on data collection and computation, the OTSP and OTSF use different data to produce state estimates that are available at different times. Overall, the OTSF uses more recent data but produces state estimates at a later time. It is thus expected that the OTSF estimates are more accurate than those of the OTSP, with latency the price paid for the enhanced accuracy of the filter.

To investigate the validity of the expected accuracy/latency tradeoff, the accuracy of the OTSP and OTSF is compared by means of a numerical example. Somewhat surprisingly, the numerical results show that the filter estimates are not uniformly better than the predictor estimates. This discrepancy is traced to the use of the initial data, suggesting a third algorithm: an optimal two-step filter that invokes an assimilation update at start-up (OTSFSU) to utilize the initial data. Aside from the initial start-up step, the OTSFSU is identical to the OTSF. We then revisit the accuracy/latency tradeoff by numerically comparing the accuracy of the OTSP and OTSFSU to determine whether the OTSFSU is uniformly more accurate than the OTSP. Spoiler alert: It is.

OPTIMAL ONE-STEP PREDICTOR

Consider the system

$$x_{k+1} = A_k x_k + B_k u_k + w_k, \tag{1}$$

$$y_k = C_k x_k + D_k u_k + v_k, \tag{2}$$

Digital Object Identifier 10.1109/MCS.2019.2961588
Date of current version: 17 March 2020

1066-033X/20©2020IEEE

Authorized licensed use limited to: University of Michigan Library. Downloaded on March 28,2020 at 16:38:43 UTC from IEEE Xplore. Restrictions apply.

Controllers and observers are invariably implemented digitally; thus discrete-time algorithms deserve special attention.

where, for all $k \geq 0$, $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^m$, $y_k \in \mathbb{R}^p$, $w_k \in \mathbb{R}^n$, and $v_k \in \mathbb{R}^p$, and the real matrices A_k, B_k, C_k, and D_k are of corresponding size. All stochastic processes are assumed to have finite second moments; the details of the densities are not relevant to the derivation below. Specifically, $w_k \in \mathbb{R}^n$ is a zero-mean, white-noise disturbance signal with the covariance $Q_k \triangleq \mathbb{E}[w_k w_k^{\mathrm{T}}] \in \mathbb{R}^{n\times n}$, and $v_k \in \mathbb{R}^p$ is a zero-mean, white sensor noise with the covariance $R_k \triangleq \mathbb{E}[v_k v_k^{\mathrm{T}}] \in \mathbb{R}^{p\times p}$. The cross-covariance between v_k and w_k is denoted by $S_k \triangleq \mathbb{E}[w_k v_k^{\mathrm{T}}] \in \mathbb{R}^{n\times p}$. The input u_k is assumed to be known; in practice, u_k is typically a control input. Consider the OOSP

$$\hat{x}_{k+1} = A_k \hat{x}_k + B_k u_k + K_k(y_k - \hat{y}_k), \tag{3}$$

$$\hat{y}_k = C_k \hat{x}_k + D_k u_k, \tag{4}$$

where $\hat{x}_k \in \mathbb{R}^n$ is an estimate of the state x_k; the optimal gain $K_k \in \mathbb{R}^{n\times p}$ is determined below.

Defining the state error

$$e_k \triangleq x_k - \hat{x}_k, \tag{5}$$

it follows that

$$\begin{aligned} e_{k+1} &= x_{k+1} - \hat{x}_{k+1} \\ &= \tilde{A}_k e_k + \tilde{w}_k, \end{aligned} \tag{6}$$

where

$$\tilde{A}_k \triangleq A_k - K_k C_k, \tag{7}$$

$$\tilde{w}_k \triangleq w_k - K_k v_k. \tag{8}$$

Next, we define the cost

$$\begin{aligned} J_k(K_k) &\triangleq \mathbb{E}[e_{k+1}^{\mathrm{T}} e_{k+1}] \\ &= \mathrm{tr}\,\mathbb{E}[e_{k+1}^{\mathrm{T}} e_{k+1}] \\ &= \mathbb{E}[\mathrm{tr}\,(e_{k+1}^{T} e_{k+1})] \\ &= \mathbb{E}[\mathrm{tr}(e_{k+1} e_{k+1}^{\mathrm{T}})] \\ &= \mathrm{tr}\,\mathbb{E}[e_{k+1} e_{k+1}^{\mathrm{T}}] \\ &= \mathrm{tr}\,P_{k+1}, \end{aligned} \tag{9}$$

where the *error covariance* is defined by

$$P_{k+1} \triangleq \mathbb{E}[e_{k+1} e_{k+1}^{\mathrm{T}}] \in \mathbb{R}^{n\times n}. \tag{10}$$

It thus follows that

$$\begin{aligned} e_{k+1} e_{k+1}^{\mathrm{T}} &= (\tilde{A}_k e_k + \tilde{w}_k)(\tilde{A}_k e_k + \tilde{w}_k)^{\mathrm{T}} \\ &= \tilde{A}_k e_k e_k^{\mathrm{T}} \tilde{A}_k^{\mathrm{T}} + \tilde{A}_k e_k \tilde{w}_k^{\mathrm{T}} + \tilde{w}_k e_k^{\mathrm{T}} \tilde{A}_k^{\mathrm{T}} + \tilde{w}_k \tilde{w}_k^{\mathrm{T}}. \end{aligned} \tag{11}$$

Next, since w_k and v_k are white-noise sequences that affect e_{k+1} but not e_k, it follows that e_k and $\tilde{w}_k$ are uncorrelated. Furthermore, since w_k and v_k have a zero mean, it follows that $\tilde{w}_k$ also has a zero mean. Therefore, $\mathbb{E}[e_k \tilde{w}_k^{\mathrm{T}}] = \mathbb{E}[e_k]\mathbb{E}[\tilde{w}_k^{\mathrm{T}}] = 0$. Taking the expected value of (11) yields

$$\begin{aligned} P_{k+1} &= \mathbb{E}[e_{k+1} e_{k+1}^{\mathrm{T}}] \\ &= \mathbb{E}[\tilde{A}_k e_k e_k^{\mathrm{T}} \tilde{A}_k^{\mathrm{T}} + \tilde{A}_k e_k \tilde{w}_k^{\mathrm{T}} + \tilde{w}_k e_k^{\mathrm{T}} \tilde{A}_k^{\mathrm{T}} + \tilde{w}_k \tilde{w}_k^{\mathrm{T}}] \\ &= \tilde{A}_k \mathbb{E}[e_k e_k^{\mathrm{T}}] \tilde{A}_k^{\mathrm{T}} + \tilde{A}_k \mathbb{E}[e_k \tilde{w}_k^{\mathrm{T}}] + \mathbb{E}[\tilde{w}_k e_k^{T}] \tilde{A}_k^{\mathrm{T}} + \mathbb{E}[\tilde{w}_k \tilde{w}_k^{\mathrm{T}}] \\ &= \tilde{A}_k P_k \tilde{A}_k^{\mathrm{T}} + \mathbb{E}[\tilde{w}_k \tilde{w}_k^{\mathrm{T}}] \\ &= \tilde{A}_k P_k \tilde{A}_k^{\mathrm{T}} + \mathbb{E}[(w_k - K_k v_k)(w_k - K_k v_k)^{\mathrm{T}}] \\ &= \tilde{A}_k P_k \tilde{A}_k^{\mathrm{T}} + \mathbb{E}[w_k w_k^{\mathrm{T}} - w_k v_k^{\mathrm{T}} K_k^{\mathrm{T}} - K_k v_k w_k^{\mathrm{T}} + K_k v_k v_k^{\mathrm{T}} K_k^{\mathrm{T}}] \\ &= \tilde{A}_k P_k \tilde{A}_k^{\mathrm{T}} + \mathbb{E}[w_k w_k^{\mathrm{T}}] - \mathbb{E}[w_k v_k^{\mathrm{T}}] K_k^{\mathrm{T}} - K_k \mathbb{E}[v_k w_k^{\mathrm{T}}] \\ &\quad + K_k \mathbb{E}[v_k v_k^{\mathrm{T}}] K_k^{\mathrm{T}} \\ &= \tilde{A}_k P_k \tilde{A}_k^{\mathrm{T}} + Q_k + K_k R_k K_k^{\mathrm{T}} - \mathbb{E}[w_k v_k^{\mathrm{T}}] K_k^{\mathrm{T}} - K_k [\mathbb{E}[w_k v_k^{\mathrm{T}}]]^{\mathrm{T}}. \end{aligned} \tag{12}$$

The cost (9) is thus given by

$$\begin{aligned} J_k(K_k) &= \mathrm{tr} P_{k+1} \\ &= \mathrm{tr}[(A_k - K_k C_k) P_k (A_k - K_k C_k)^{\mathrm{T}} + Q_k + K_k R_k K_k^{\mathrm{T}} \\ &\quad - S_k K_k^{\mathrm{T}} - K_k S_k^{\mathrm{T}}] \\ &= \mathrm{tr}[A_k P_k A_k^{\mathrm{T}} - K_k C_k P_k A_k^{\mathrm{T}} - A_k P_k C_k^{\mathrm{T}} K_k^{\mathrm{T}} + K_k C_k P_k C_k^{\mathrm{T}} K_k^{\mathrm{T}} \\ &\quad + Q_k + K_k R_k K_k^{\mathrm{T}} - S_k K_k^{\mathrm{T}} - K_k S_k^{\mathrm{T}}] \\ &= \mathrm{tr}[K_k (C_k P_k C_k^{\mathrm{T}} + R_k) K_k^{\mathrm{T}} - K_k C_k P_k A_k^{\mathrm{T}} - A_k P_k C_k^{\mathrm{T}} K_k^{\mathrm{T}} \\ &\quad + A_k P_k A_k^{\mathrm{T}} + Q_k - S_k K_k^{\mathrm{T}} - K_k S_k^{\mathrm{T}}] \\ &= \mathrm{tr}[K_k (C_k P_k C_k^{\mathrm{T}} + R_k) K_k^{\mathrm{T}}] - 2\mathrm{tr}\, K_k (C_k P_k A_k^{\mathrm{T}} + S_k^{\mathrm{T}}) \\ &\quad + \mathrm{tr}(A_k P_k A_k^{\mathrm{T}} + Q_k). \end{aligned} \tag{13}$$

To minimize $J_k(K_k)$, note that

$$\frac{\mathrm{d}J_k(K_k)}{\mathrm{d}K_k} = 2(C_k P_k C_k^{\mathrm{T}} + R_k) K_k^{\mathrm{T}} - 2C_k P_k A_k^{\mathrm{T}} - 2S_k^{\mathrm{T}}. \tag{14}$$

Setting the derivative to zero yields the *optimal one-step predictor gain*

$$K_k = (A_k P_k C_k^{\mathrm{T}} + S_k)(C_k P_k C_k^{\mathrm{T}} + R_k)^{-1}. \tag{15}$$

By substituting (7) and (15) in (12), the *one-step predictor* is thus given by

$$\hat{x}_{k+1} = A_k \hat{x}_k + B_k u_k + (A_k P_k C_k^{\mathrm{T}} + S_k)(C_k P_k C_k^{\mathrm{T}} + R_k)^{-1}(y_k - C_k \hat{x}_k), \tag{16}$$

$$P_{k+1} = A_k P_k A_k^{\mathrm{T}} - (A_k P_k C_k^{\mathrm{T}} + S_k)(C_k P_k C_k^{\mathrm{T}} + R_k)^{-1} \cdot (C_k P_k A_k^{\mathrm{T}} + S_k^{\mathrm{T}}) + Q_k. \tag{17}$$

Authorized licensed use limited to: University of Michigan Library. Downloaded on March 28,2020 at 16:38:43 UTC from IEEE Xplore. Restrictions apply.

Optimal Two-Step Predictor and Two-Step Filter With Correlated Disturbance and Sensor Noise

This sidebar gives the update equations for the optimal two-step predictor (OTSP) and optimal two-step filter (OTSF) in the case where $S_k \neq 0$. For the OTSP, the *assimilation update* is given by

$$x_k^{\mathrm{a}} = x_k^{\mathrm{f}} + P_k^{\mathrm{f}} C_k^{\mathrm{T}} (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1} (y_k - C_k x_k^{\mathrm{f}}), \tag{S1}$$

$$P_k^{\mathrm{a}} = P_k^{\mathrm{f}} - P_k^{\mathrm{f}} C_k^{\mathrm{T}} (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1} C_k P_k^{\mathrm{f}}, \tag{S2}$$

which are identical to (20) and (21). The *forecast update* is given by

$$x_{k+1}^{\mathrm{f}} = A_k x_k^{\mathrm{a}} + B_k u_k + S_k (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1} (y_k - C_k x_k^{\mathrm{f}}), \tag{S3}$$

$$\begin{aligned} P_{k+1}^{\mathrm{f}} = {} & A_k P_k^{\mathrm{a}} A_k^{\mathrm{T}} - S_k (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1} C_k P_k^{\mathrm{f}} A_k^{\mathrm{T}} \\ & - A_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1} S_k^{\mathrm{T}} \\ & - S_k (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1} S_k^{\mathrm{T}} + Q_k, \end{aligned} \tag{S4}$$

which are (22) and (23) with additional terms involving S_k. Defining the *OTSP gain*

$$K_k^{\mathrm{f}} \triangleq P_k^{\mathrm{f}} C_k^{\mathrm{T}} (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1}, \tag{S5}$$

(S1) and (S2) can be written as

$$x_k^{\mathrm{a}} = x_k^{\mathrm{f}} + K_k (y_k - C_k x_k^{\mathrm{f}}), \tag{S6}$$

$$P_k^{\mathrm{a}} = (I - K_k C_k) P_k^{\mathrm{f}}, \tag{S7}$$

which are identical to (25) and (26).

For the OTSF, the *forecast update* is given by

$$x_{k+1}^{\mathrm{f}} = A_k x_k^{\mathrm{a}} + B_k u_k + S_k (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1} (y_k - C_k x_k^{\mathrm{f}}), \tag{S8}$$

$$\begin{aligned} P_{k+1}^{\mathrm{f}} = {} & A_k P_k^{\mathrm{a}} A_k^{\mathrm{T}} - S_k (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1} C_k P_k^{\mathrm{f}} A_k^{\mathrm{T}} \\ & - A_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1} S_k^{\mathrm{T}} \\ & - S_k (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1} S_k^{\mathrm{T}} + Q_k, \end{aligned} \tag{S9}$$

which are (27) and (28) with additional terms involving S_k. The *assimilation update* is given by

$$\begin{aligned} x_{k+1}^{\mathrm{a}} = {} & x_{k+1}^{\mathrm{f}} + P_{k+1}^{\mathrm{f}} C_{k+1}^{T} (C_{k+1} P_{k+1}^{\mathrm{f}} C_{k+1}^{T} + R_{k+1})^{-1} \\ & \cdot (y_{k+1} - C_{k+1} x_{k+1}^{\mathrm{f}}), \end{aligned} \tag{S10}$$

$$P_{k+1}^{\mathrm{a}} = P_{k+1}^{\mathrm{f}} - P_{k+1}^{\mathrm{f}} C_{k+1}^{\mathrm{T}} (C_{k+1} P_{k+1}^{\mathrm{f}} C_{k+1}^{\mathrm{T}} + R_{k+1})^{-1} C_{k+1} P_{k+1}^{\mathrm{f}}, \tag{S11}$$

which are identical to (29) and (30). In terms of the *OTSF gain*

$$K_{k+1}^{\mathrm{f}} \triangleq P_{k+1}^{\mathrm{f}} C_{k+1}^{\mathrm{T}} (C_{k+1} P_{k+1}^{\mathrm{f}} C_{k+1}^{\mathrm{T}} + R_{k+1})^{-1}, \tag{S12}$$

(S10) and (S11) can be written as

$$x_{k+1}^{\mathrm{a}} = x_{k+1}^{\mathrm{f}} + K_{k+1} (y_{k+1} - C_{k+1} x_{k+1}^{\mathrm{f}}), \tag{S13}$$

$$P_{k+1}^{\mathrm{a}} = (I - K_{k+1} C_{k+1}) P_{k+1}^{\mathrm{f}}, \tag{S14}$$

which are identical to (32) and (33).

The OTSF in the traditional notation is

$$\begin{aligned} \hat{x}_{k+1|k} = {} & A_k \hat{x}_{k|k} + B_k u_k + S_k (C_k P_{k|k-1} C_k^{T} + R_k)^{-1} \\ & \cdot (y_k - C_k \hat{x}_{k|k-1}), \end{aligned} \tag{S15}$$

$$\begin{aligned} P_{k+1|k} = {} & A_k P_{k|k} A_k^{T} - S_k (C_k P_{k|k-1} C_k^{T} + R_k)^{-1} \\ & \cdot C_k P_{k|k-1} A_k^{T} - A_k P_{k|k-1} C_k^{T} (C_k P_{k|k-1} C_k^{T} + R_k)^{-1} S_k^{T} \\ & - S_k (C_k P_{k|k-1} C_k^{T} + R_k)^{-1} S_k^{T} + Q_k, \end{aligned} \tag{S16}$$

$$\begin{aligned} \hat{x}_{k+1|k+1} = {} & \hat{x}_{k+1|k} + P_{k+1|k} C_{k+1}^{T} (C_{k+1} P_{k+1|k} C_{k+1}^{T} + R_{k+1})^{-1} \\ & \cdot (y_{k+1} - C_{k+1} \hat{x}_{k+1|k}), \end{aligned} \tag{S17}$$

$$\begin{aligned} P_{k+1|k+1} = {} & P_{k+1|k} - P_{k+1|k} C_{k+1}^{T} (C_{k+1} P_{k+1|k} C_{k+1}^{T} + R_{k+1})^{-1} \\ & \cdot C_{k+1} P_{k+1|k}, \end{aligned} \tag{S18}$$

where (S15) and (S16) are (38) and (39) with additional terms involving S_k, and (S17) and (S18) are identical to (40) and (41). The OTSF gain (S12) can be written as

$$K_{k+1}^{\mathrm{f}} = P_{k+1|k}^{\mathrm{f}} C_{k+1}^{\mathrm{T}} (C_{k+1} P_{k+1|k}^{\mathrm{f}} C_{k+1}^{\mathrm{T}} + R_{k+1})^{-1}, \tag{S19}$$

and thus (S17) and (S18) can be written as

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1}^{\mathrm{f}} (y_{k+1} - C_{k+1} \hat{x}_{k+1|k}), \tag{S20}$$

$$P_{k+1|k+1} = (I - K_{k+1}^{\mathrm{f}} C_{k+1}) P_{k+1|k}, \tag{S21}$$

which are identical to (43) and (44).

Note that the error-covariance propagation equation is independent of data. Furthermore, (16) and (17) can be written in terms of K_k as

$$\hat{x}_{k+1} = A_k \hat{x}_k + B_k u_k + K_k (y_k - C_k \hat{x}_k), \tag{18}$$

$$P_{k+1} = A_k P_k A_k^{\mathrm{T}} - K_k (C_k P_k A_k^{\mathrm{T}} + S_k^{\mathrm{T}}) + Q_k. \tag{19}$$

OPTIMAL TWO-STEP PREDICTOR

As an alternative but equivalent implementation of the optimal predictor, (16) and (17) can be implemented as the OTSP. For simplicity in this and subsequent sections, consider the case where $S_k = 0$. The case where S_k is nonzero is discussed in "Optimal Two-Step Predictor and Optimal Two-Step Filter With Correlated Disturbance and Sensor Noise." The *assimilation update* is given by

$$x_k^{\mathrm{a}} = x_k^{\mathrm{f}} + P_k^{\mathrm{f}} C_k^{\mathrm{T}} (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1} (y_k - C_k x_k^{\mathrm{f}}), \tag{20}$$

$$P_k^{\mathrm{a}} = P_k^{\mathrm{f}} - P_k^{\mathrm{f}} C_k^{\mathrm{T}} (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1} C_k P_k^{\mathrm{f}}, \tag{21}$$

and the *forecast update* is given by

$$x_{k+1}^{\mathrm{f}} = A_k x_k^{\mathrm{a}} + B_k u_k, \tag{22}$$

$$P_{k+1}^{\mathrm{f}} = A_k P_k^{\mathrm{a}} A_k^{\mathrm{T}} + Q_k. \tag{23}$$

Figure 1 shows a timing diagram for the real-time implementation of the OTSP. By substituting (20) into (22) and (21) into (23), it can be seen that $x_k^{\mathrm{f}} = \hat{x}_k$ and $P_k^{\mathrm{f}} = P_k$. Defining the *OTSP gain*

$$K_k^{\mathrm{f}} \triangleq P_k^{\mathrm{f}} C_k^{\mathrm{T}} (C_k P_k^{\mathrm{f}} C_k^{\mathrm{T}} + R_k)^{-1}, \tag{24}$$

(20) and (21) can be written as

Authorized licensed use limited to: University of Michigan Library. Downloaded on March 28,2020 at 16:38:43 UTC from IEEE Xplore. Restrictions apply.

$$x_k^{\mathrm{a}} = x_k^{\mathrm{f}} + K_k(y_k - C_k x_k^{\mathrm{f}}), \tag{25}$$

$$P_k^{\mathrm{a}} = (I - K_k C_k) P_k^{\mathrm{f}}. \tag{26}$$

It is interesting to note that the computational requirements of the OTSP are more burdensome than the OOSP, and yet the OTSP and OOSP produce identical state estimates. Therefore, there is no implementation advantage of the OTSP over the OOSP. However, the value of the OTSP is in providing a framework for the OTSF, as discussed next.

OPTIMAL TWO-STEP FILTER

In contrast to the OTSP in (20)–(23), the OTSF is given by the *forecast update*

$$x_{k+1}^{\mathrm{f}} = A_k x_k^{\mathrm{a}} + B_k u_k, \tag{27}$$

$$P_{k+1}^{\mathrm{f}} = A_k P_k^{\mathrm{a}} A_k^{\mathrm{T}} + Q_k, \tag{28}$$

and the *assimilation update*

$$x_{k+1}^{\mathrm{a}} = x_{k+1}^{\mathrm{f}} + P_{k+1}^{\mathrm{f}} C_{k+1}^{\mathrm{T}} (C_{k+1} P_{k+1}^{\mathrm{f}} C_{k+1}^{\mathrm{T}} + R_{k+1})^{-1} \cdot (y_{k+1} - C_{k+1} x_{k+1}^{\mathrm{f}}), \tag{29}$$

$$P_{k+1}^{\mathrm{a}} = P_{k+1}^{\mathrm{f}} - P_{k+1}^{\mathrm{f}} C_{k+1}^{\mathrm{T}} (C_{k+1} P_{k+1}^{\mathrm{f}} C_{k+1}^{\mathrm{T}} + R_{k+1})^{-1} C_{k+1} P_{k+1}^{\mathrm{f}}. \tag{30}$$

Note that (20) and (21) are slightly different from (29) and (30); specifically, (20) and (21) use the measurement y_k, whereas (29) and (30) use the measurement y_{k+1}. Moreover, the forecast and assimilation updates of the OTSF appear in reverse order compared to the forecast and assimilation updates of the OTSP. This reversal explains the index k in (20) and (21) and the index $k+1$ in (29) and (30).

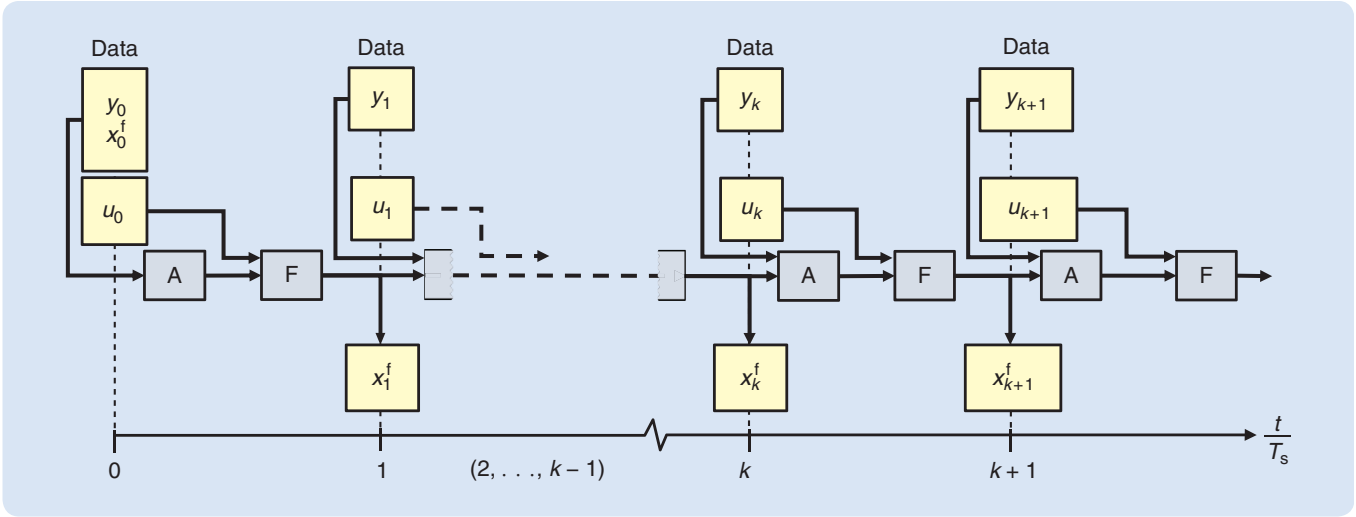


FIGURE 1 The timing diagram for the optimal two-step predictor (OTSP), where A denotes an assimilation update and F denotes a forecast update. The measurement y_k and applied control u_k are available at step k for computation. For all $k \geq 1$, the OTSP produces an estimate of x_k at step k without latency.

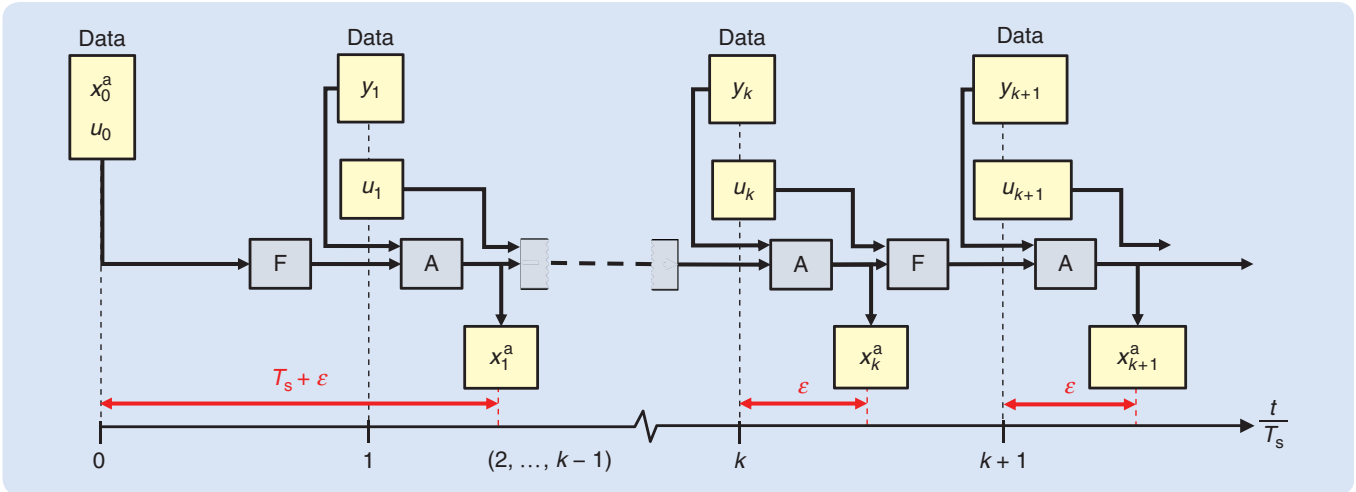


FIGURE 2 The timing diagram for the optimal two-step filter (OTSF), where A denotes an assimilation update and F denotes a forecast update. The measurement y_k and applied control u_k are available at step k for computation. For all $k \geq 1$, the OTSF produces an estimate of x_k with latency ε.

Authorized licensed use limited to: University of Michigan Library. Downloaded on March 28,2020 at 16:38:43 UTC from IEEE Xplore. Restrictions apply.

Next, note that, since x^{a}_{k+1} depends on y_{k+1} and the required computation cannot be performed instantaneously, the estimate x^{a}_{k+1} of x_{k+1} is not available at step $k+1$ but at some time afterward, where the latency depends on the computer speed and architecture. Therefore, if the latency of x^{a}_k as an estimate of x_k is critical in a real-time application, the OTSP may be a better choice than the OTSF. Figure 2 provides a timing diagram for the real-time implementation of the OTSF.

In terms of the *OTSF gain*

$$K^{\mathrm{f}}_{k+1} \triangleq P^{\mathrm{f}}_{k+1}C^{\mathrm{T}}_{k+1}(C_{k+1}P^{\mathrm{f}}_{k+1}C^{\mathrm{T}}_{k+1}+R_{k+1})^{-1}, \tag{31}$$

(29) and (30) can be written as

$$x^{\mathrm{a}}_{k+1} = x^{\mathrm{f}}_{k+1} + K_{k+1}(y_{k+1} - C_{k+1}x^{\mathrm{f}}_{k+1}), \tag{32}$$

$$P^{\mathrm{a}}_{k+1} = (I - K_{k+1}C_{k+1})P^{\mathrm{f}}_{k+1}. \tag{33}$$

Note that, aside from a shift in the index, the OTSF gain is identical to the OTSP gain.

OPTIMAL ONE-STEP FILTER

The OOSF can be obtained by substituting (27) and (28) into (29) and (30). Specifically, the OOSF is given by

$$\begin{aligned} x^{\mathrm{a}}_{k+1} = {} & A_k x^{\mathrm{a}}_k + B_k u_k + (A_k P^{\mathrm{a}}_k A^{\mathrm{T}}_k + Q_k)C^{\mathrm{T}}_{k+1} \\ & \cdot [C_{k+1}(A_k P^{\mathrm{a}}_k A^{\mathrm{T}}_k + Q_k)C^{\mathrm{T}}_{k+1} + R_{k+1}]^{-1} \\ & \cdot [y_{k+1} - C_{k+1}(A_k x^{\mathrm{a}}_k + B_k u_k)], \end{aligned} \tag{34}$$

$$\begin{aligned} P^{\mathrm{a}}_{k+1} = {} & A_k P^{\mathrm{a}}_k A^{\mathrm{T}}_k + Q_k - (A_k P^{\mathrm{a}}_k A^{\mathrm{T}}_k + Q_k) \\ & \cdot C^{\mathrm{T}}_{k+1}[C_{k+1}(A_k P^{\mathrm{a}}_k A^{\mathrm{T}}_k + Q_k)C^{\mathrm{T}}_{k+1} + R_{k+1}]^{-1} \\ & \cdot C_{k+1}(A_k P^{\mathrm{a}}_k A^{\mathrm{T}}_k + Q_k). \end{aligned} \tag{35}$$

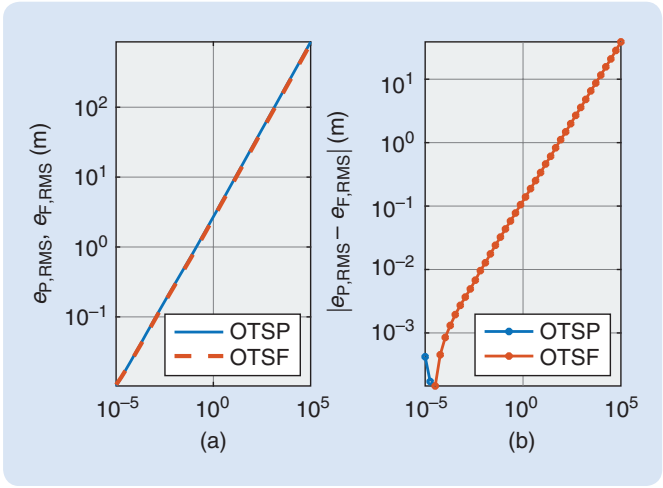


FIGURE 3 The root-mean-square (RMS) position-estimation errors for the optimal two-step predictor (OTSP) and optimal two-step filter (OTSF) versus the disturbance-covariance scaling α_{d}. Part (a) shows the RMS position-estimation errors for the OTSP and OTSF, $e_{\mathrm{P,RMS}}$ and $e_{\mathrm{F,RMS}}$, respectively. Part (b) shows $|e_{\mathrm{P,RMS}} - e_{\mathrm{F,RMS}}|$. In all cases where $e_{\mathrm{P,RMS}} > e_{\mathrm{F,RMS}}$, the OTSF is more accurate than the OTSP; these are shown in red. In all cases where $e_{\mathrm{P,RMS}} < e_{\mathrm{F,RMS}}$, the OTSP is more accurate than the OTSF; these are shown in blue. Note that the OTSP is more accurate than the OTSF for $\alpha_{\mathrm{d}} < 10^{-4.5}$, whereas the OTSF is more accurate than the OTSP for $\alpha_{\mathrm{d}} > 10^{-4.5}$.

OPTIMAL TWO-STEP FILTER AND TRADITIONAL KALMAN FILTER

We now restate the OTSF in standard notation to show that it is precisely the Kalman filter. Defining

$$\hat{x}_{k|k-1} \triangleq x^{\mathrm{f}}_k, \quad P_{k|k-1} \triangleq P^{\mathrm{f}}_k, \tag{36}$$

$$\hat{x}_{k|k} \triangleq x^{\mathrm{a}}_k, \quad P_{k|k} \triangleq P^{\mathrm{a}}_k, \tag{37}$$

(27)–(30) can be written as

$$\hat{x}_{k+1|k} = A_k\hat{x}_{k|k} + B_k u_k, \tag{38}$$

$$P_{k+1|k} = A_k P_{k|k} A^{\mathrm{T}}_k + Q_k, \tag{39}$$

$$\begin{aligned} \hat{x}_{k+1|k+1} = {} & \hat{x}_{k+1|k} + P_{k+1|k}C^{\mathrm{T}}_{k+1}(C_{k+1}P_{k+1|k}C^{\mathrm{T}}_{k+1} + R_{k+1})^{-1} \\ & \cdot (y_{k+1} - C_{k+1}\hat{x}_{k+1|k}), \end{aligned} \tag{40}$$

$$\begin{aligned} P_{k+1|k+1} = {} & P_{k+1|k} - P_{k+1|k}C^{\mathrm{T}}_{k+1}(C_{k+1}P_{k+1|k}C^{\mathrm{T}}_{k+1} + R_{k+1})^{-1} \\ & \cdot C_{k+1}P_{k+1|k}, \end{aligned} \tag{41}$$

which is the standard notation for the Kalman filter [2]. Furthermore, the OTSF gain (31) can be written as

$$K^{\mathrm{f}}_{k+1} = P^{\mathrm{f}}_{k+1|k}C^{\mathrm{T}}_{k+1}\left(C_{k+1}P^{\mathrm{f}}_{k+1|k}C^{\mathrm{T}}_{k+1} + R_{k+1}\right)^{-1}, \tag{42}$$

and thus (40) and (41) can be written as

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K^{\mathrm{f}}_{k+1}(y_{k+1} - C_{k+1}\hat{x}_{k+1|k}), \tag{43}$$

$$P_{k+1|k+1} = (I - K^{\mathrm{f}}_{k+1}C_{k+1})P_{k+1|k}. \tag{44}$$

An alternative notation [1], [3], [4] is given by

$$\hat{x}^{+}_k \triangleq x^{\mathrm{f}}_k, \quad P^{+}_k \triangleq P^{\mathrm{f}}_k, \tag{45}$$

$$\hat{x}^{-}_k \triangleq x^{\mathrm{a}}_k, \quad P^{-}_k \triangleq P^{\mathrm{a}}_k. \tag{46}$$

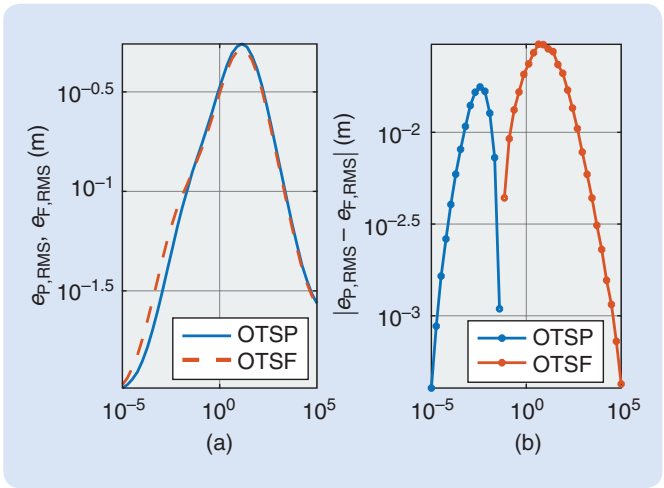


FIGURE 4 The root-mean-square (RMS) position-estimation errors for the optimal two-step predictor (OTSP) and optimal two-step filter (OTSF) versus the sensor-noise covariance α_{sn}. Part (a) shows the RMS position-estimation errors for the OTSP and OTSF, $e_{\mathrm{P,RMS}}$ and $e_{\mathrm{F,RMS}}$, respectively. Part (b) shows that the OTSP is more accurate than the OTSF for $\alpha_{\mathrm{sn}} < 10^{-2}$, whereas the OTSF is more accurate than the OTSP for $\alpha_{\mathrm{sn}} > 10^{-2}$.

Authorized licensed use limited to: University of Michigan Library. Downloaded on March 28,2020 at 16:38:43 UTC from IEEE Xplore. Restrictions apply.

OPTIMAL TWO-STEP PREDICTOR VERSUS OPTIMAL TWO-STEP FILTER

In this section, we investigate whether or not the OTSF estimates are more accurate than the OTSP estimates. Consider an undamped oscillator with a mass of 4 kg and stiffness of 2 N/m, sampled at 4 Hz using a zero-order hold input. The velocity is measured, and the position is to be estimated. The sampled-data system representing this arrangement has the dynamics

$$\begin{bmatrix} x_{k+1} \\ \dot{x}_{k+1} \end{bmatrix} = \begin{bmatrix} 0.9844 & 0.2487 \\ -0.1243 & 0.9844 \end{bmatrix} \begin{bmatrix} x_k \\ \dot{x}_k \end{bmatrix} + \begin{bmatrix} 0.0078 \\ 0.0622 \end{bmatrix} u_k + w_k, \tag{47}$$

$$y_k = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} x_k \\ \dot{x}_k \end{bmatrix} + v_k, \tag{48}$$

where $x_k \in \mathbb{R}$ is the position in meters, $y_k = \dot{x}_k \in \mathbb{R}$ is the velocity in meters/second, u_k is the applied force in Newtons, $w_k \in \mathbb{R}^2$ is the disturbance, and $v_k \in \mathbb{R}$ is the sensor noise.

To investigate the effect of the disturbance w_k, sensor noise v_k, and initial condition x_0 on the estimates of the position produced by the OTSP and OTSF, let

$$u_k \sim \mathcal{N}(0,1),\ w_k \sim \mathcal{N}(0, \alpha_{\mathrm{d}} I),\ v_k \sim \mathcal{N}(0, \alpha_{\mathrm{sn}}),\ x_0 \sim \mathcal{N}(0, \alpha_{\mathrm{ic}} I), \tag{49}$$

where $\alpha_{\mathrm{d}}, \alpha_{\mathrm{sn}}, \alpha_{\mathrm{ic}}$ are varied individually, with the remaining variables fixed. Specifically, we first vary $\alpha_{\mathrm{d}} \in [10^{-5}, 10^{5}]$, with $\alpha_{\mathrm{sn}} = \alpha_{\mathrm{ic}} = 10^{-5}$ fixed. Next, we vary $\alpha_{\mathrm{sn}} \in [10^{-5}, 10^{5}]$, with $\alpha_{\mathrm{d}} = \alpha_{\mathrm{ic}} = 10^{-5}$ fixed. Finally, we vary $\alpha_{\mathrm{ic}} \in [10^{-5}, 10]$, with $\alpha_{\mathrm{d}} = \alpha_{\mathrm{sn}} = 10^{-5}$ fixed. In all cases, the initial states of the OTSP and OTSF are set to zero and $P_0^{\mathrm{f}} = I_2$, $P_0^{\mathrm{a}} = I_2$.

In each case and for each choice of $\alpha_{\mathrm{d}}, \alpha_{\mathrm{sn}}, \alpha_{\mathrm{ic}}$, 10,000 simulations are run with randomly generated values of u_k, w_k, and v_k and the initial condition x_0. For each simulation, the root mean square (RMS) position-estimation errors for the OTSP and OTSF are computed for $0 \le k \le 80$. The RMS position-estimation errors for the OTSP and OTSF are then averaged across the 10,000 simulations. The RMS position-estimation errors for the OTSP and OTSF are defined to be $e_{\mathrm{P,RMS}}(\alpha_{\mathrm{d}}, \alpha_{\mathrm{sn}}, \alpha_{\mathrm{ic}})$ and $e_{\mathrm{F,RMS}}(\alpha_{\mathrm{d}}, \alpha_{\mathrm{sn}}, \alpha_{\mathrm{ic}})$, respectively. Specifically, we compute

$$e_{\mathrm{P,RMS}} = \frac{1}{10{,}000} \sum_{j=1}^{10{,}000} \sqrt{\frac{1}{80} \sum_{i=1}^{80} (x_{i,j}^{\mathrm{f}} - x_{i,j})^2}, \tag{50}$$

$$e_{\mathrm{F,RMS}} = \frac{1}{10{,}000} \sum_{j=1}^{10{,}000} \sqrt{\frac{1}{80} \sum_{i=1}^{80} (x_{i,j}^{\mathrm{a}} - x_{i,j})^2}, \tag{51}$$

where $x_{i,j}^{\mathrm{f}}$ is the position estimate produced by the OTSP at the ith step for the jth simulation, $x_{i,j}^{\mathrm{a}}$ is the position estimate produced by the OTSF at the ith step for the jth simulation,

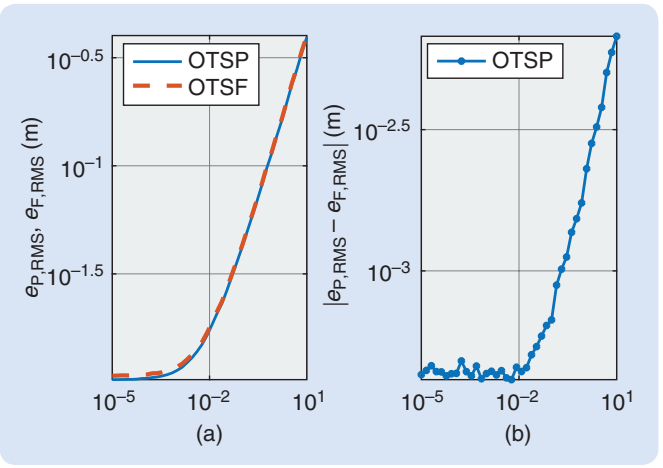


FIGURE 5 The root-mean-square (RMS) position-estimation errors for the optimal two-step predictor (OTSP) and optimal two-step filter (OTSF) versus the initial-condition-covariance scaling, α_{ic}. Part (a) shows the RMS position-estimation errors for the OTSP and OTSF, $e_{\mathrm{P,RMS}}$ and $e_{\mathrm{F,RMS}}$, respectively. Part (b) shows that the OTSP is uniformly more accurate than the OTSF.

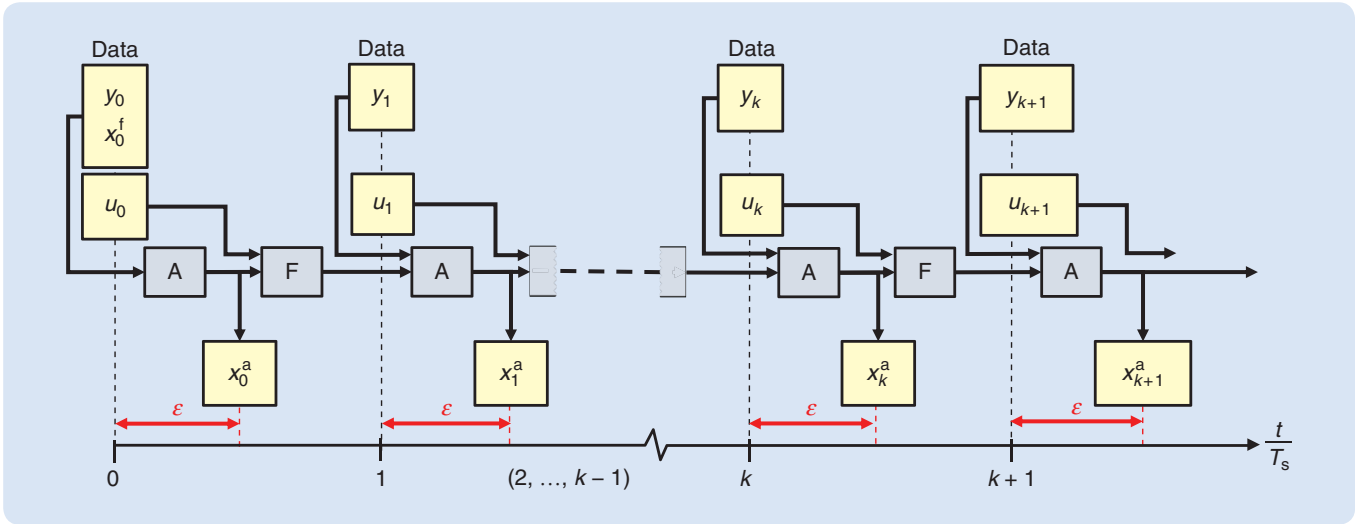


FIGURE 6 The timing diagram for the optimal two-step filter with start-up (OTSFSU), where A denotes an assimilation update and F denotes a forecast update. The measurement y_k and applied control u_k are available at step k for computation. For all $k \ge 0$, the OTSFSU produces an estimate of x_k with latency ε.

Authorized licensed use limited to: University of Michigan Library. Downloaded on March 28,2020 at 16:38:43 UTC from IEEE Xplore. Restrictions apply.

The filter can be realized only as a two-step algorithm, whereas the predictor can be realized as either a one- or two-step algorithm.

and $x_{i,j}$ is the true position at the ith step for the jth simulation. Furthermore, $|e_{P,RMS} - e_{F,RMS}|$ is plotted to show the accuracy of the OTSF position estimate relative to the OTSP position estimate in meters. Specifically, $e_{P,RMS} - e_{F,RMS}$ is positive in the case where the OTSF RMS position-estimation error is smaller than the OTSP RMS position-estimation error and vice versa.

For the three cases considered above, Figures 3(a), 4(a), and 5(a) show $e_{P,RMS}, e_{F,RMS}$ versus $\alpha_d, \alpha_{sn}, \alpha_{ic}$, respectively. Figures 3(b), 4(b), and 5(b) graph $|e_{P,RMS} - e_{F,RMS}|$ versus $\alpha_d, \alpha_{sn}, \alpha_{ic}$, respectively, where the values of $|e_{P,RMS} - e_{F,RMS}|$ are color coded based on the sign of $e_{P,RMS} - e_{F,RMS}$. Figure 5 shows that the OTSP position estimate is more accurate than the OTSF position estimate for all choices of α_{ic}. This situation is surprising since we expect the latency of the OTSF estimates to be offset by greater accuracy. We thus seek a variation of the OTSF that produces the expected improved accuracy in return for latency, as discussed in the next section.

TABLE 1 The initial estimate $\hat{x}_0$ and real-time data y and u used by the estimator to approximate x_k. The time at which the estimate of x_k becomes available is given in terms of the step k, sample time T_s, and latency ε.

Estimator	$\hat{x}_0$	y Data	u Data	$\hat{x}_k$	When Available?
OTSP	x_0^f	y_{k-1}	u_{k-1}	x_k^f	kT_s
OTSF	x_0^a	y_k	u_{k-1}	x_k^a	$kT_s + \varepsilon$
OTSFSU	x_0^f	y_k	u_{k-1}	x_k^a	$kT_s + \varepsilon$

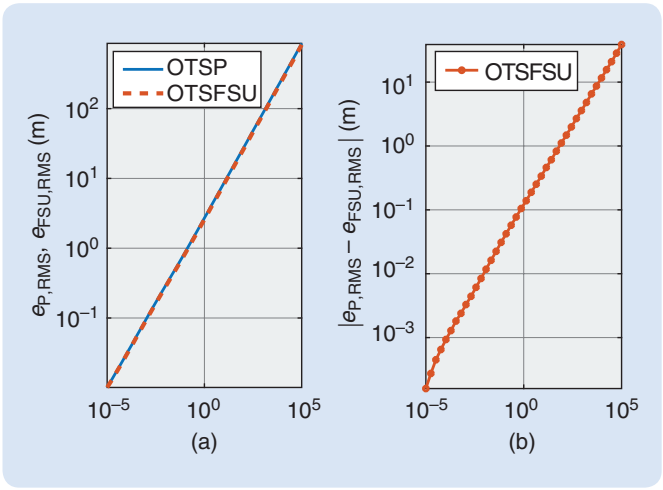


FIGURE 7 The root-mean-square (RMS) position-estimation errors for the optimal two-step predictor (OTSP) and optimal two-step filter with start-up (OTSFSU) versus the disturbance-covariance scaling, α_d. Part (a) shows the RMS position-estimation errors for the OTSP and OTSFSU $e_{P,RMS}$, and $e_{FSU,RMS}$, respectively. Part (b) shows that the OTSFSU is uniformly more accurate than the OTSP.

OPTIMAL TWO-STEP FILTER WITH START-UP

As shown in Figure 5, the OTSP may be more accurate than the OTSF. This is surprising because the OTSF produces latency estimates and uses more recent data, leading to the expectation of more accurate state estimates. This phenomenon can be traced to the fact that OTSF does not use the measurement y_0. In fact, at step $k = 0$ (20) uses y_0, whereas at step $k = 0$ (29) does not use y_0. This situation suggests the possibility of a variant of the OTSF that employs an additional assimilation update before the initial forecast update; otherwise, all subsequent updates of the OTSFSU are identical to the OTSF. Figure 6 presents a timing diagram for the real-time implementation of the OTSFSU. The next section investigates the accuracy of the OTSFSU compared to the OTSP to determine if the OTSFSU is more accurate, as compensation for its inherent latency.

REAL-TIME IMPLEMENTATION

To clarify the timing of the computation required for the OTSP, OTSF, and OTSFSU, note that for the OTSP the state estimate is x_k^f, whereas for the OTSF and OTSFSU the state estimate is x_k^a. Figures 1, 2, and 6 show timing diagrams for the real-time implementation of the OTSP, OTSF, and OTSFSU. Note that the OTSP is the only estimator that produces an estimate of x_k at step k. In contrast, the OTSF and OTSFSU produce estimates of x_k at time $kT_s + \varepsilon$, where T_s is the sample time of the digital implementation and $\varepsilon \in (0, T_s)$ is the latency (which is platform dependent). As pictured in Figure 2, the first state estimate produced by the OTSF is an estimate of x_1; this estimate is not available until $t = T_s + \varepsilon$. In addition, this estimate of x_1 does not use the data y_0, as can be seen by the absence of y_0 in Figure 2. Alternately, as displayed in Figure 6, the OTSFSU uses y_0 to produce an estimate of x_0 at $t = \varepsilon$. Table 1 lists the data used by each algorithm and the associated latency.

OPTIMAL TWO-STEP PREDICTOR VERSUS OPTIMAL TWO-STEP FILTER WITH START-UP

We now compare the OTSP and OTSFSU using the same procedure employed to compare the OTSP and OTSF. For the OTSFSU, define $e_{FSU,RMS}(\alpha_d, \alpha_{sn}, \alpha_{ic})$, and plot $|e_{P,RMS} - e_{FSU,RMS}|$. Specifically, compute

Authorized licensed use limited to: University of Michigan Library. Downloaded on March 28,2020 at 16:38:43 UTC from IEEE Xplore. Restrictions apply.

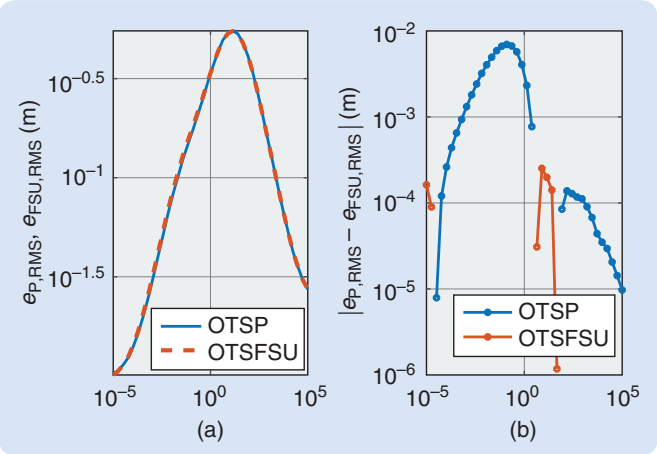


FIGURE 8 The root-mean-square (RMS) position-estimation errors for the optimal two-step predictor (OTSP) and optimal two-step filter with start-up (OTSFSU) versus the sensor-noise-covariance scaling, α_{sn}. Part (a) shows the RMS position-estimation errors for the OTSP and OTSFSU, $e_{P,RMS}$ and $e_{FSU,RMS}$, respectively. Part (b) shows that the OTSP is more accurate than the OTSFSU for $10^{-4.5} < \alpha_{sn} < 10^{-2}$ and $\alpha_{sn} > 10^{2}$.

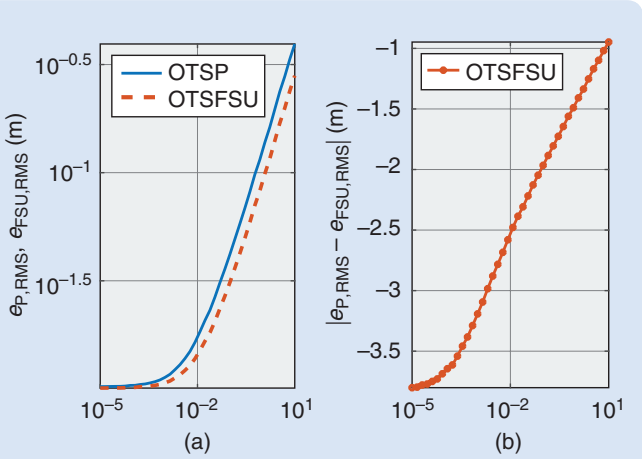


FIGURE 9 The root-mean-square (RMS) position-estimation errors for the optimal two-step predictor (OTSP) and optimal two-step filter with start-up (OTSFSU) versus the initial-condition-covariance scaling, α_{ic}. Part (a) shows the RMS position-estimation errors for the OTSP and OTSFSU, $e_{P,RMS}$ and $e_{FSU,RMS}$, respectively. Part (b) shows that the OTSFSU is uniformly more accurate than the OTSP.

$$e_{FSU,RMS} = \frac{1}{10,000} \sum_{j=1}^{10,000} \sqrt{\frac{1}{80} \sum_{i=1}^{80} (x_{i,j}^{a} - x_{i,j})^2}, \tag{52}$$

where $x_{i,j}^{a}$ is the position estimate produced by OTSFSU at the ith step for the jth simulation, and $x_{i,j}$ is the true position at the ith step for the jth simulation. Figures 7–9 show that the OTSFSU position estimate is more accurate than the OTSP position estimate for all choices of α_d and α_{ic}. Furthermore, as displayed in Figure 8, the OTSP shows no distinct advantage over the OTSFSU for variations of α_{sn}.

CONCLUSIONS

The goal of this article was to untangle the relationship between the OTSP and OTSF. A fundamental distinction between the filter and the predictor is that the filter can be realized only as a two-step algorithm, whereas the predictor can be realized as either a one- or two-step algorithm. An analysis by a timing diagram shows that the estimates produced by the OTSF take advantage of more recent data but have latency due to the time needed for computation. Partially contradicting the expected latency/accuracy tradeoff, a numerical example showed that the OTSF is not uniformly better than the OTSF. This discrepancy was traced to the nonuse of initial data by the OTSF, which suggested a variation of the OTSF that implements an additional assimilation update at start-up; otherwise the OTSF and the OTSFSU are identical. Revisiting the numerical example showed that the OTSFSU is uniformly more accurate than the OTSP. As with the OTSF, this additional accuracy comes at the price of latency.

ACKNOWLEDGMENTS

This research was partially supported by the Air Force Office of Scientific Research under Dynamic Data-Driven Applications Systems grant FA9550-16-1-0071 (http://www.1dddas.org/).

AUTHOR INFORMATION

Syed Aseem Ul Islam (aseemisl@umich.edu) received the B.S. degree in aerospace engineering from the Institute of Space Technology, Islamabad, Pakistan, and he is pursuing the Ph.D. degree in flight dynamics and control from the University of Michigan, Ann Arbor. His research interests include data-driven adaptive control for aerospace applications.

Ankit Goel received the B.E. degree in mechanical engineering from the Delhi College of Engineering, India, and the Ph.D. degree from the University of Michigan, Ann Arbor. His research interests include the data-driven estimation and control of high-dimensional complex systems.

Dennis S. Bernstein received the Sc.B. degree from Brown University, Providence, Rhode Island, and the Ph.D. degree from the University of Michigan, Ann Arbor, where he is a professor in the Aerospace Engineering Department. His research interests include the identification, estimation, and control for aerospace applications. He is the author of *Scalar, Vector, and Matrix Mathematics*, published by Princeton University Press.

REFERENCES

[1] A. Jazwinski, *Stochastic Processes and Filtering Theory*. New York: Elsevier, 1970.

[2] B. Anderson and J. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice Hall, 1979.

[3] J. Crassidis and J. Junkins, *Optimal Estimation of Dynamic Systems*. Boca Raton, FL: CRC Press, 2004.

[4] D. Simon, *Optimal State Estimation: Kalman, H-Infinity, and Nonlinear Approaches*. Hoboken, NJ: Wiley, 2006.

[5] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, 1960. doi: 10.1115/1.3662552.

[6] B. Teixeira, "Kalman filters," *IEEE Control Syst. Mag.*, vol. 28, no. 2, pp. 16–18, Apr. 2008. doi: 10.1109/MCS.2008.4472371.

Authorized licensed use limited to: University of Michigan Library. Downloaded on March 28,2020 at 16:38:43 UTC from IEEE Xplore. Restrictions apply.

The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/1477-7835.htm

Selection of sustainable transport system: a case study

Abhishek Singh
Department of Civil and Environmental Engineering, Delhi Technological University, Delhi, India
Amulya Gurtu
Cofrin School of Business, University of Wisconsin Green Bay, Green Bay, Wisconsin, USA, and
Rajesh Kumar Singh
Management Development Institute, Gurgaon, India

Received 30 March 2020
Revised 28 June 2020
Accepted 28 June 2020

Abstract

Purpose – Sustainable mobility will be the key to the survival of mankind in the 21st century. Cities with debt-ridden and poorly managed transport systems have to change to the ever-growing demands of the public transport system. The low cost of transport has been a key factor in sustainable development for any city. This study is trying to propose a framework for the selection of sustainable transport in context to an Indian case of NCR Delhi, India.

Design/methodology/approach – The present study has identified eight criteria for the selection of a sustainable transport system. Criteria for selection of sustainable transport are CO_2 emissions, cost of fuel, energy efficiency, cost of maintenance, number of accidents, congestion, number of injuries and road noise. Three alternatives of transport considered for this study are state-run bus, pooled car and *Shuttl* (App-based buses). The analytic hierarchy process (AHP) is used to prioritize the criteria. The study has further illustrated a framework for the selection of sustainable transport based on these criteria.

Findings – It is observed that CO_2 emissions are the most important criterion for a sustainable transport system. It is followed by a reduction in congestion and the number of injuries. Three alternatives of transport considered for this study are state-run bus, pooled car and *Shuttl* (a service provider). *Shuttl* has been found to be the most sustainable transport system.

Research limitations/implications – AHP is not able to capture the vagueness in decision-making. Therefore, fuzzy AHP can be considered for further detailed analysis as future scope of study in a different context of a sustainable transportation system. Major implications for policymakers and stakeholders are that development of public transport in cities should be done after considering different dimensions of sustainable operations.

Originality/value – The study has proposed a unique framework for the selection of a sustainable transport system by the public based on sustainability criteria. Findings will help policymakers in formulating strategies for developing sustainable transport system.

Keywords Transport system, City commute, Mobility, Sustainability, Analytic hierarchy process (AHP)

Paper type Research paper

1. Introduction

Transportation, besides a growing population and industrialization, increased the pollution significantly (Nakamura, 2004). Therefore, transportation systems need to be environmentally sustainable (Gupta and Singh, 2020). Even manufacturing organizations are trying to use the concept of reverse logistics for ensuring a sustainable transportation system (Agrawal and Singh, 2019). Sustainable transportation meets the requirements of the present without creating problems for future transportation systems (Black, 1996). Usually, a sustainable transport system works on the principle of the triple bottom line (e.g., World

The authors would like to thank reviewers and editor of the journal for valuable comments to improve the quality of this manuscript.

Management of Environmental Quality: An International Journal
© Emerald Publishing Limited
1477-7835
DOI 10.1108/MEQ-03-2020-0059

Bank, 1996; Loo, 2002; Schipper, 2003). The public transportation system plays an essential social function, and it needs to be designed responsibly for stopping global climate change (Haase and Nuissi, 2007). Traditionally, transportation systems focus on performance metrics such as vehicle mobility and congestion, and sustainability measures are not given due weightage (Zheng *et al.*, 2013). Without having an effective performance framework, the selection of a sustainable transportation system is a big challenge (Awasthi *et al.*, 2011). A sustainable transport system must provide mobility and accessibility to all urban residents in an environmentally friendly manner. This is a complex and challenging task when the needs and demands of people belonging to different income groups are not only different but also often conflicting. A mode of transportation is decided by individuals based on economic criteria, trip time involved, convenience, comfort and safety. According to May *et al.* (2001), sustainable transportation should ensure neat and clean streets, safety to the environment and support to a vibrant economy.

Zhao *et al.* (2020) have observed that sustainable transportation-related issues have emerged as an important topic of research. For a sustainable transportation system, long-term strategies need to be developed (Singh and Agrawal, 2018). Therefore, research in this area is highly significant for the economic growth of developing countries like India. Vehicles on roads create maximum urban air pollution. Air pollution is a major problem in metro cities. National Capital Region of Delhi, India (NCR), is considered one of the worse affected metro cities in the world due to vehicular pollution. It is being observed that after the lockdown in India due to COVID 19 pandemic in 2020, pollution level has drastically reduced in NCR Delhi and other metro cities. Therefore, controlling road transportation and the selection of optima alternatives by the public may play a vital role in controlling environmental pollution. The population and income growth in Indian cities over the last few decades have a marked effect on the mobility of its citizens. India's transport infrastructure demand has increased by almost eight times since 1980, which is higher than any other economy in Asia. This growth has been positive in many ways, including the development of thriving auto industry, growth of public infrastructure and allied economic growth. Fifteen of the top 20 polluted cities in the world are from India (Hindu, 2019).

Additionally, many cities in India are also ranked amongst the most congested cities in the world. The average speed for vehicles in some metro cities is reported as low as 17 km/h (10.6 m.p.h.). Congestion results in the reduction of productivity, fuel waste and pollution. A high-level estimate of the economic loss due to congestion, for the top four metros of India, is over USD 22 billion annually. Multiple efforts at the city, state and national levels have been taken to alleviate traffic congestion to improve productivity and to reduce pollution.

As an example, the Pradhan Mantri Gram Sadak Yojana (PMGSY) is one of the best examples to make the commute between rural and urban areas of India better. The emphasis on this scheme in the last three years has led to a record construction of all-weather roads at a rate of 130 km/day (81.3 miles/day). To address the twin challenges of pollution and congestion in urban India, several initiatives have been launched in various cities. Further, multiple national policy initiatives, such as the National Urban Transport Policy and the National Electric Mobility Mission Plan 2020, have been designed for improving mobility. While there are multiple examples of such initiatives, the magnitude and complexity of our mobility challenges require a comprehensive action. If India continues to grow with present carbon emission average rate of 6.8%, which was the trend between 2006 and 2015, it will supersede the European Union's total emissions by 2020 (EC-JRC report, 2016). Delhi, Gurgaon, Noida, Faridabad and Ghaziabad constitute the NCR of Delhi, India. NCR is a big and extended city with a population of about 46 million. Almost 10 million people commute to and from neighboring cities daily for different purposes, making it even more populated. The city roads are congested and do not serve the purpose of mobility for the increasing population. The distances between the office complexes and residential complexes are 20–

50 km (12–31 miles) in NCR, and most of the people prefer driving to work. Although Metro has reduced the problem to some extent, it does not provide last-mile connectivity. The government is trying to develop different alternatives of transport in metro cities, but sustainability is yet not the biggest concern for policymakers. Therefore, people use road transport choosing from the limited options like state-run buses, carpools or *Shuttl* (App-based office buses). The selection of sustainable transport by the public is a challenge, and it requires awareness also. This study identified the criteria for sustainable transport using analytical hierarchy process (AHP) and suggests a framework for the selection of the most sustainable transportation system in NCR Delhi, India. The AHP methodology and why it was used are explained in Section 3 (Research Methodology). The remainder of the paper has been organized as follows. Section 2 provides the literature review, Section 3 provides the research methodology, Section 4 shares the results and discusses them and, finally, Section 5 concludes the discussion with some limitations and possible future research.

2. Literature review

The primary objective of transportation is to move people, equipment and supplies safely and efficiently from one place to another (Gurtu *et al.*, 2019). The transport sector is the biggest contributor of the emissions in the USA and accounted for 29% of all emissions in 2017. Further, light-duty vehicles, which include passenger vehicles, caused 59% of emissions in the USA in 2017 (US EPA, 2020). Therefore, it is necessary to pay attention to sustainability in transportation. The concept of sustainable transport stemmed from sustainable development (Hall, 2002). Stefanieca *et al.* (2020) assessed the sustainability of inland transportation in China. Sustainable transport has two definitions. The first focuses on environmental problems and resource depletion, while the second encompasses social and economic welfare. The latter is preferred as it allows us to take a comprehensive view of all the effects of the transport sector (Litman and Burwell, 2006) and encourages people to seek integrated solutions to sustainable transport (Zhou, 2012). A sustainable transportation system is less costly to society. It should meet the transportation needs of all the sections of society. A sustainable transport system leads to sustainable development. According to Daly (1991), for sustainable development, the use of renewable and non-renewable resources should be less than its regeneration. "Another corollary to this definition is that the rate of generating pollution (to air, water, and land) must be kept the same or less than the time required to absorb and completely disintegrate that pollution in the environment" (Gurtu, 2014). In other words, emissions should be assimilated into the environment Mohan and Tiwari (1999). The transportation sector accounts for approximately 25% of worldwide carbon dioxide emissions (Van den Berg and De Langen, 2017).

In metro cities, most people use their own cars (Cosgrove *et al.*, 2009), and it causes traffic congestion (Arnott and Small, 1994). Transport sector contributes significantly to CO_2 emissions and climate change, but serious efforts have not been made by policymakers to control this problem (Marsden and Rye, 2010). Cruz and Katz-Gerro (2016) have tried to analyze few cases of transport companies and found that selection of green technologies and green practices by transport companies is influenced by government policies and economic constraints. Therefore, effective government policies and support for public transport may help in sustainable growth of cities. Vicente *et al.* (2020) have tried to analyze factors influencing passenger loyalty toward public transport services by using structural equation modeling approach. They have observed that public transport providers' commitment to environmental sustainability does have a direct positive effect on the passenger loyalty. Other researchers have also found in their studies that public prefer to choose environment-friendly solutions even if they cost more (European Commission, 2008; Kennell, 2016). According to Kennell (2016), investment in green operations may have more ROI if properly

communicated to public about its green initiatives. Currie *et al.* (2018) have tried to measure sustainability of public transport in terms of social, economic and environmental factors.

Energy is of utmost importance for all functions in supply chain management, logistics and transportation. Sustainable transportation uses energy efficiently. Road freight transport consumes energy significantly (Browne and Allen, 1998). According to Aßmann and Sieber (2005), worldwide transport consumes about 20% of global primary energy. However, according to a more recent estimate, the global transportation sector consumes 25% of all the energy (The Maritime Executive, 2020). Also, oil prices are dropping now. However, due to the high price of oil in the past, the cost of fuel was around 35% of the total costs of carriers (Gurtu, 2019; Thuermer, 2008). Higher fuel cost or lower energy efficiency also impact the costs of goods sold (Tanowitz and Rutchik, 2008).

Disturbance due to sound is known as noise pollution. It depends on the intensity of different sound sources (Müller-Wenk, 2004). Noise level above 70 decibels (dB) can cause many ailments such as high blood pressure, grouchiness and risk of cardiac infarctions. Noise from transport vehicles also causes vibrations to buildings and homes. Usually, vehicles cause two types of noise: engine noise and rolling noise. Proper maintenance of vehicles and the use of high-quality fuel help in reducing noise pollution.

Traffic accidents cause medical emergencies, casualties and emotional and financial impacts on individuals and their families. They impact family income, work productivity and losses to insurance companies (Elvik, 2000). Distracted driving such as texting and talking over the phone is an avoidable cause of road accidents (Henley *et al.*, 2018). Technological innovations in vehicles, compliance with traffic rules and following maintenance guidelines help in reducing traffic accidents.

Congestion index is the most prominent measure for the evaluation of a transportation system (Boarnet *et al.*, 1998; Vaziri, 2002). It tries to compare the time required for traveling during peak and off-peak hours. Due to congestion, traffic jams happen. People reach late to their offices. Work is not completed on time. It overall impacts the performance of all systems. It also causes more CO_2 emissions, and the fuel efficiency reduces. The cost of transportation also increases.

Emissions are increasing exponentially due to the growth of vehicles on the road. One of the reasons for the increase in the number of vehicles is the increase in the earnings of people. This problem can be tackled with the use of better technology and high quality of fuel. Alternative fuels may also be explored. Oil is the most commonly used source of road transport. In place of oil, transportation systems may explore alternative sources of energy such as biofuels, solar energy, electricity and hydrogen (Gurtu *et al.*, 2015, 2016, 2019). Research needs to be done for comparison of emissions from different sources. It will also help in the selection of sustainable fuel for road transport.

Travel behavior and the choice of the mode of transportation have a significant influence on the sustainability of transport systems (Schneider, 2013), but the influence is mediated by various factors, such as transport policies, car ownership, household income, age (Mattioli, 2014), culture, travel distance (Cattaneo *et al.*, 2018) and weather conditions (Kim and Macdonald, 2016). Song *et al.* (2017) attempted to assess the expected environmental costs of accessibility; Colvile *et al.* (2004) attempted to evaluate the road transport sustainability regarding the exposure to air pollution from traffic through modeling the movement of air, vehicles and exhaust emissions; and Salehi *et al.* (2017) developed a method to minimize total carbon emissions taking into consideration speed control. The transportation system should be sustainable in terms of social, economic and environmental measures. For environmental sustainability, fuel efficiency should be high. Emissions should be less harmful to society. For economic sustainability, the cost of transportation should be minimized by optimizing speed, the number of persons in vehicles and the time required for traveling. For social sustainability, the transportation system should ensure safety, social interaction and

availability (Shi and Zhou, 2012). Any sustainable system must also be energy-efficient and must create minimum waste (Reddy *et al.*, 2000; Zheng *et al.*, 2013).

From the above literature review, it can be inferred that a reduction in transport emissions is crucial for environmental sustainability. This will also help to reduce many health-related, particularly respiratory, ailments in areas with heavy traffic. This understanding was also confirmed by the subject-matter experts. Based on this understanding and many rounds of discussions with experts, a total of eight indicators for sustainable transport have been finalized. These are CO_2 emissions, congestion, number of injuries, cost of fuel, number of accidents, cost of maintenance, energy efficiency and road noise.

3. Research methodology

AHP is used in this study for prioritization of criteria of sustainable transport and for evaluating the most sustainable transport. AHP was developed by Saaty (1980) to solve problems, which are complex, and a solution depends on many factors or criteria. Managers use AHP in various strategic and tactical decision-making situations. Singh (2013) has used AHP for prioritization of factors for a coordinated supply chain. Chhabra *et al.* (2017) have used AHP for analyzing alternatives for green logistics. Bhandari *et al.* (2018) have used AHP for justification of advanced manufacturing technologies for small and medium-sized enterprises.

Analytic Hierarchy Process (AHP): AHP helps decision-makers to analyze the relative importance of criteria for effective decision-making (Saaty, 1980). These criteria are compared on the Saaty's (1980) nine-point rating scale (Table 1). The steps used to find the importance of rating of criteria using AHP are discussed as follows:

Step 1. To prepare a pairwise comparison matrix, each criterion is compared with respect to another on the 9-point rating scale.

Step 2. To produce the normalized matrix, each element of each column is divided by the sum of all the elements of the corresponding column.

Step 3. Average value of each row of matrix C' (i.e. the matrix produced in step-2) is calculated.

Step 4. The consistency of the decision matrix is checked by the consistency ratio. The value of maximum eigenvalue (λ_{max}) is required for finding the consistency index (CI) as follows.

(1) Assume that the pairwise comparison matrix is represented by P1 and the principal matrix is represented by P2

Intensity of importance	Definition
1	Equal importance
3	Weak importance of one over another
5	Essential or strong important
7	Demonstrated importance
9	Absolute importance
2, 4, 6, 8	Intermediate values between the two adjacent judgment
Reciprocals of above non-zero	If activity *i* has one of the above non-zero numbers assigned to it when compared with activity *j*, then *j* has the reciprocal value when compared with *i*

Table 1. Tomas Saaty's (1994) nine-point scale

(2) P3 may be defined as the product of P1 and P2, while P4 is defined as P3/P2

(3) let λmax be the average of the elements of P4

Step 5. After finding the maximum eigenvalue of the matrix, the value of CI is calculated as shown in Equation (1),

$$\mathrm{CI} = \frac{\lambda_{\max} - n}{n - 1} \tag{1}$$

Step 6. Finally, the consistency ratio is calculated, as shown in Equation (2).

$$\mathrm{CR} = \frac{\mathrm{CI}}{\mathrm{RI}(n)} \tag{2}$$

where RI (n) is a random index, which varies with the value of n (Table 2). If the value of CR becomes greater than 0.1, then go to step 1and check the consistency of the decisions taken for pairwise comparison. If it is less than 0.1, it means decisions are consistent.

4. Results and discussion

AHP is used to select the most sustainable alternatives for the transportation system in NCR Delhi, India. This section presents the analysis and development of a framework for selecting environment-friendly alternatives.

4.1 Phase 1 – AHP model for analysis

In this case, the goal is to evaluate alternatives in sustainable mobility transportation for NCR Delhi, India. The criteria for the AHP model have been selected based on a literature review (Castillo *et al.*, 2010) and discussion held with industry experts from the urban transportation and environment department. The criteria for selection of alternative are CO_2 emissions, energy efficiency, cost of operation, road noise, number of accidents, number of injuries, congestion and cost of fuel. For this study, authors have considered three transport alternatives. These are state-run buses, car-pool (App-based transport) and *Shuttl*. The formulation of AHP model (Figure 1) consists of a goal, main factors and result (alternatives). The framework has three stages which include the problem at the first stage, criteria as the second stage and the three alternatives – state-run buses, car-pool and *Shuttl* practices – as the third stage. Figure 1 shows the framework for selection of sustainable transport using AHP.

4.2 Phase 2 – Collection and measurement of primary data

After the model of AHP for hierarchy has been formulated, the next phase is about measuring and collecting the primary data. This was done by making a team of five experts and assigning the pairwise comparison to the main factors used in the AHP hierarchy. These experts were chosen and interviewed from different domains of the transportation industry.

Table 2. List of random index values

n	2	3	4	5	6	7	8	9
RI(n)	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45

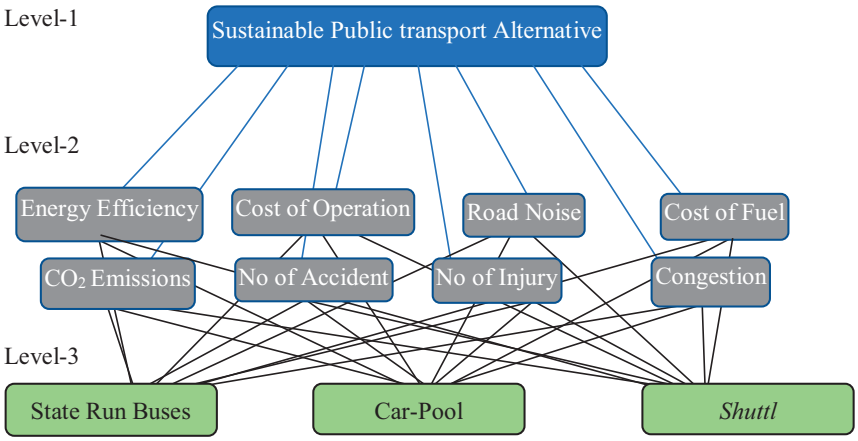


Figure 1.
AHP framework for the selection of sustainable transport alternative

Two of them were fleet managers, and three of the experts were strategists for sustainable transport. The nine-point scale developed by Saaty (1994) for AHP was used to assign relative scores to a pairwise comparison between main factors. The experts were asked to assign a relative score to each comparison using a scale mentioned in Table 2, and the process was continued for all levels of hierarchy. For collecting the primary data from experts, a questionnaire was prepared, which comprised all main factors and was used to obtain the pairwise comparison judgment from the experts. In the case of no consensus, importance was given to the decision which had a majority.

4.3 Phase 3 – formulation of pairwise comparison matrices

In this step, pairwise comparison matrices are constructed for each level of attributes. The governing element for a lower-level element is the higher-level element. Comparison is made between the elements in the lower level to each other. This is done on the basis of their effect on the governing element above it. This gives a square matrix of judgments. The pairwise comparison is made in terms of domination of an element over another element to express the judgments as integers. If element 1 dominates over element 2, then the whole integer is entered in row 1 column 2, and its reciprocal is entered in row 2 column 1. If the compared elements are the same, value 1 is assigned to both positions. Following the above method, pairwise comparison for all criteria shown in level 2 is made, and relative weights for all criteria are found as shown in Table 3. It is observed that CO_2 emission is the most important sustainability criteria, with a relative weight of 0.3258. It is followed by congestion (0.2084), the number of injuries (0.1511), cost of fuel (0.1072), number of accidents (0.0839), cost of maintenance (0.0571), energy efficiency (0.0392) and road noise (0.0273).

Criteria	Criteria weights (priority vector)
CO_2 emissions	0.3258
Cost of fuel	0.1072
Energy efficiency	0.0392
Cost of maintenance	0.0571
Number of accidents	0.0839
Congestion	0.2084
Number of injuries	0.1511
Road noise	0.0273

Table 3.
Weights of different criteria for the selection of sustainable transport

The next phase of the study is to find a solution for the selection of sustainable transport on the basis of the above-mentioned sustainability criteria. First, with respect to each criterion, a pairwise comparison matrix for three selected alternatives is made. The results are shown in Table 4. Values of priority vector in the last column show the relative weight. The importance of different alternatives differs from criteria to criteria. For example, in terms of CO_2 emissions, *Shuttl* is the most important alternative, with a relative weight of 0.5571. In terms of congestion, state-run bus is the most important alternative, with a weight of 0.6333.

Similarly, in terms of other criteria, the most important alternatives have been identified. In Table 5, the relative weights of each criterion and weight for all alternatives are summarized. After getting the relative weights of all alternatives against each sustainability criteria, global weights for each alternative are calculated. For each choice, a set of global priority weight was obtained by multiplying local weights of lower factors with the weights of all higher parent factors and then adding all the multiplied products. The results are given in Table 6. The sum of global weights of an alternative against all criteria gives the global

Table 4.
Pairwise comparison judgment matrices

Normalized pairwise matrix of alternatives for carbon emissions				
Alternatives	State-run bus	Pooled car	*Shuttl*	Priority vector (P.V.)
State-run bus	0.3000	0.3750	0.2857	0.3202
Pooled car	0.1000	0.1250	0.1429	0.1226
Shuttl	0.6000	0.5000	0.5714	0.5571
Normalized pairwise matrix of alternatives for congestion				
State-run bus	0.6522	0.5556	0.6923	0.6333
Pooled car	0.1304	0.1111	0.0769	0.1062
Shuttl	0.2174	0.3333	0.2308	0.2605
Normalized pairwise matrix of alternatives for number of injury				
State-run bus	0.2308	0.2727	0.2258	0.2431
Pooled car	0.0769	0.0909	0.0968	0.0882
Shuttl	0.6923	0.6364	0.6774	0.6687
Normalized pairwise matrix of alternatives for cost of fuel consumed				
State-run bus	0.1667	0.0769	0.2727	0.1721
Pooled car	0.5000	0.2308	0.1818	0.3042
Shuttl	0.3333	0.6923	0.5455	0.5237
Normalized pairwise matrix of alternatives for number of accidents				
State-run bus	0.2308	0.2727	0.2258	0.2431
Pooled car	0.0769	0.0909	0.0968	0.0882
Shuttl	0.6923	0.6364	0.6774	0.6687
Normalized pairwise matrix of alternatives for cost of maintenance				
State-run bus	0.1250	0.0769	0.1818	0.1279
Pooled car	0.5000	0.3077	0.2727	0.3601
Shuttl	0.3750	0.6154	0.5455	0.5119
Normalized pairwise matrix of alternatives for energy efficiency				
State-run bus	0.1429	0.1000	0.1818	0.1416
Pooled car	0.4286	0.3000	0.2727	0.3338
Shuttl	0.4286	0.6000	0.5455	0.5247
Normalized pairwise matrix of alternatives for road noise				
State-run bus	0.2308	0.3750	0.2105	0.2721
Pooled car	0.0769	0.1250	0.1579	0.1199
Shuttl	0.6923	0.5000	0.6316	0.6080

desirability index of the concerned alternative. Global desirability indexes for all alternatives are summarized in Table 7. It is observed that the global desirability index of *Shuttl* is highest, that is, 0.5155. It is followed by state-run buses, with a global desirability index of 0.3322. The last alternative is pooled cars, with a global desirability index of 0.1523.

Therefore, it can be concluded that, overall, *Shuttl* is the most suitable transportation system in NCR considering the criteria of sustainability. It is followed by state-run buses, and lastly by pooled car system. *Shuttl* is most preferred also due to its last-mile connectivity and comfortability to passengers. The *Shuttl* company is located in NCR and specializes in providing safe, comfortable and hassle-free travel to daily commuters. The company has aggregated over 1500 AC buses and 500 cabs across the country. Electric vehicles have also been inducted in the fleet. With the experience gained in the past three and half years, the company is poised to take a quantum leap in its endeavor to benchmark itself against the very best in the industry; it has worked with different governments and municipalities to provide the best solution in public private partnership (PPP) model. Due to the rigorous quality controls and client-oriented approach, the company has acquired its client list with some very reputed and renowned names. For the past few years, *Shuttl* company has started putting efforts toward the implementation of green practices in various activities. Therefore, it is felt that other alternatives should also adopt green practices to make their transport system sustainable.

Table 5. Weights of criteria for alternatives

Criteria for selection of alternative	Level 2 weights	Priority vector state-run bus	Priority vector pooled car	Priority vector *Shuttl*
CO_2 emissions	0.3258	0.3202	0.1226	0.5571
Cost of fuel	0.1072	0.1721	0.3042	0.5237
Energy efficiency	0.0392	0.1416	0.3338	0.5247
Cost of maintenance	0.0571	0.1279	0.3601	0.5119
Number of accidents	0.0839	0.2431	0.0882	0.6687
Congestion	0.2084	0.6333	0.1062	0.2605
Number of injuries	0.1511	0.2431	0.0882	0.6687
Road noise	0.0273	0.2721	0.1199	0.6080

Table 6. Global weights of alternatives

	State-run bus	Pooled car	*Shuttl*
CO_2 emissions	0.1043	0.0399	0.1815
Cost of fuel	0.0184	0.0326	0.0561
Energy efficiency	0.0056	0.0131	0.0206
Cost of maintenance	0.0073	0.0206	0.0293
Number of accidents	0.0204	0.0074	0.0561
Congestion	0.1320	0.0221	0.0543
Number of injury	0.0367	0.0133	0.1010
Road noise	0.0074	0.0033	0.0166
Total global weight	*0.3322*	*0.1523*	*0.5155*

Table 7. Global desirability index of alternatives

Global desirability for state-run bus	0.3322
Global desirability for pooled car	0.1523
Global desirability for *Shuttl*	0.5155

5. Conclusions, limitations, future scope and managerial implications

The purpose of this study was to identify the criteria for sustainable transport system in metro cities and suggests a framework for the selection of the most sustainable transportation system in NCR Delhi, India. We have used AHP for illustrating the proposed framework. This study has been done for policymakers in designing sustainable transportation in an existing metro which has severe traffic congestion. In the present situation, the development of sustainable transport system is a challenge. In this study, three alternatives of transport are considered in context to NCR Delhi, India. These are state-run buses, pooled cars and *Shuttl*. This study is trying to illustrate a framework for the selection of sustainable transport.

In the present era of globalization, ensuring sustainable transport is a big challenge. Green practices in transportation can achieve both economic and environmental objectives. In the era of environmentalism, cities and municipalities having sustainable transport options will surely benefit. The benefits will ultimately lead to performance improvements, increased public comfort, higher profits and overall sustainable growth of cities. We have considered eight criteria for the selection of a sustainable transport system. Criteria for selection of sustainable transport are CO_2 emissions, cost of fuel, energy efficiency, cost of maintenance, number of accidents, congestion, number of injuries and road noise. Based on the pairwise comparison of different criteria through AHP, it is observed that CO_2 emission is the most important criterion, followed by a reduction in congestion and number of injuries. After finding the weights of different criteria, the global desirability index of all three alternatives is calculated. It is found that *Shuttl* has the highest global desirability index. It means that in terms of these criteria, *Shuttl* is the most sustainable alternative.

Findings will be very useful primarily for policymakers and for city planners in developing sustainable transport systems in metro cities. These findings will also have a major impact on the selection of sustainable transport mode by public in cities. Usually public use their own vehicles or other means of transport due to unavailability of sustainable transport. It will further help in terms of improving energy efficiency and reducing fuel consumption, reduction of carbon footprint, maintaining the environment, among others. The results of this research will help future researchers to prioritize various factors according to their benefits and to have a holistic approach toward green logistics. Once the decisions are made at higher levels, it will motivate the people at the mid-level and lower level to work toward sustainable activities. This ultimately will improve the society and the quality of life.

There are certain limitations to the present research. First limitation is that we have considered limited number of sustainability criteria; these criteria may be increased to develop a more comprehensive framework. Framework based on triple bottom line approach can be also used as a part of future studies. Apart from it, ranking of these criteria may also vary from city to city. Second limitation is that AHP is not able to capture the vagueness in decision-making. Therefore, fuzzy AHP can be considered for further detailed analysis as future scope of study in a different context of a sustainable transportation system. Findings may be further validated through case studies for other metro cities with heavy traffic congestions. Further, a similar study can be conducted in another country, and results can be compared to analyze the similarities and differences in the preferences. We can also study about the behavior of public for sustainable transport selection through empirical studies.

References

Agrawal, S. and Singh, R.K. (2019), "Forecasting product returns and reverse logistics performance: structural equation modelling", *Management of Environmental Quality: An International Journal*, doi: 10.1108/MEQ-05-2019-0109.

Arnott, R. and Small, K. (1994), "The economics of traffic congestion", *American Scientist*, Vol. 82 No. 5, pp. 446-455.

Awasthi, A., Chauhan, S.S. and Omrani, H. (2011), "Application of fuzzy TOPSIS in evaluating sustainable transportation systems", *Expert Systems with Applications*, Vol. 38, pp. 12270-12280.

Aßmann, D. and Sieber, N. (2005), "Transport in developing countries: renewable energy versus energy reduction", *Transport Reviews*, Vol. 25 No. 6, pp. 719-38.

Bhandari, D., Singh, R.K. and Garg, S.K. (2018), "Justification of advanced manufacturing technologies for small and medium enterprises from auto component sector: AHP approach", *International Journal of Productivity and Quality Management*, Vol. 23 No. 4, pp. 473-491.

Black, W.R. (1996), "Sustainable transportation: a US perspective", *Journal of Transport Geography*, Vol. 4, pp. 151-159.

Boarnet, M.G., Kim, E.J. and Parkany, E. (1998), *Measuring Traffic Congestion*.

Browne, M. and Allen, J. (1998), "Strategies to reduce the use of energy by road freight transport in cities", *Transport Logistics*, Vol. 1 No. 3, pp. 195-209.

Castillo, N.H. and Pitfield, D.E. (2010), "ELASTIC _ a methodological framework for identifying and electing sustainable transport indicators", *Transportation Research Part D: Transport and Environment*, Vol. 15 No. 4, pp. 179-188.

Cattaneo, M., Malighetti, P., Morlotti, C. and Paleari, S. (2018), "Students' mobility attitudes and sustainable transport mode choice", *International Journal of Sustainability in Higher Education*, Vol. 19 No. 5, pp. 942-962.

Chhabra, D., Garg, S.K. and Singh, R.K. (2017), "Analysing alternatives for Green Logistics in an Indian automotive organisation: a case study", *Journal of Cleaner Production*, Vol. 167 No. 20, pp. 962-969.

Colvile, R.N., Kaur, S., Britter, R., Robins, A., Bell, M.C. and Shallcross, D. (2004), "Sustainable development of urban transport systems and human exposure to air pollution", *The Science of the Total Environment*, Vol. 334, pp. 481-e487.

Cosgrove, D., Gargett, D. and Mitchell, D. (2009), *Urban Passenger Transport: How People Move about in Australian Cities - Information Sheet 31*, Bureau of Infrastructure, Transport and Regional Economics, Canberra, Australia, ISSN 1440–9593.

Cruz, I.S. and Katz-Gerro, T. (2016), "Urban public transport companies and strategies to promote sustainable consumption practices", *Journal of Cleaner Production*, Vol. 123, pp. 28-33.

Currie, G., Truong, L. and Gruyter, C.D. (2018), "Regulatory structures and their impact on the sustainability performance of public transport in world cities", *Research in Transportation Economics*, Vol. 69, pp. 494-500.

Daly, H.E. (1991), *Steady State Economics*, Island Press, Washington, DC.

EC-JRC report (2016), *PBL Netherlands Environmental Assessment Agency European Commission, Joint Research Centre. Trends in Global CO2 Emissions: 2016 Report*, PBL Publishers, Netherlands, pp. 14/30-32.

Elvik, R. (2000), "How much do road accidents cost the national economy?", *Accident Analysis and Prevention*, Vol. 32 No. 6, pp. 849-851, doi: 10.1016/s0001-4575(00)00015-4.

European Commission (2008), "Special eurobarometer 295: attitudes of European citizens towards the environment, brussels, European commission", available at: http://ec.europa.eu/public_opinion/archives/ebs/ebs_295_en.pdf.

Gupta, A. and Singh, R.K. (2020), "Managing operations by a logistics company for sustainable service quality: Indian perspective", *Management of Environmental Quality International Journal*, doi: 10.1108/MEQ-11-2019-0246.

Gurtu, A. (2014), *A Framework for Reducing Greenhouse Gas (GHG) Emissions through Carbon Pricing for Offshore Outsourcing*, Ph.D., Ryerson University, Toronto, available at: http://digital.library.ryerson.ca/islandora/object/RULA%3A3388 (accessed 28 January 2015).

Gurtu (2019), "A pioneering approach to reducing fuel cost and carbon emissions from transportation", *Transportation Journal*, Vol. 58 No. 4, doi: 10.5325/transportationj.58.4.0309.

Gurtu, A., Jaber, M.Y. and Searcy, C. (2015), "Impact of fuel price and emissions on inventory policies", *Applied Mathematical Modelling*, Vol. 39 Nos 3-4, pp. 1202-1216, doi: 10.1016/j.apm.2014.08.001.

Gurtu, A., Searcy, C. and Jaber, M.Y. (2016), "A framework for reducing global manufacturing emissions", *The Journal of Environment and Development*, Vol. 25 No. 2, pp. 159-190, doi: 10.1177/1070496515623821.

Gurtu, A., Searcy, C. and Jaber, M.Y. (2019), "Transportation and sustainable supply chains", in Sarkis, J. (Ed.), *Handbook on the Sustainable Supply Chain*, 1 ed., Edward Elgar Publishing, pp. 410-428.

Haase, D. and Nuissi, H. (2007), "Does urban sprawl drive changes in the water balance and policy? The case of Leipzig (Germany), 1870–2003", *Landscape and Urban Planning*, Vol. 80 Nos 1–2, pp. 1-13.

Hall, R.P. (2002), *Introducing the Concept of Sustainable Transportation to the U.S. DOT through the Reauthorization of TEA-21*, Massachusetts Institute of Technology, Cambridge, MA.

Henley, W.H., Ranganathan, S.K. and Gurtu, A. (2018), "Improving effectiveness of public service advertisements to prevent texting and driving of American youth", *International Journal of Nonprofit and Voluntary Sector Marketing*, Vol. 23 No. 4, pp. 1-9, doi: 10.1002/nvsm.1626.

Hindu (2019), *Fifteen of the 20 Most Polluted Cities in the World Are in Indi*, 5th March 2019.

Kennell, B. (2016), "Environmental concern empowers the people", available at: http://www.huffingtonpost.com/brian-kennell/environmental-concern-emp_b_8105580.html (accessed 18 October 2016).

Kim, H. and Macdonald, E. (2016), "Does wind discourage sustainable transportation mode choice? Findings from San Francisco, California, USA", *Sustainability*, Vol. 8 No. 3, p. 257.

Litman, T. and Burwell, D. (2006), "Issues in sustainable transportation", *International Journal of Global Environmental Issues*, Vol. 6 No. 4, p. 331e347.

Loo, B.P.Y. (2002), "Role of stated preference methods in planning for sustainable urban transportation: state of practice and future prospects", *Journal of Urban Planning and Development*, Vol. 128 No. 4, pp. 210-224.

Marsden, G. and Rye, T. (2010), "The governance of transport and climate change", *Journal of Transport Geography*, Vol. 18, pp. 669-678.

Mattioli, G. (2014), "Where sustainable transport and social exclusion meet: households without cars and car dependence in great britaBritainin", *Journal of Environmental Policy and Planning*, Vol. 16 No. 3, pp. 379-400.

May, A.D., Page, M. and Hull, A. (2001), "Developing a set of decision-support tools for sustainable urban transport in the UK", *Transport Policy*, Vol. 15, pp. 328-340.

Mohan, D. and Tiwari, G. (1999), "Sustainable transport systems: linkages between environmental issues, public transport, non-motorised transport and safety", *Economic and Political Weekly*, Vol. 34 No. 25, pp. 1589-1596.

Müller-Wenk, R. (2004), "A method to include in lca road traffic noise and its health effects", *International Journal of Life Cycle Assessment*, Vol. 9 No. 2, pp. 76-85, doi: 10.1007/bf02978566.

Nakamura, H., Hayashi, Y. and May, A.D. (Eds) (2004), "Urban transport and the environment—an international perspective", *World Conference on Transportation Research Society and Institute for Transport Policy Studies*, Elsevier, Amsterdam.

Reddy, A.K., Anand, Y.P. and D'Sa, A. (2000), "Energy for a sustainable road/rail transport system in India", *Energy for Sustainable Development*, Vol. 4 No. 1, pp. 29-44.

Saaty, T.L. (1980), *The Analytic Hierarchy Process*, McGraw-Hill, New York, NY.

Saaty, T.L. (1994), "How to make a decision: the analytic hierarchy process", *Interface*, Vol. 24 No. 6, pp. 19-43.

Salehi, M., Jalalian, M. and Siar, M.M.V. (2017), "Green transportation scheduling with speed control: trade-off between total transportation cost and carbon emission", *Computers and Industrial Engineering*, Vol. 113, pp. 392-e404.

Schipper, L. (2003), "Sustainable urban transport in the 21st century: a new agenda", *Proceedings of Conference: Transportation, Energy, and Environmental Policy: Managing Transitions, held in Monterey, CA, 9/11/2001–9/12/01*, Transportation Research Board, Washington, DC.

Schneider, R.J. (2013), "Theory of routine mode choice decisions: an operational framework to increase sustainable transportation", *Transport Policy*, Vol. 25, pp. 128-e137.

Shi, J. and Zhou, N. (2012), "A quantitative transportation project investment evaluation approach with both equity and efficiency aspects", *Research in Transportation Economics*, Vol. 36 No. 1, pp. 93-100, doi: 10.1016/j.retrec.2012.03.00.

Singh, R.K. (2013), "Prioritising the factors for coordinated supply chain using analytic hierarchy process (AHP)", *Measuring Business Excellence*, Vol. 17 No. 1, pp. 80-98, Emerald.

Singh, R.K. and Agrawal, S. (2018), "Analyzing disposition strategies in reverse supply chains: fuzzy TOPSIS approach", *Management of Environmental Quality: An International Journal*, Vol. 29 No. 3, pp. 427-443.

Song, Y., Miller, H.J., Stempihar, J. and Zhou, X.S. (2017), "Green accessibility: estimating the environmental costs of network-time prisms for sustainable transportation planning", *Journal of Transport Geography*, Vol. 64, pp. 109-e119.

Stefanieca, A., Hosseinib, K., Xiec, J. and Lia, Y. (2020), "Sustainability assessment of inland transportation in China: a triple bottom line-based network DEA approach", *Transportation Research Part D*, Vol. 80 No. 2020, p. 102258.

Tanowitz, M. and Rutchik, D. (2008), *Squeezing Opportunity Out of Higher Fuel Costs*, Supply Chain Management Review, October, pp. 34-40.

The Maritime Executive (2020), "Transport uses 25 percent of world energy", available at: https://www.maritime-executive.com/article/transport-uses-25-percent-of-world-energy (accessed 28 March 2020).

Thuermer, K.E. (2008), "Air Cargo braces for a slowdown", *Supply Chain Management Review*, December.

US EPA (2020), US Environment Protection Agency Report, 2020.

Van den Berg, R. and De Langen, P.W. (2017), "Environmental sustainability in container transport: the attitudes of shippers and forwarders", *International Journal of Logistics Research and Applications*, Vol. 20 No. 2, pp. 146-162.

Vaziri, M. (2002), "Development of highway congestion index with fuzzy set models", *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1802 No. 1, pp. 16-22, doi: 10.3141/1802-03.

Vicente, P., Sampaio, A. and Reis, E. (2020), "Factors influencing passenger loyalty towards public transport services: does public transport providers' commitment to environmental sustainability matter?", *Case Studies on Transport Policy*, Vol. 8, pp. 627-638.

World Bank (1996), *Sustainable Transport: Priorities for Policy Reform*, The International Bank for Reconstruction and Development/ The World Bank, Washington, DC.

Zhao, X., Ke, Y., Zuo, J., Xiong, W. and Wu, P. (2020), "Evaluation of sustainable transport research in 2000-2019", *Journal of Cleaner Production*, Vol. 256, p. 120404.

Zheng, J., Garrick, N.W., Atkinson-Palombo, C., McCahill, C. and Marshall, W. (2013), "Guidelines on developing performance metrics for evaluating transportation sustainability", *Research in Transportation Business and Management*, Vol. 7, pp. 4-13, doi: 10.1016/j.rtbm.2013.02.001.

Zhou, J. (2012), "Sustainable transportation in the US: a review of proposals, policies, and programs since 2000", *Frontiers of architectural research*, Vol. 1 No. 2, pp. 150-e165.

About the authors

Abhishek Singh is doing his graduation in environment engineering from Delhi Technological University, Delhi, India. He has keen interest in environment-related research issues.

Amulya Gurtu is an Associate Professor of Supply Chain Management at the Cofrin School of Business, University of Wisconsin, Green Bay, WI, USA. His research interests are in global manufacturing and supply chains and their impact on the environment and society. He obtained his PhD from Ryerson University, Canada, and an MBA from Western University, Canada. His research has been published in the *Journal of Business and Economic Perspectives,* the *International Journal of Physical Distribution and Logistics Management, Transportation Journal, the International Journal of Nonprofit and Voluntary Sector Marketing*, *Operations and Supply Chain Management*, *Services Marketing Quarterly*, the *International Journal of Inventory Research*, the *Journal of Cleaner Production*, the *Journal of Environment and Development, Management Research Review, Sustainable Production and Consumption* and *Applied Mathematical Modelling*. He served as the Editor-in-Chief of the *International Journal of Sustainable Entrepreneurship and Corporate Social Responsibility* (IJSECSR) published by IGI Global, and he is currently serving as an associate editor of the *International Journal of Systems Science: Operations and Logistics* published by Taylor and Francis.

Dr. Rajesh Kumar Singh is professor in operations management area at Management Development Institute (MDI), Gurgaon, India. His areas of interest include supply chain management, Industry 4.0 and circular economy. He has about 170 research papers published in reputed international/national journals and conferences. He has published papers in journals such as *Annals of Operations Research, Resources, Conservation and Recycling, Journal of Cleaner Production, Industrial Management and Data Systems, Production planning and Control, International Journal of Production Research, Singapore Management Review, International Journal of Productivity and Performance Management, Benchmarking: An International Journal, Journal of Modelling in Management, Competitiveness Review: An International Business Journal, Global Journal of Flexible Systems and Management, International Journals of Productivity and Quality Management, International Journal of Services and Operations Management, International Journal of Automotive Industry and Management, International Journal of Logistic Systems Management, IIMB Management Review and Productivity Promotion*. Presently, he serving as an associate editor of *International Journal of Consumer Studies* published by Wiley. He has also worked as a guest editor for many reputed journals. Rajesh Kumar Singh is the corresponding author and can be contacted at: rksdce@yahoo.com

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**

International Journal of Mechanical and Production Engineering Research and Development (IJMPERD)
ISSN (P): 2249–6890; ISSN (E): 2249–8001
Vol. 10, Issue 3, Jun 2020, 3397-3406
© TJPR Pvt. Ltd.

TRANS STELLAR
•Journal Publications •Research Consultancy

SIX SIGMA APPLICATION IN KETCHUP INDUSTRY FOR DAMAGED GLASS BOTTLES: A CASE STUDY

YATHARTH SHARMA, SARMAD HASSAN & RAJAT KUMAR

Delhi Technological University, New Delhi, India

ABSTRACT

The aim of this paper is to examine the problem of loss incurred due to damage of glass bottles in the tomato sauce/ketchup industry which not only results in depreciating economy to the manufacturer but also degrades the reputation of a company and degrades its image. This paper uses the DMAIC methodology which is a part of Six Sigma for evaluating the problem. The identification of the root causes was done using Kaoru Ishikawa Diagram (Fish Bone Diagram). The weight of individual root cause was calculated using Analytical Hierarchy Process (AHP). Pareto Chart was established to figure out the root causes of the greatest consequence, which were having a very significant impact on the deliverables. Subsequently, brainstorming was done with the experts of this field including the department of Quality Control and logistics was organized which facilitated in figuring out the solutions by reviewing the process map. This paper has been the first attempt to find and review the root causes of the problem of Glass Bottles damage in this particular industry and suggest solutions so that there is a decline in the quantity of damages using DMAIC approach given under Six Sigma methodology.

KEYWORDS: AHP, DMAIC, Industrial Management, Lean Management, Quality Control & Six Sigma

Received: Jun 20, 2020; **Accepted**: Jul 01, 2020; **Published**: Jun 14, 2020; **Paper Id.**: IJMPERDJUN2020323

Original Article

1. INTRODUCTION

Six Sigma is a useful methodology for problem-solving. It gives a valuable measurement approach. Six Sigma has a mathematical base, which, with proper utilization of methodologies, can help in refining the quality of both the process and the product. The basic goal of this methodology is the application of a strategy based on measurement focused on improving the process and reducing variation by applying specific projects. Managers in the food industry have to thus focus on maintaining profitability in a reducing market, along with providing efficient service and good quality products to increasingly sophisticated buyers. In order to accomplish this apparently unattainable goal, companies can choose two strategies. First, they can improve their system's operational efficiency. Second, they can take proper steps to improve its operational quality.

2. CASE STUDY

This case study is performed at Company XYZ which is a leading ketchup bottle manufacturer in India. The DMAIC methodology is applied below. To the best of our knowledge, no research work is done till date to resolve the problem of damage of Glass Bottles in Ketchup industry.

2.1 Define Phase

2.1.1 Problem Statement

To reduce the quantity of damaged glass bottles from 9,830 in the preceding year to 4,200 in the next year.

2.1.2 Business Case

Company XYZ has experienced 9,830 damaged glass bottle units in the preceding year worth of INR 1,100,000/-. This has resulted in the followingdirect cash losses to the company. These losses are mentioned in table 1.

Table 1

Particular	Loss Amount (INR)
Loss of Glass bottle cost	70337 /-
Loss of cost of extra engagement of labor	44322 /-
Loss in revenue due to devaluation of the product	126632 /-
Total Loss	**241,291 /-**

Therefore, loss from each bag damaged= INR 241,291 / 9830 =INR 24.54/-

2.1.3 Process Map

The sequence in which the movement of bag takes place and its sub process map is shown in figure. 1

2.2 Measure Phase

After this tool of data analysis, we use AHP to find the weightage of each root cause followed by use of Pareto Chart to prioritize our resources and direct attention to vital root causes.

The cause & effect relationship shown in figure 2 below can occur at the vendors point, in transit between vendors point and main store, at the main store, in transit between main store and filling station and at the filling station. Whereas the causes & effects relationship shown in figure. 3 can occur at the filling station, in transit between filling station and warehouse, in transit between various warehouse positions, in transit between warehouse and platform, at platform, in transit between platform and truck, at truck during transit between plant and point of unloading and at the point of unloading.

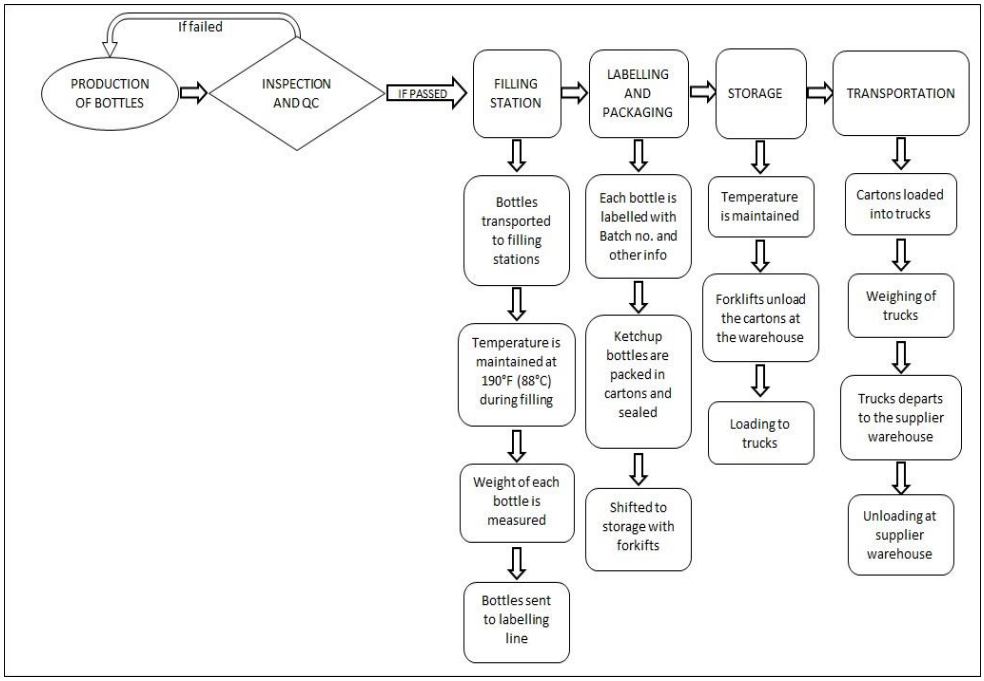


Figure 1: Process Map and Sub-Process Map (Glass Bottle).

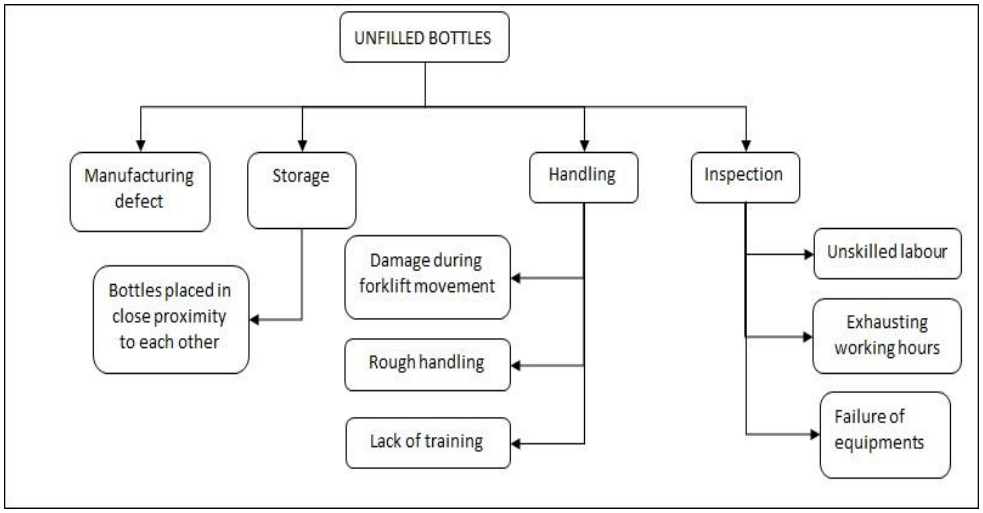


Figure 2: Fish Bone Diagram (Unfilled Bottles).

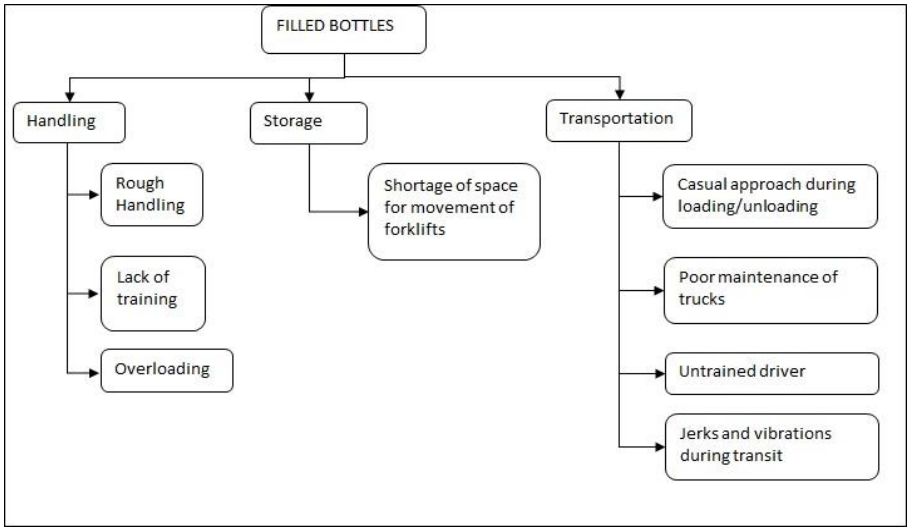


Figure 3: Fish Bone Diagram (Filled Bottle).

2.3 Analysis Phase

Due to similarity between the root causes, they are combined together under six main heads, which from here onwards will be identified as the root causes. First root cause is Storage defects (ST) which includes improper management of space which may cause defects during movement of forklifts. Second root cause is Transportation (T) which is very important to consider since the products are fragile and can break if not taken care properly. It includes improper loading/unloading practice, poor maintenance of trucks, too much jerk and vibration during transit, etc. Third root cause is Production (P) where defects can arrive due to unskilled labor, improper production techniques, etc. Fourth root cause is Packaging (PG). Fifth root cause is lack of Attention and Supervision (S) at multiple stages like filling station, loading/unloading. Sixth root cause is issues during Inspection (IN), this may be due to non-maintenance of Inspecting machines, false positive inspection, etc.

2.3.1 Weight Age Calculation using ahp

Weightage of every individual root cause is calculated using AHP (Analytic Hierarchy Process). AHP is designed tosolve

complex multi criteria problems. Many quantitative approaches have been taken into observation to evaluate new technology implementation. Table 2 depicts the pairwise combination of each of the root cause. The fraction obtained in table 2 is converted into decimals as presented in table 3. On the basis of these pairwise combinations, iterations are performed till we obtain consistency ratio of less than 0.1 (If the consistency ratio is more than 0.1, then the data provided is not reliable). These iterations is shown in table 4.

Table 2: Pairwise Combinations of Each Root Cause

	ST	T	PR	PG	S	IN
ST	1	1/4	1/3	1/2	1/3	1/5
T	4	1	2	5	3	4
PR	3	1/2	1	2	1	1/3
PG	2	1/5	1/2	1	1/5	1/2
S	3	1/3	1	5	1	3
IN	5	1/4	3	2	1/4	1
SUM	18	2.533	7.833	15.500	5.783	9.033

Table 3: Pairwise Combination of Root Cause In Decimals

	ST	T	PR	PG	S	IN
ST	0.055	0.098	0.042	0.032	0.057	0.022
T	0.222	0.394	0.255	0.322	0.518	0.442
PR	0.166	0.197	0.127	0.129	0.172	0.036
PG	0.111	0.078	0.063	0.064	0.034	0.055
S	0.166	0.131	0.127	0.322	0.172	0.332
IN	0.277	0.098	0.382	0.129	0.043	0.110

Table 4: Checking For Consistency

	ST	T	PR	PG	S	IN	Weighted Sum Average	Criteria Weight	λ
ST	0.051	0.090	0.046	0.034	0.070	0.035	0.326	0.051	6.331
T	0.206	0.359	0.277	0.340	0.627	0.695	2.504	0.359	6.967
PR	0.154	0.180	0.138	0.136	0.209	0.058	0.875	0.138	6.325
PG	0.103	0.072	0.069	0.068	0.042	0.087	0.441	0.068	6.476
S	0.154	0.120	0.138	0.340	0.209	0.521	1.483	0.209	7.099
IN	0.257	0.090	0.415	0.136	0.052	0.174	1.125	0.174	6.473

After taking the average value ofλ, we get $\lambda(avg)$=6.611

$$Consistency\ Index\ (CI) = \frac{\lambda(\text{avg})-n}{n-1} = 0.122366$$

$$Consistency\ Ratio = \frac{CI}{RI} = 0.098682$$

Where RI is Random Index and is fixed as 1.24 for matrix size (Factors) of 6. Since the consistency ratio is less than 0.10, we can say that our matrix is reasonably consistent.

ST	0.05
T	0.36
PR	0.14
PG	0.07
S	0.21
IN	0.17

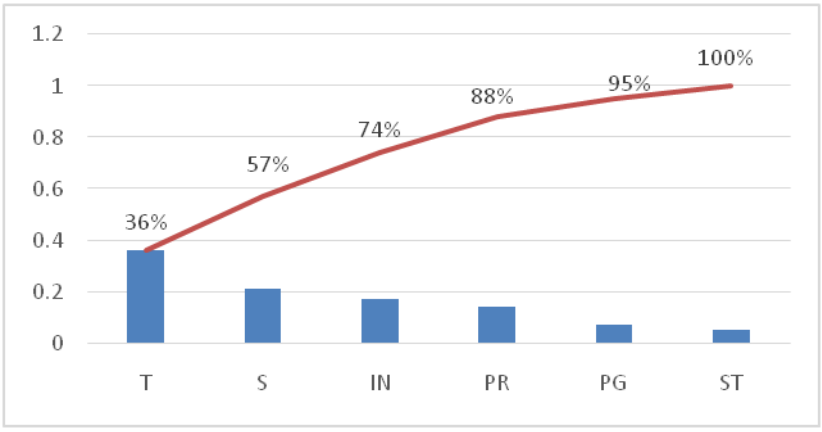


Table 5: Values of Pareto Chart Figure 4: Pareto Chart (Weightage of Each Root Cause).

2.3.2 Analysis of Weight Age of each Root Cause

Figure 4 shows analysis of root causes through their weightage obtained from AHP, as shown in table 5

It can be understood from the pare to chart that Transportation (T) and Storage (S) combined accounts for more than half of the loss due to damaged bottles. Other major root cause being Inventory (IN). Remaining root causes (PG & ST) account for just 20% of the loss. Giving due priority to T, S, IN can cover upto 3/4th of the total loss due to damaged bottles. CE being the only factor on which the company will have least control because the product is FMCG and it spent less time in company warehouse.

2.3.3 Analysis of Loss that is Associated with each Root Cause

Through the weightage calculated from AHP, we now deduce the loss related with each root cause by multiplying their weightage with the total no. of bags damaged bottles. This gives us the no. of bottles damaged due to each root cause or the loss related with each root cause in INR, since, we know the cost of each bottle damaged (As calculated in the business case).Table 8 shows this analysis. Furthermore, table 8 is represented in the form of a Pareto chart shown in figure. 6 which high lights the importance of each root cause in monetary terms. From figure 6 it is clearly visible that our target of reducing loss due to bottle damage by 43% is easily achievable by primarily focusing on 3 main root causes, namely, T, S and IN.

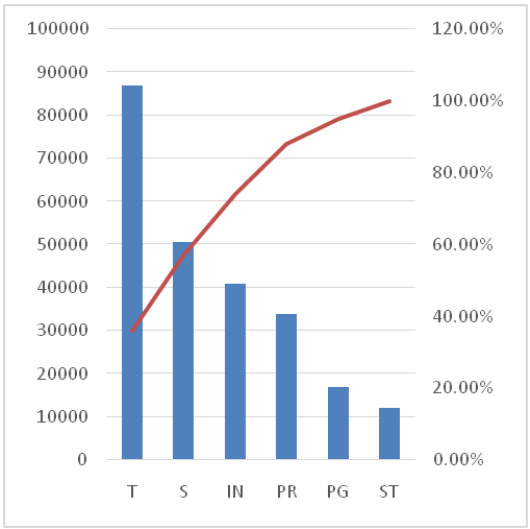


Figure 5: Pareto Chart (Losses Linked With Each Root Cause).

Table 6: Values of Pareto Chart

Root Cause	Weight Age	No. of Bottles Damaged	Loss Amount (INR)
ST	0.05	492	12064.55
T	0.36	3539	86864.76
PR	0.14	1376	33780.74
PG	0.07	688	16890.37
S	0.21	2064	50671.11
IN	0.17	1671	41019.47

Since cost of each glass bottle damaged = INR 24.54 /- (Illustrated in the Business Case)

2.3.4 Date wise Analysis of 12-Months Data

It is important to analyze the point of time during the month at which glass bottle damage takes place. This helps us in verifying the practicality of our interpretations of figure 4 and figure 5. Table 8 gives a 12 months data of summation of glass bottle damage on each of the date. Table 8 is represented in the form of a paretochart in Figure. 6 which is not following the decreasing trend since we need to analyze the trend across the month. It is clearly visible that in the second half of the month, due to excess movement of the glass bottles and lower rate of attention by staff on each bottle, more bottles get damaged. This verifies our claim that over majority of the losses due to bag damage can be prevented if three root causes T, S and IN are properly focused on, along with brief attention to the issue of maintenance.

2.4 Improve Phase

In order to reduce loss due to bags getting damaged, following solutions are suggested in table 7.

Table 7: Suggestions from the Improve Phase

S. No.`	Solutions	Target Cause
1	Proper loading &unloading procedures must be followed	T
2	Training to the workforce and regularly briefing by supervisor in order for cheering up morale.	S
3	Trucks must be serviced properly so that its suspension system does not becomes stiff	T
4	More number of testing must be done,	IN
4	Proper training must be provided to the driver	T
5	Solving the points of bottlenecks in order to become more efficient, automated monitoring of the production processes	PR
6	reducing false positive reports of the empty bottles	IN
7	Gentle handling of the bottles, prevention of collision of bottles from each other	PG
8	Warehouse and production plant must be organized properly so that space management can be done	ST
9	Prevent overloading of the trucks and forklift	T/ST/S
10	Give proper guidelines for proper handling of the bottles	S
11	Give workforce adequate breaks and manage proper work-life balance	S

2.5 Control Phase

The aim of this project was to bring back the focus of the management to surge in losses due to surge in the amount of glass bottles getting damaged. Thus, this project has identified not only the root causes but also direct financial loss due to each of the root cause of this problem. Solutions based on interpretations of the pareto chart have been accepted and implemented. For sustenance of results, the new process was standardized, documented, and distributed at every department to execute on a daily basis. Additionally, the ongoing performance was reviewed till it became part of an everyday daily process.

Table 8: Pareto Chart (Date wise Analysis of 12-Months Data

Date	Damaged Bottles	Date	Damaged Bottles	Date	Damaged Bottles	Date	Damaged Bottles
1	146	9	235	17	378	25	376
2	204	10	137	18	598	26	312
3	116	11	278	19	478	27	289
4	118	12	248	20	446	28	256
5	198	13	221	21	364	29	221
6	273	14	464	22	398	30	309
7	489	15	443	23	287	31	451
8	263	16	489	24	345		

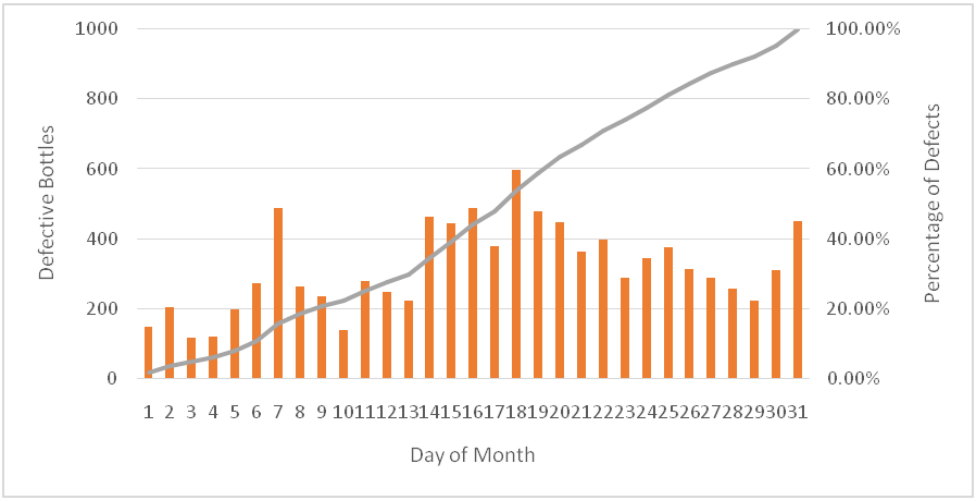


Figure 6: Pareto Chart (Date wise Analysis of 12 Month Data).

3. CONCLUSIONS

This study is performed to classify and examine the main causes of the problem of Glass Bottle damage in Ketchup/Tomato Sauce industry and suggest solutions to reduce damages.

In order to recognize opportunities for improvement, a process map was documented. Root causes were recognized with the support of fish-bone diagram. Weightage of each root cause was calculated using analytical hierarchy process and a pare to chart aided to focus attention on the most pressing root causes.

Accordingly, brainstorming was performed with the production, warehouse, logistics and quality control section was done which helped in framing the solutions by examining the process map.

Although this is one of the first attempts to focus on the problem of Glass Bottle damage, academics and practitioner could use this study for future study in other industries where Glass Bottle are used. Future research can be conducted by calculating the weightage of each root cause using paired comparison method or balanced score card method. A follow up on the level of lean implementation and cultural change in the companies addressed in this study is advised. Also, specifically concerning other causes, future work may focus on more accurately quantifying the impact of machine design changes and the impact in inventory reduction. Finally, another interesting research avenue may be comparing Lean Management implementation results among companies of different sectors. This would allow further understanding the

potential and importance of process innovation in different industrial contexts.

REFERENCES

1. *Antony J (2008). Can Six Sigma be effectively deployed in SMEs? Int. J.Prod. Perform. Manage., 57(5): 420-423.*
2. *Antony J, Banuelas R (2001). Six Sigma: A business strategy for manufacturing organizations. Manuf. Eng., 80(3): 119-121.*
3. *Antony J, Banuelas R (2002). Key ingredients for the effective implementation of Six Sigma programmes. Measuring Bus. Excellence., 6(4): 20-28.*
4. *AHN, Kook Chan. "Impact Response of the Film Mechanical Properties of the Coated Glass Plate."International Journal of Mechanical and Production Engineering Research and Development (IJMPERD) 9. 2, Apr 2019, 499-506*
5. *Bergman B, Kroslid D (2000). Six Sigma : a revival of the Profound Knowledge of variation. Proceedings of the 3rd International Conference on Building People and Organizational Excellence (Aarhus, Denmark), pp. 260- 266.*
6. *Brewer PC, CPA, Eighme JE (2005). Using six sigma to improve the finance function: Here are some tips for success. Strat. Finance., 85(5): 27-33.*
7. *Kang, H. D., and Kook Chan Ahn. "A Study on the Dynamic Behaviour of the Coating Tempered Glass Plate Under Impact." Int. J. of Mechanical and Production Engineering Research and Development (IJMPERD), 8 (6) (2018): 193-200.*
8. *Breyfogle FW (1999). Implementing Six Sigma: Smarter Solutions using Statistical Methods, New York: John Wiley.*
9. *Chen SC, Chen KS, Hsia TC (2005). Promoting customer satisfactions by applying six sigma: An example from the automobile industry. Qual. Manage. J., 12(4): 21-33.*
10. *Mohan, N. A. N. J. A. N. G. U. D., Sathyashankara Sharma, and R. I. T. E. S. H. Bhat. "A comprehensive study of glass fibre reinforced Polymer (gfrp) drilling." International Journal of Mechanical and Production Engineering Research and Development 9.1 (2019): 1-10.*
11. *Kumar M, Antony J (2008). Comparing the quality management practices in UK SMEs. Ind. Manage. Data Syst., 108(9): 1153-1166.*
12. *Hellsten U, KlefsjoÈ B (2000). TQM as a management system consisting of values, techniques and tools. The TQM Magazine., 12: 238- 244.*
13. *Ahirrao, S. P., Kashinath N. Borse, and SonaliBagrecha. "Eco friendly pavement blocks of waste glass, fly ash, and dust." International Journal of Civil, Structural, Environmental and Infrastructure Engineering Research and Development 3.5 (2013): 77-80.*
14. *Starbird D (2002). Business excellence: Six Sigma as a management system: A DMAIC approach to improving Six Sigma management processes. Annual Quality Congress Proceedings, Milwaukee, pp. 47-55.*
15. *Heymans, B. (2015). Lean manufacturing and the food industry. Retrieved June 5, 2015, from http://www.flowmakers.com/articles/Articlefoodindustryandkaizen.pdf*
16. *Hirano, H. (2009). JIT implementation manual – The complete guide to just-in-time manufacturing: Volume 2 – Waste and the 5S's. New York: Productivity Press.*
17. *George, M. L. (2003). Lean six sigma for service: How to use lean speed and six sigma quality to improve services and transactions. McGraw-Hill.*
18. *Kennedy, I., Plunkett, A., &Haider, J. (2013). Implementation of lean principles in a food manufacturing company. In A.*

Azevedo (Ed.)

19. *Advances in Sustainable and Competitive Manufacturing Systems, Lecture Notes in Mechanical Engineering (pp. 1579-1590). Springer. doi: 10.1007/978-3-319-00557-7_127*
20. *Ozden, B., "Use of AHP in decision making for flexible manufacturing systems", Journal of Manufacturing Technology Management, Vol. 16 No.7, pp. 808-819, 2005*
21. *Sanders, D. Hild, R., "Six Sigma on business processes: Common organizational issues" Quality Engineering, Vol.12 No.4, pp.603–610, 2000*
22. *Henry, J. R. (2012). Achieving lean changeover: Putting SMED to work. Boca Raton, FL: Productivity Press*

AUTHOR'S PROFILE



Yatharth Sharma received the B.Tech Degree in Mechanical Engineering from Delhi Technological University, in 2019. He was a former member of the team DTU AUV, SAE and ISHRAE. His interests include playing badminton, counselling on career paths and Industrial Engineering applications. He worked with Hilti India as GET and in Mercedes-Benz, EIL, Air India and Pawan Hans as intern and developed skills related to the defence industry equipment manufacturing.



Rajat Kumar received the B.Tech Degree in Mechanical Engineering from Delhi Technological University, in 2019. He was a former BAJA member of the team DelTech BAJA. His interests include designing the transmission system in BAJA vehicles, application of Input-Output Model for End-of-Life-Vehicles in India, thermodynamics and Industrial Engineering applications. He worked with Bharat Electronics Limited as intern and developed skills related to the defence industry equipment manufacturing.



Sarmad Hassan received the B.Tech Degree in Mechanical Engineering from Delhi Technological University, in 2019. He was a former member of ISHRAE. His interests include designing UAV and Robotic systems and Industrial

Engineering applications. He worked with Omnipresent Robot Tech as intern and now as a design engineer and developed skills related to the defence industry equipment manufacturing.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/342754860

Sodium hydroxide effect on the mechanical properties of flyash-slag based geopolymer concrete

Article *in* Structural Concrete · July 2020

DOI: 10.1002/suco.202000068

CITATIONS
0

READS
29

2 authors, including:



Manvendra Verma
Delhi Technological University
4 PUBLICATIONS **1** CITATION

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

behavior of Geopolymer concrete View project

All content following this page was uploaded by Manvendra Verma on 09 July 2020.

The user has requested enhancement of the downloaded file.

SPi	Journal Code	Article ID	Dispatch: 28-JUN-20	CE:
	SUCO	202000068	No. of Pages: 12	ME:

Received: 31 January 2020 | Revised: 30 April 2020 | Accepted: 19 May 2020

DOI: 10.1002/suco.202000068

TECHNICAL PAPER

fib WILEY

Sodium hydroxide effect on the mechanical properties of flyash-slag based geopolymer concrete

Manvendra Verma | Nirendra Dev

Department of Civil Engineering, Delhi Technological University, Delhi, India

Correspondence
Manvendra Verma, Department of Civil Engineering, Delhi Technological University, Delhi 110042, India.
Email: manvendraverma_2k16phdce05@dtu.ac.in

Funding information
Delhi Government

Abstract

Geopolymer concrete is an innovative, sustainable, cementless, and eco-friendly concrete that directly reduces the carbon footprints due to the total replacement of the cement from the concrete. A very excessive amount of CO_2 produces in the production of cement. In the experimental investigation, analysis of the mechanical properties or engineering properties of the GPC of the different molarity of NaOH (8–16 M), and also different Na_2SiO_3/NaOH ratio (0.5–3.0) analyses in the destructive testing of the GPC. Examine the curing temperature effect on the engineering properties of the GPC. After the experimental investigation, oven-cured specimens got a higher engineering strength compared to the ambient-cured specimens of the same mix designs. The 14 M mix design got the optimum point for engineering strength among the various molar concentrated mix designs. In the case of the alkaline ratio, the 2.5 got the optimum point of the engineering strength among all ratios of the sodium silicate to sodium hydroxide. The highest compressive strength, splitting tensile, and flexural strength in the all mix designs got are 35.7 N/mm^2, 5.2 N/mm^2, and 5.6 N/mm^2 respectively at 56 days after oven-curing. Based on results, proposed the correlation equation between the splitting strength and compressive strength and equation between the flexural strength and compressive strength.

KEYWORDS

compressive strength, curing condition, flexural strength, geopolymer concrete, molarity of sodium hydroxide, splitting tensile

1 | INTRODUCTION

Concrete is the second most abundant material used in the world after the water. Conventionally concrete used the Portland cement as a binding material in the concrete. India produced cement around 502 million tons in 2018. Cement is not eco-friendly, so many environmental issues associated with the production of cement in the industry. Cement production contributes around 8% of carbon dioxide in greenhouse gas emission. Cement production reduces the natural resource and also polluted

Abbreviations: AA, alkali activated; AAFS, alkali activated flyash slag; ASTM, American Society for testing and materials; EDS, energy-dispersive X-ray spectroscopy; GGBFS, ground granulated blast furnace slag; GPC, geopolymer concrete; OPC, ordinary Portland cement; SEM, scanning electron microscope; XRD, X-ray diffraction.

Discussion on this paper must be submitted within two months of the print publication. The discussion will then be published in print, along with the authors' closure, if any, approximately nine months after the print publication.

 © 2020 *fib*. International Federation for Structural Concrete

the environment that is very hazardous for our future. There are limited natural resources available on the earth, so we must go for sustainable development.

On the other hand, abundance of industrial solid waste like as flyash, slag and rice husk ash available on the earth. From the industrial production, around 200 million tons of flyash produced annually in the thermal power plant and 20 million tons of slag produced annually in steel manufacturing plants in India. Flyash and slag are wastes of the industry, so they dumped the industrial solid wastes in the ground using cultivated land and also polluted the environment due to the limited mode of use of the industrial solid wastes like as fly ash and slag.

GPC made by the flyash got more density as compare to made by the metakaolin due to its spherical and porous structure of the flyash particle. It also has very high fineness and increases the workability of the GPC. So, they got a higher solid to liquid ratio. For increasing the workability of the Geopolymer mortar increase the dosage of water or superplasticiser but after an optimum point dosage of superplasticiser create the adverse effect and reduce the strength of the mortar cubes. Moreover, increase the dosage of sodium silicate, and NaOH concentrations degrade the workability of mortar. In the AAFS mortar, autogenous shrinkage generates due to the self-desiccation after the hardening but reducing the volume due to the chemical shrinkage after made end product of the reaction in small size. Increasing the dosage of sodium silicate or slag content, reduced the total porosity and also reduced the mesopores volume, resulting in increasing the density of the mortar. The Flyash-Slag based mortar shows the higher drying shrinkage than the OPC samples, due to the higher mesopores volume of Flyash-Slag based mortar compared to OPC mortar.[1]

AAFS mortar showed higher flexural strength and compressive strength as compared to the OPC mortars. The lesser amount of water absorption as compared to OPC mortar and higher the drying shrinkage as compared to OPC mortar, increasing the dosage of flyash degrade the drying shrinkage of AAFS mortar. The end products of the AAFs mortar are semi-crystalline silicate hydrate and amorphous aluminosilicate gel.[2] The Flow of AA mortar rises with the increment of the water to binder ratio and degrades of the fine aggregate to binder ratio. If the fine aggregate to binder ratio higher than 2.5, the workability of the AA mortar sharply reduced. Generally, the slag based AA mortar got a higher strength as compared to the flyash based AA mortar when both type of mortar binder activated by the same NaOH and Na_2SiO_3. Increment of the water content directly reduced the strength of the AA mortar.[3,4] The compressive strength of Geopolymer mortar with the larger quantity of aggregate can be enlarged by the optimizing the amount of alkali because the increment of the proportion of the aggregate in the geopolymer mortar degrade the amount of the geopolymerization but not directly effect on the compressive strength.[5] The thickness of the geopolymer paste has affected the bonding between the aggregate and Geopolymer paste in which the thickness of the paste higher destroy by the vapor evaporation under the elevated temperature, but the paste is too thin paste cannot build the bond correctly. So, with the increase of aggregate size required to change the thickness of the paste.[6] High calcium flyash used as binder resulted in the high strength mortar up to 86 MPa, excellent volumetric stability with the higher flowability.[1] The molarity of the NaOH in the mix design of GPC directly increases the strength and also reduce the workability of the plastic concrete.[7,8] The molarity of the NaOH increase in the mix reduces the setting time of the GPC.[9] The NaOH and sodium silicate mix ratio of 1:1 use to activate the binder. It can generally achieve the strength up to the M50.[10] Alkali activated slag concrete has a higher capability in term of the corrosion resistivity to the OPC concrete. It also reduces the total embodied energy of the construction.[11]

2 | EXPERIMENTAL PROGRAM

A total of 10 numbers mix designs of the GPC cast and tested for the various properties. All the mix samples cured in two conditions ambient and oven, to analyze the effect of the curing condition on the mix designs. GPC made by the Flyash and GGBFs as a binder in 75/25 ratio by weight respectively, binders activated by the NaOH and Na_2SiO_3 at various ratio and with the different molarity of the Sodium hydroxide.

2.1 | Materials specifications

Sodium hydroxide present in the white delinquent flakes forms with the minimum purity level of 97.0% in the material sample purchased from the Fisher Scientific Delhi. Sodium silicate present in compact, tacky/slightly cloudy liquid form with the minimum assay of Na_2O titrimetric is 10.0% and SiO_2, gravimetric is 25.5–28.5% in the solution of the sample purchased from the central drug house, Delhi. Locally available m-sand or stone dust used as a fine aggregate in the mix design of the tests. Stone dust is a well-graded medium sand type of zone II. Its water absorption, specific gravity, silt content and bulk density are 1.21%, 2.62, 6%, and 1,610 kg/m^3, respectively and locally crushed 10 mm and 20 mm coarse aggregate

Uncorrected Proofs

TABLE 1 Chemical composition of flyash and GGBFS

Characteristics	SiO_2	Al_2O_3	CaO	Fe_2O_3	MgO	SO_3	LOI
Flyash (%)	45.8	21.4	13.7	12.6	1.3	1.9	0.1
GGBFS (%)	34.52	20.66	32.43	0.57	10.09	0.77	0.3

as per ASTM standards.[12–18] In the mix design, use the high performance super plasticizing admixture conplast SP430 manufactured by the Fosroc constructive solutions. Conplast SP 430 conforms with BS 5075, BS: EN 934-2 and with ASTM C494 as type A and type F, depending on the dosage used.[19] GGBFS generated by the isolation of molten slag with the help of high-pressure water jets. Separation of the molten slag prevents crystallization, so the end products found in the form of glassy and granular aggregates. That end product crushed, pulverized and separated for the multipurpose of the applications. GGBFS used in the project taken from the Bhilai steel plant, Durg, Chhattisgarh in India.[20]

Table 1 presents the chemical compositions present in the sample materials and Figure 1 shows the scanning electron microscope picture in which that shows the porous undefined shape of particles of GGBFS. Flyash is produced in the thermal power stations by the combustion of coals for the electricity generation. Flyash collects from the Gurunanak Dev thermal plant, Bhatinda, Punjab in India. Flyash is the type of class f by the classification as per the ASTM C618-33512.[21]

Table 1 presents the chemical compositions present in the sample materials and Figure 2 shows the scanning electron microscope picture in which that shows the porous spherical particles of flyash. Figures 3 and 4 show the EDS graph shows the elemental graph of the flyash and GGBFS, respectively in which shows the mineral elements present in the flyash and GGBFS samples.

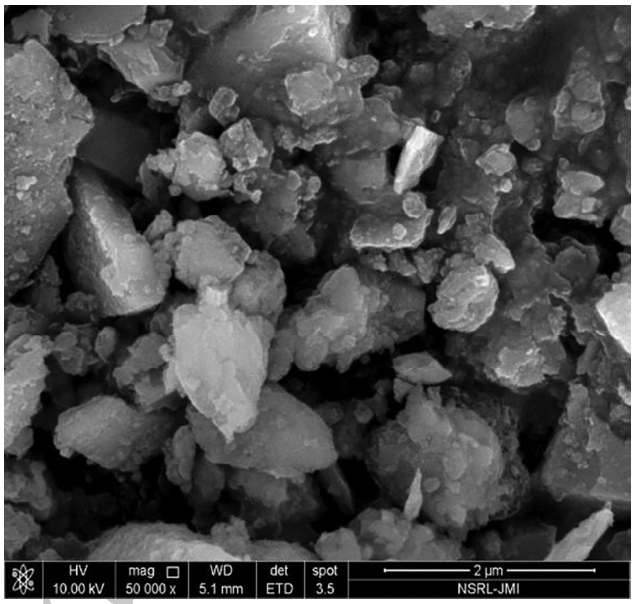


FIGURE 1 SEM image of GGBFS

2.2 | Specimens preparation

All the raw material mix in the 40-l capacity mixture for one time of the batch. The Alkaline chemical made 24 hr before the mixing, sodium hydroxide, sodium silicate and superplasticiser mix before 24 hr of casting. At the time of mixing first pour the coarse aggregate in the mixture then after pouring the sand or fine aggregate and mix properly then after pouring the chemical in the mix and at last pour the extra water as per requirement during the mixing, mix the concrete 3–5 min in the pan mixture.[22] The proper mixed concrete plastic paste cast in the molds are as cube, cylinder and beam shape. As per the Indian standards cubes dimension is 150 mm*150 mm*150 mm, cylinders size is 150 mm*300 mm as per diameter*height and prism in size of 100 mm*100 mm*500 mm as per width*height*length for the testing the mechanical properties of the GPC.[23] The specimens demold after the 3 days of casting then after cured in the oven for the 24 hr and ambient cured. Table 2 describes the all mix composition of the raw materials for the GPC.

The mix design from 1 to 5 shows the variation of the molarity of NaOH adds in the mix design. M1 shows 8 M NaOH dosage and is increases continuously in increasing order with the mix design. The optimum point in term of the strength got from the destructive test at the 14 M dosage in the mix design. After the molarity analysis in the mix design, check the alkaline ratio from 0.5 to 3.0, whereas the sodium silicate to sodium hydroxide called an alkaline ratio. The mix design from M6 to M10 shows the variation in the alkaline ratio from 0.5 to 3.0 in the increasing order.

3 | RESULTS AND DISCUSSION

In the experimental investigation analysis, the compressive strength, splitting tensile, flexural tensile strength and modulus of elasticity of the test results of the specimens of the mix designs. The specimens of all mixes tested at 7 days, 14 days, 28 days, 42 days, and 56 days after the casting.

Uncorrected

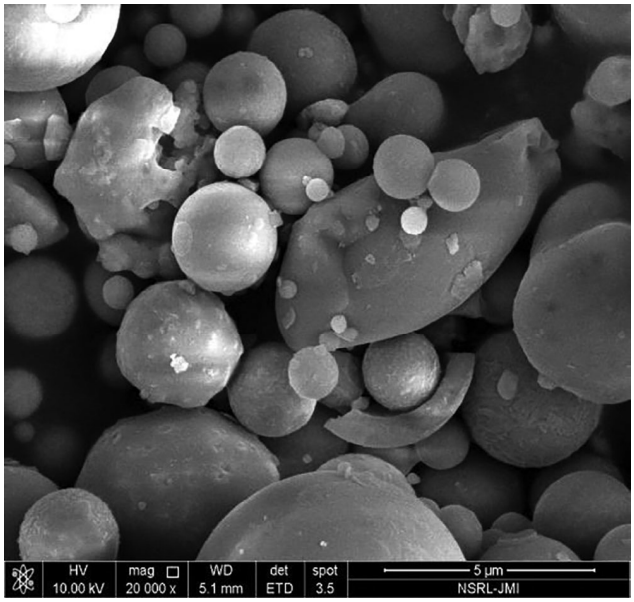


FIGURE 2 SEM image of flyash

3.1 | Effect of sodium hydroxide molar concentration

Sodium hydroxide works as a binder (flyash and GGBFS) activation in the mix designs of the GPC; the changes in the molar concentration of NaOH directly affect the mix design of the GPC. Analysis, the effect of the molarity of NaOH in mix designs of the GPC by varying 8–16 M and examine the test results of the specimens of all design mixes by the destructive strength test.

3.1.1 | Compressive strength

The compressive strength of the GPC varies with the varying molarity of the NaOH and curing condition. The compressive strength rises with the increment in the molarity of NaOH in the mix design but beyond an optimum point decreases the compressive strength in the oven-cured specimens. The highest compressive strength of oven cured 14 M mix is 34.2 N/mm^2 at 56 days but in the ambient cured specimen got at 16 M mix is 25 N/mm^2 at 56 days. Figure 5a shows the compressive strength of ambient cured specimens and Figure 5b shows the compressive strength of oven cured specimens with the variation of the molarity of NaOH in the design mix of GPC. The highest compressive strength of design mixes of 8 M, 10 M, 12 M, 14 M and 16 M are 13.9 N/mm^2, 15.5 N/mm^2, 21.8 N/mm^2, 24.0 N/mm^2, and 25.0 N/mm^2, respectively at 56 days after ambient-curing but after oven-curing compressive strength of 8 M, 10 M, 12 M, 14 M, and 16 M are 23.2 N/mm^2, 25.0 N/mm^2, 27.9 N/mm^2, 34.2 N/mm^2, and 31.0 N/mm^2, respectively at 56 days.

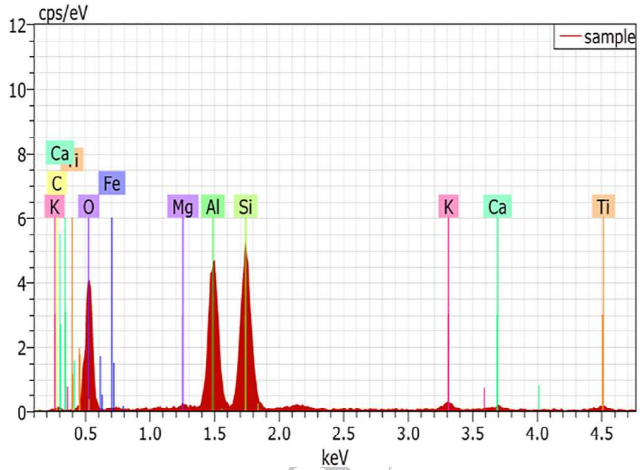


FIGURE 3 EDS graph of GGBFS

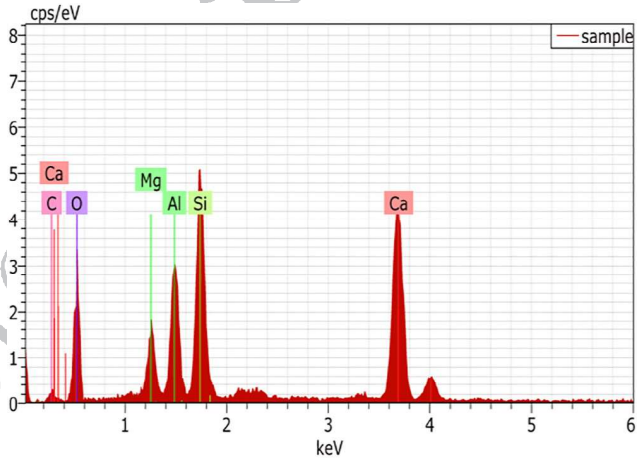


FIGURE 4 Compressive strength of (a) ambient-cured GPC and (b) oven-cured GPC

3.1.2 | Splitting strength

The splitting tensile strength test used to find the indirect tensile strength of the design mix of the GPC. Figure 6a shows the splitting strength of ambient-cured specimens of mix designs and Figure 6b shows the splitting strength of oven-cured samples of the mix designs, in which varies the molarity of NaOH in the mix design of GPC from 8 M to 16 M. In the ambient cured specimens 16 M got the highest splitting tensile strength is 3.6 N/mm^2 at 56 days, but in oven-cured samples 14 M got the most upper splitting tensile strength is 5.2 N/mm^2 at 56 days. The highest splitting tensile strength of 8 M, 10 M, 12 M, 14 M and

Color Figure - Print and Online
Color Figure - Print and Online
Uncorrected Proof

TABLE 2 Mix design

Mix	Fly ash (kg/m^3)	GGBFS (kg/m^3)	Coarse aggregate (kg/m^3)	Fine aggregate (kg/m^3)	NaOH solution (kg/m^3)	Sodium silicate (kg/m^3)	Super plasticizer (kg/m^3)	Extra water (kg/m^3)
M1	303.75	101.25	1,269	683	81.0 (8 M)	81.0	4.05	20.25
M2	303.75	101.25	1,269	683	81.0 (10 M)	81.0	4.05	20.25
M3	303.75	101.25	1,269	683	81.0 (12 M)	81.0	4.05	20.25
M4	303.75	101.25	1,269	683	81.0 (14 M)	81.0	4.05	20.25
M5	303.75	101.25	1,269	683	81.0 (16 M)	81.0	4.05	20.25
M6	303.75	101.25	1,269	683	108.0 (14 M)	54.0	4.05	20.25
M7	303.75	101.25	1,269	683	64.8 (14 M)	97.2	4.05	20.25
M8	303.75	101.25	1,269	683	54.0 (14 M)	108.0	4.05	20.25
M9	303.75	101.25	1,269	683	46.28 (14 M)	115.72	4.05	20.25
M10	303.75	101.25	1,269	683	40.5 (14 M)	121.5	4.05	20.25

Color Figure - Print and Online

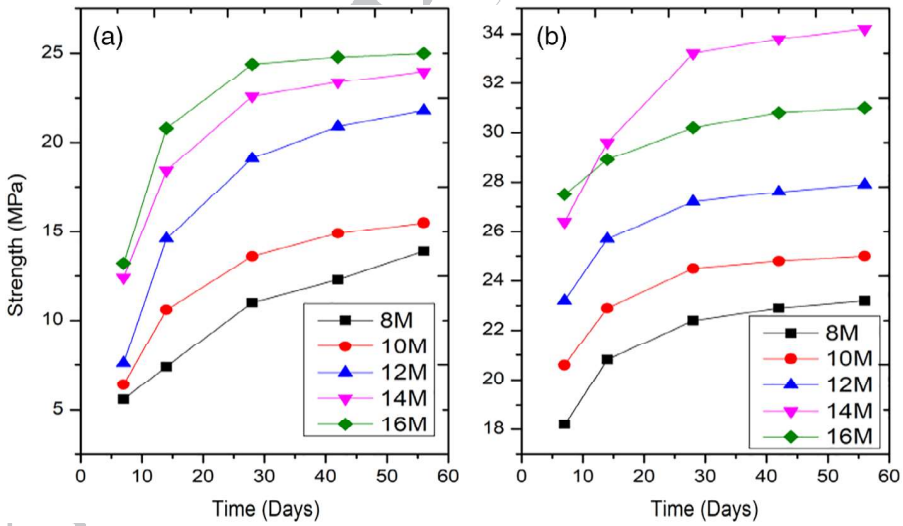


FIGURE 5 Splitting strength of (a) ambient cured GPC and (b) oven cured GPC

Color Figure - Print and Online

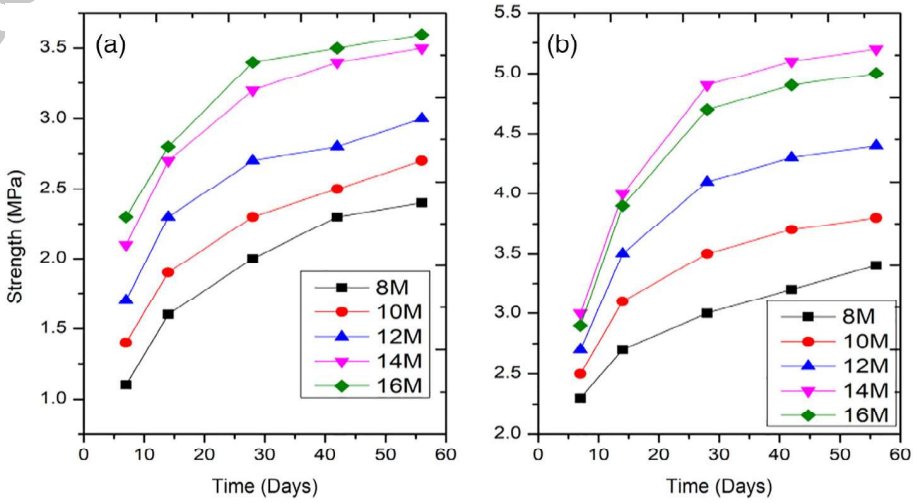


FIGURE 6 Flexural strength of (a) ambient-cured GPC and (b) oven-cured GPC

Uncorrected Proofs

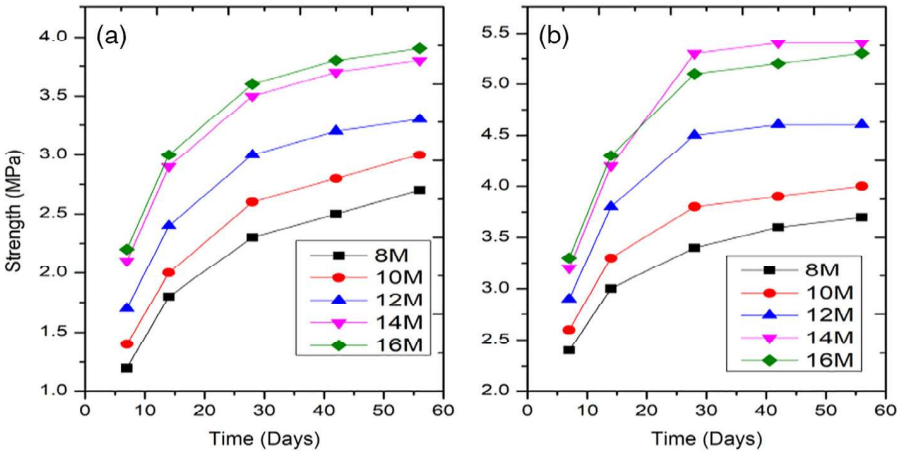


FIGURE 7 Graph between density and molarity of NaOH of GPC

16 M are 2.4 N/mm^2, 2.7 N/mm^2, 3.0 N/mm^2, 3.5 N/mm^2 and 3.6 N/mm^2 respectively at 56 days after ambient-curing but after oven-curing splitting tensile strength of 8 M, 10 M, 12 M, 14 M and 16 M are 3.4 N/mm^2, 3.8 N/mm^2, 4.4 N/mm^2, 5.2 N/mm^2 and 5.0 N/mm^2 respectively at 56 days.

3.1.3 | Flexural strength

Flexural tensile strength to find bending tension of the beam specimen. Figure 7 shows the flexural tensile strength of mixes ambient cured samples and oven cured samples, respectively. Similarly, the oven-cured samples got more flexural tensile strength as compared to the ambient-cured specimens, the highest strength of oven-cured specimens of 14 M mix is 5.4 N/mm^2 at 56 days, but in ambient-cured specimens, the most upper strength of 16 M mix is 3.9 N/mm^2. In both curing conditions, the flexural tensile strength increases with the increment of the molarity of NaOH in mix design but in the oven-cured strength decreases beyond the 14 M NaOH in mixes. The highest flexural tensile strength of 8 M, 10 M, 12 M, 14 M, and 16 M are 2.7 N/mm^2, 3.0 N/mm^2, 3.3 N/mm^2, 3.8 N/mm^2, and 3.9 N/mm^2, respectively at 56 days after ambient-curing but after oven-curing flexural tensile strength of 8 M, 10 M, 12 M, 14 M, and 16 M are 3.7 N/mm^2, 4.0 N/mm^2, 4.6 N/mm^2, 5.4 N/mm^2, and 5.3 N/mm^2, respectively at 56 days.

3.1.4 | Density, Poisson ratio, and modulus of elasticity

Figure 8 presented the density of the mix designs to the analysis of the effect of the molarity of NaOH in mix designs of GPC specimens. All the specimens of various mix designs with the variation of molarity 8–16 M cured

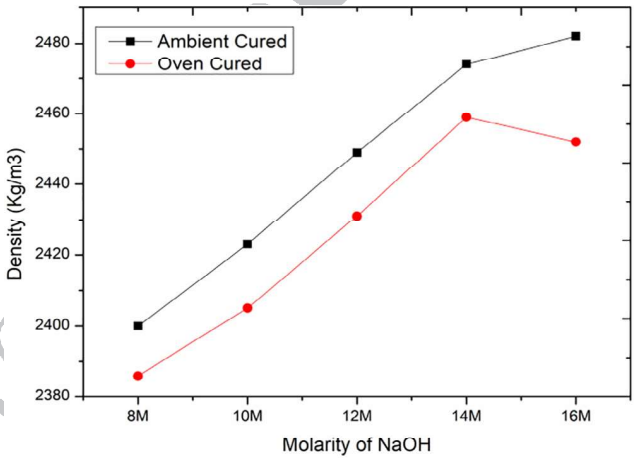


FIGURE 8 Graph between the modulus of elasticity and molarity of NaOH of GPC

under both conditions ambient and oven-curing. Analysis of the variation in results of the Poisson ratio, modulus of elasticity and density drawn out from the samples testing. Density, Poisson ratio and modulus of elasticity tested after the 28 days of the casting. Density finds out through the weight of the cube specimens of the various mixes. The Poisson ratio finds through the testing of the cylindrical specimens tested in the UTM machine by applying statistical compression loading along the longitudinal direction and got the vertical and horizontal strain. The Elastic modulus finds through the stress–strain curve made by the cyclic loading on the cylindrical specimen about one-third of the fracture load statically. The ambient cured sample has a higher density than the oven-cured GPC specimens of different mix designs. A mix of molarity of NaOH in mixes rise in the density of the design mix specimens, but after the increase beyond the 14 M slightly decreases the density of the mix designs specimens. 16 M mix designs samples got the highest average density of 2,482 kg/m^3 at 28 days of ambient

Color Figure - Print and Online
Uncorrected Proofs

cured specimens, whereas the oven-cured samples got the average density of 2,474 kg/m^3. The Poisson ratio of the ambient-cured GPC specimens are slightly excessive than oven-cured specimens, but the modulus of elasticity of the GPC mix designs higher for the oven-cured samples.

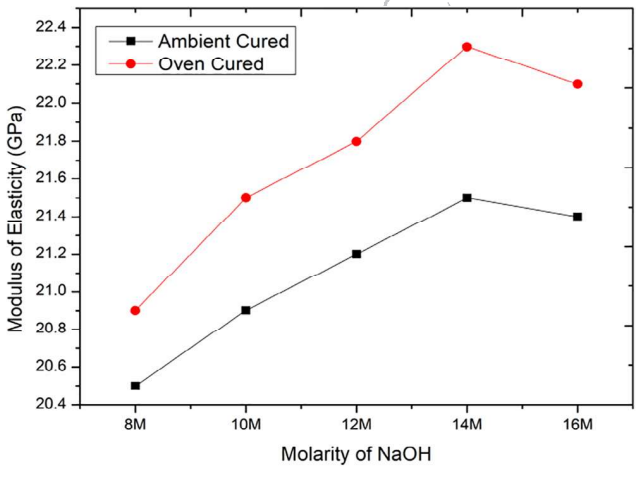
Figure 9 shows the maximum modulus of elasticity of the 14 M mix design of oven cured specimens is 22.3 GPa, whereas the ambient cured specimens modulus of elasticity is 21.5 GPa of the 14 M mix design of GPC.

3.2 | Effect of sodium silicate to sodium hydroxide (Na_2SiO_3/NaOH) ratio

Sodium silicate works as a binder (flyash and GGBFS) activation with the sodium hydroxide in the mix designs of the GPC, the changes in the ratio of Na_2SiO_3/NaOH directly affect the properties of the mix design of the GPC. Analysis, the effect of the rate of Na_2SiO_3/NaOH in mix designs of the GPC by varying the ratio 0.5–3.0 and examine the test results of the specimens of all design mixes by the destructive strength test.

3.2.1 | Compressive strength

The compressive strength of the GPC varies with the ratio of Na_2SiO_3/NaOH and curing condition. The compressive strength increases with the increment in the ratio in the mix design but beyond a point decreases the compressive strength in both curing condition samples. The highest compressive strength of oven cured 14 M mix is 35.7 N/mm^2 at 56 days and in the ambient cured specimen got 25.8 N/mm^2 at 56 days. Figure 10a shows the compressive strength of ambient cured specimens and in Figure 10b shows the compressive strength of oven cured specimens with the ratio of Na_2SiO_3/NaOH in design mix of GPC. The highest compressive strength of design mixes of ratio 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0 are 20.3 MPa, 24.0 MPa, 25.0 MPa, 25.4 MPa, 25.8 MPa, and 23.4 MPa respectively at 56 days after ambient-curing but after oven-curing compressive strength of the ratio 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0 are 24.6 MPa, 34.2 MPa, 34.7 MPa, 35.1 MPa, 35.7 MPa, and 32.5 MPa, respectively at 56 days.



Color Figure – Print and Online

FIGURE 9 Compressive strength of (a) ambient-cured GPC and (b) oven-cured GPC

3.2.2 | Splitting strength

The splitting tensile strength test to find the indirect tensile strength of the design mix of the GPC. Figure 11 shows the splitting tensile strength of ambient-cured specimens and oven cured samples of mix designs, in which varies the ratio of Na_2SiO_3/NaOH in the mix design of GPC from 0.5 to 3.0. In the both curing conditions specimens got the highest splitting tensile strength at the ratio of 2.5 and the uppermost strength of ambient cured and oven cured specimens are 4.0 MPa, 5.2 MPa respectively at 56 days in the mixes of GPC. The highest splitting tensile strength of design mixes of ratio 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0 are 2.7 MPa, 3.5 MPa, 3.6 MPa, 3.8 MPa, 4.0 MPa, and 3.6 MPa, respectively at 56 days after ambient-curing but after oven-curing compressive strength of the ratio 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0 are 3.5 MPa, 5.2 MPa, 5.0 MPa, 5.1 MPa, 5.2 MPa, and 5.0 MPa, respectively at 56 days.

3.2.3 | Flexural strength

Flexural tensile strength to find bending tension of the beam specimen. Figure 12 shows the flexural tensile strength of mixes ambient cured samples and oven cured samples, respectively. Similarly, the oven-cured samples got more flexural tensile strength as compared to the ambient-cured specimens, the highest strength of oven-cured specimens of the ratio of 2.5 in the mix is 5.6 MPa at 56 days, but in ambient cured samples the most upper strength of the ratio 2.5 in the mix is 4.0 MPa. In both curing conditions, the flexural tensile strength rises with the increment of the rate of sodium silicate to sodium hydroxide in the mix design but in the oven-cured strength decreases beyond the ratio of 2.5 in design mixes. The highest flexural tensile strength of design mixes of ratio 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0 are 3.0 MPa, 3.8 MPa, 4.0 MPa, 4.0 MPa, 4.0 MPa, and 3.9 MPa, respectively at 56 days after ambient-curing but after

Uncorrected Proofs

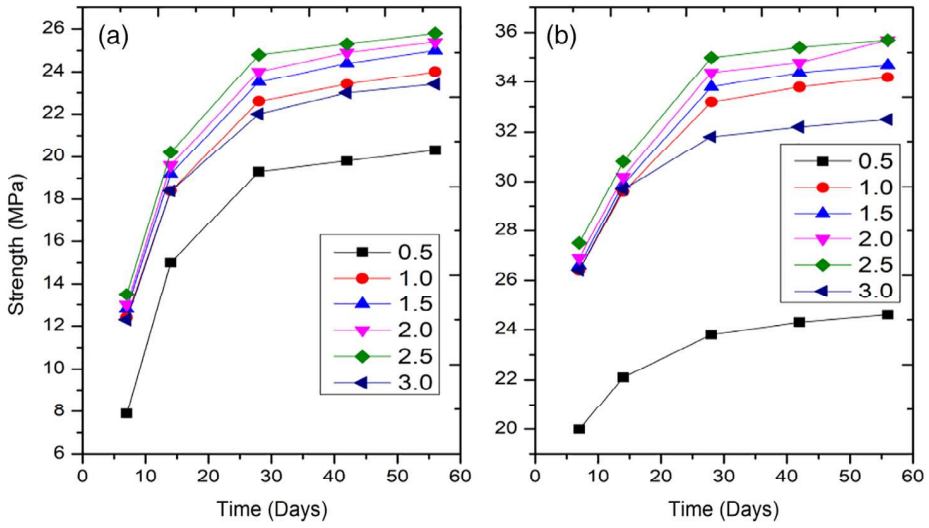


FIGURE 10 Splitting strength of (a) ambient cured GPC and (b) oven-cured GPC

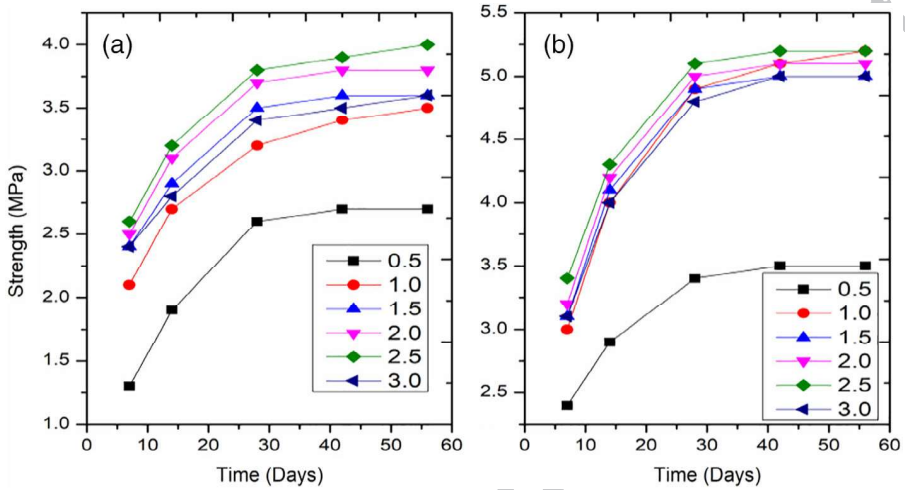


FIGURE 11 Flexural strength of (a) ambient-cured GPC and (b) oven-cured GPC

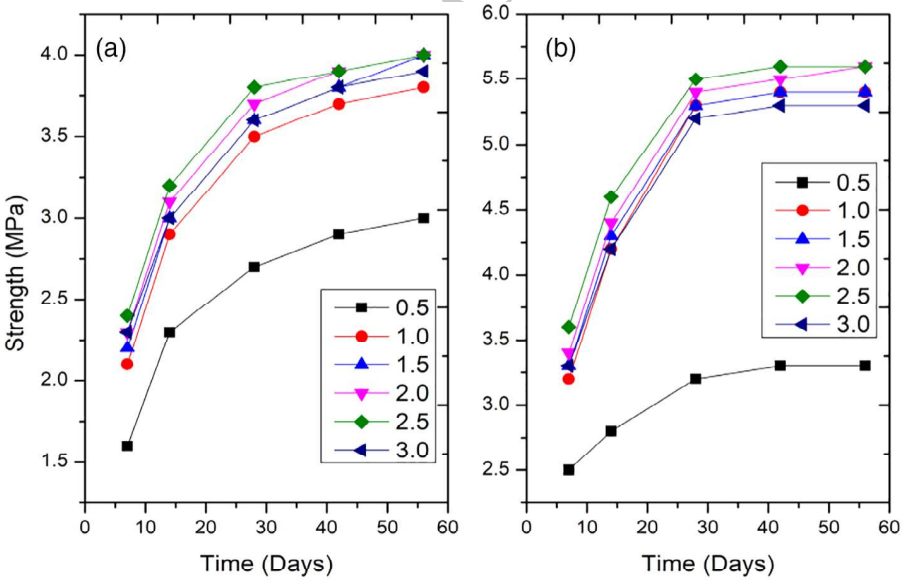
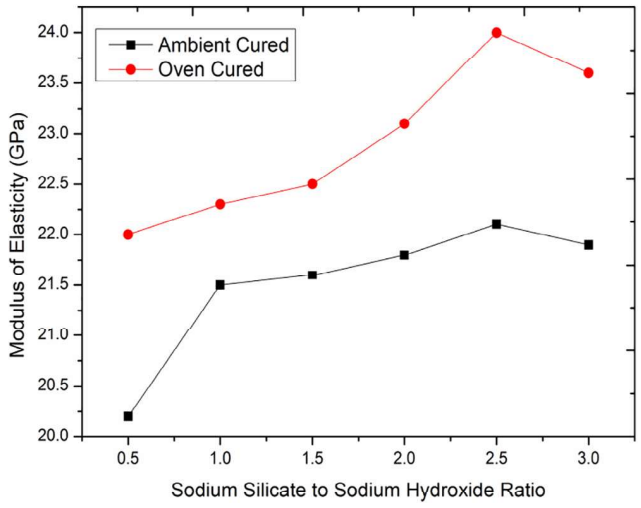


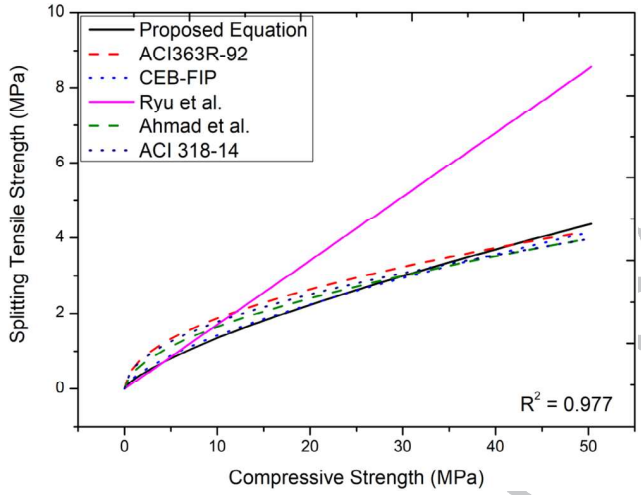
FIGURE 12 Graph between the density and Na_2SiO_3/NaOH ratio in GPC

Color Figure - Print and Online
Color Figure - Print and Online
Color Figure - Print and Online
Proofs



Color Figure - Print and Online

FIGURE 13 Graph between the modulus of elasticity and Na_2SiO_3/NaOH ratio in GPC



Color Figure - Print and Online

FIGURE 14 Relationship between splitting tensile strength and compressive strength

oven-curing compressive strength of the ratio 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0 are 5.4 MPa, 5.4 MPa, 5.6 MPa, 5.6 MPa, 5.3 MPa, and 5.0 MPa, respectively at 56 days.

3.2.4 | Density, Poisson ratio, and modulus of elasticity

Figure 13 represent the effect of the ratio of Na_2SiO_3/NaOH on the density of mix design of GPC; Ambient cured specimens have higher density compared to oven-cured samples and got the highest density are 2,494 kg/m^3 at the ratio of 2.5 of ambient cured specimens. Figure 14 represents the effect of the ratio of Na_2SiO_3/NaOH on the modulus of elasticity of mix design of GPC. Oven cured samples have an excessive modulus of elasticity as compared to the ambient cured samples and got the highest modulus of elasticity at the ratio of 2.5 are 24 GPa of oven cured samples.

4 | CORRELATION BETWEEN THE MECHANICAL PROPERTIES

After the long-term experimental investigation, GPC mechanical properties show the correlation to each other. Based on the results out from the destructive experiment test of various mixes, proposed the correlation equation among the compressive strength, splitting strength and flexural strength.

4.1 | Relationship between splitting tensile strength and compressive strength

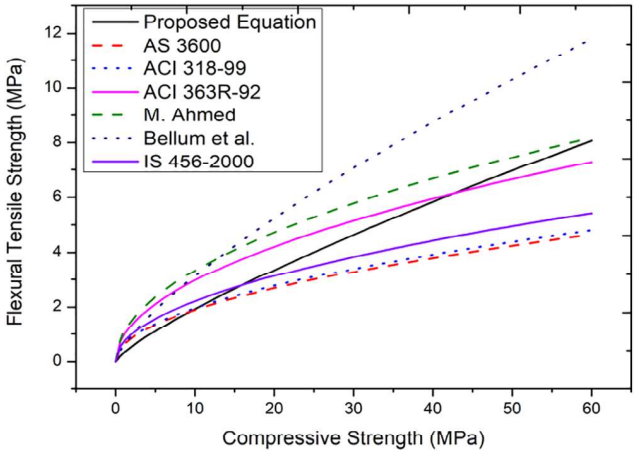
The splitting tensile strength of the GPC measured through the cylindrical specimen of 150 mm dia × 300 mm height by a static load on the transverse side on the specimens and compressive strength from the 150 mm × 150 mm size cube test in the UTM machine in the laboratory. From the results, a correlation Equation (1) proposed between the splitting and compressive strength and compared with the other correlation equations given by the code or other authors. The ACI committee gives Equation (2) in the state of the art of high strength concrete[24] and a committee of euro code international gives Equation (3) in the model code 1990.[25] Equation (4) introduced by an ACI standard and report for building code requirement for structural concrete[26] and Ahmed et al. proposed Equation (5) in the international journal of structural engineering in 2014.[27] Ryu et al. introduce Equation (6) of the correlation of flyash based GPC in the journal construction and building material in 2013.[10] All the correlation are very similar and shows that splitting strength are around to12-13% of the compressive strength of the mixes specimens. Figure 15 graph shows the various equations and compared with the proposed equation of empirical correlation.

$$\text{Proposed Eq.}: f_{st} = 0.228 f_c^{0.82} \quad (1)$$

$$\text{ACI363R} - 92: f_{st} = 0.59\sqrt{f_c} \quad (2)$$

$$\text{CEB} - \text{FIP}: f_{st} = 0.301 f_c^{0.67} \quad (3)$$

Uncorrected Proofs



Color Figure : Print and Online

FIGURE 15 Relationship between flexural strength and compressive strength

$$\text{ACI } 318-14: f_{st} = 0.56 f_c^{0.5} \quad (4)$$

$$\text{Ahmed et al.}: f_{st} = 0.462 f_c^{0.55} \quad (5)$$

$$\text{Ryu et al.}: f_{st} = 0.17 f_c^{0.75} \quad (6)$$

4.2 | Relationship between flexural strength and compressive strength

Flexural strength of the concrete is fundamental in structural engineering because it shows the tensile strength of concrete and bending behavior of the specimens. From the experimental results, proposed the correlation Equation (7) between the Flexural strength and compressive strength of the mixes and compared to the other correlation Equations (8)–(13) given by the codes and authors. ACI committee introduces Equation (8) in the publication of the state of the art of high strength concrete.[24] A standard report presented on the code for the design of concrete in the structure introduces Equation (9) of correlation.[26] Australian code for the reinforced concrete design presented the correlation Equation (10), and the Indian code introduced Equation (11) between the flexural strength and compressive strength.[28,29] Ahmed et al. proposed Equation (12) in the international journal of structural engineering and Bellum et al. also proposed Equation (13) between the correlation of flexural strength and compressive strength.[27,30] Figure 16 shows the proposed equation shows that the flexural strength is around 15–16% of the compressive strength of the same mix of GPC.

$$\text{Proposed Equation}: f_{fs} = 0.3066 f_c^{0.799} \quad (7)$$

$$\text{ACI363R} - 92: f_{fs} = 0.94\sqrt{f_c} \quad (8)$$

$$\text{ACI } 318-99: f_{fs} = 0.62\sqrt{f_c} \quad (9)$$

$$\text{AS } 3600: f_{fs} = 0.6\sqrt{f_c} \quad (10)$$

$$\text{IS } 456-2000: f_{fs} = 0.7\sqrt{f_c} \quad (11)$$

$$\text{M.Ahmed}: f_{fs} = 1.055\sqrt{f_c} \quad (12)$$

$$\text{Bellum et al.}: f_{fs} = 0.57 f_c^{0.74} \quad (13)$$

where f_c are characteristic strength in MPa, f_{fs} are flexural strength in MPa.

5 | CONCLUSIONS

The conclusion drained through the testing of the various mix designs specimens in the form of cube cylinder and prism molds by the destructive and nondestructive testing of ambient-cured and oven-cured specimens, in which the molarity of NaOH and Na_2SiO_3/NaOH varying in the design mix of GPC.

- The compressive strength rises with the enlargement in molarity of NaOH in the mix design but beyond a point decreases the compressive strength in the oven-cured specimens. The highest compressive strength of oven cured 14 M mix is 34.2 N/mm^2 at 56 days but in the ambient-cured specimen got at 16 M mix is 25 N/mm^2 at 56 days.
- In the ambient cured specimens 16 M got the highest splitting tensile strength is 3.6 N/mm^2 at 56 days, but in oven-cured samples 14 M got the most upper splitting tensile strength is 5.2 N/mm^2 at 56 days.
- Similarly, the oven-cured specimens got more flexural tensile strength as compared to the ambient-cured specimens, the highest strength of oven-cured specimens of 14 M mix are 5.4 N/mm^2 at 56 days, but in ambient-cured specimens, the most upper strength of 16 M mix is 3.9 N/mm^2.
- The compressive strength enlarges with the increment in the ratio in the mix design but beyond a point decreases the compressive strength in both curing condition samples. The highest compressive strength of oven cured 14 M mix is 35.7 N/mm^2 at 56 days and in

Uncorrected Proofs

the ambient cured specimen got 25.8 N/mm^2 at 56 days.

- In the both curing conditions specimens got the highest splitting tensile strength at the ratio of 2.5 and the uppermost strength of ambient cured and oven cured specimens are 4.0 MPa, 5.2 MPa respectively at 56 days in the mixes of GPC.
- In both curing conditions, the flexural tensile strength rises with the increment of the rate of sodium silicate to sodium hydroxide in mix design but in the oven-cured strength decreases beyond the ratio of 2.5 in design mixes.
- Based on results, proposed the correlation Equations (1) between the splitting strength and compressive strength are $f_{st} = 0.228 f_c^{0.82}$ and Equation (7) between the flexural strength and compressive strength are $f_{fs} = 0.3066 f_c^{0.799}$.

ACKNOWLEDGMENTS

This work supported by the Civil Engineering Department, Delhi Technological University funded by the Delhi Government. Moreover, the help of Sandeep Mishra for XRD analysis and Ankesh Kumar and Jawed Alam for the another in the work that is related to the central instrumentation facility of the Delhi Technological University.

ORCID

Manvendra Verma https://orcid.org/0000-0002-0840-841X

REFERENCES

1. Chindaprasirt P, Chareerat T, Sirivivatnanon V. Workability and strength of coarse high calcium fly ash geopolymer. Cem Concr Compos. 2007;29:224–229.
2. Nath P. School of Civil and Mechanical Engineering Study of Fly Ash Based Geopolymer Concrete Cured in Ambient Condition; 2014.
3. Yang K-H, Song J-K, Lee K-S, Ashour AF. Flow and compressive strength of alkali-activated mortars. ACI Mater J. 2009;106 (1):50–58.
4. Yang K-H, Song J-K. Workability loss and compressive strength development of cementless mortars activated by combination of sodium silicate and sodium hydroxide. J Mater Civ Eng. 2009;21:119–127.
5. Temuujin J, Van Riessen A, Mackenzie KJD. Preparation and characterisation of fly ash-based geopolymer mortars. Construct Build Mater. 2010;24(10):1906–1910.
6. Shu-guang HU, Jing WU, Wen Y, Yong-Jia HE. Preparation and properties of geopolymer-lightweight aggregate refractory concrete. J Cent South Univ Technol. 2009;16:914–918.
7. Patankar SV, Ghugal YM, Jamkar SS. Mix design of fly ash-based geopolymer concrete. Advances in structural engineering: Materials, volume three, 2015.
8. Patankar SV, Ghugal YM, Jamkar SS. Effect of concentration of sodium hydroxide and degree of heat curing on fly ash-based geopolymer mortar effect of concentration of sodium hydroxide and degree of heat curing on fly ash-based geopolymer mortar. Indian J Mater Sci. 2014;2014:938789.
9. Lee NK, Lee HK. Setting and mechanical properties of alkali-activated fly ash/slag concrete manufactured at room temperature. Construct Build Mater. 2013;47:1201–1209.
10. Ryu GS, Lee YB, Koh KT, Chung YS. The mechanical properties of fly ash-based geopolymer concrete with alkaline activators. Construct Build Mater. 2013;47:409–418.
11. Awoyera P, Adesina A. A critical review on the application of alkali-activated slag as a sustainable composite binder. Case Stud Constr Mater. 2019;11:e00268.
12. B. Statements. Standard Test Method for Surface Moisture in Fine Aggregate 1. pp. 1–3, 2019.
13. C. Ag, B. Statements. Standard Test Method for Relative Density (Specific Gravity) and Absorption of Coarse. pp. 1–5, 2019.
14. B. Statements. Standard Test Method for Materials Finer than 75-μm (No. 200) Sieve in Mineral Aggregates by Washing 1, no. 200, pp. 7–10, 2019.
15. C. Ag, B. Statements, T. A. Evaluation. Standard Test Method for Bulk Density ('Unit Weight') and Voids in Aggregate 1. pp. 1–5, 2019.
16. B. Statements, C. Ag. Standard Test Method for Sieve Analysis of Fine and Coarse Aggregates 1. pp. 1–5, 2019.
17. S. Container. Standard Test Method for Moisture in Mineral Aggregate Used on Built-Up Roofs 1, Vol. 89, no. Reapproved, 2017, pp. 1–2, 2019.
18. C. Ag, B. Statements, W. Pycnometer. Standard Test Method for Relative Density (Specific Gravity) and Absorption of Fine, Vol. I, pp. 1–6, 2019.
19. C. Aggregates. Standard Specification for Chemical Admixtures for Concrete 1. pp. 1–10, 2019.
20. P.-I. I. Metals. Yearbook, 2015, Vol. 2015, no. 0712, pp. 1–10, 2016.
21. T. Fly, C. Ag. Standard Specification for Coal Fly Ash and Raw or Calcined Natural Pozzolan for Use, pp. 1–5, 2019.
22. C. Ag, T. C. Vertically, M. Cabinets, M. Rooms, M. H. Concrete, T. A. Evaluation, Standard Practice for, pp. 1–8, 2019.
23. I. Standard. Specification for Moulds for Use in Tests of Cement and Concrete, no. 1952, 2008.
24. Russell HG, Anderson AR, Banning JO, et al. State-of-the-Art Report on High-Strength Concrete Reported by ACI Committee 363, vol. 92, no. Reapproved, 1997.
25. ceb-fip-model-code-1990-design-code.pdf, 1990.
26. A. C. I. Committee. Building Code Requirements for Structural Concrete (ACI 318-14), 2014.
27. Ahmed M, El Hadi KM, Mallick J, Hasan MA, Ahmed A. Evaluating the co-relationship between concrete flexural tensile strength and compressive strength. Int J Struct Eng. 2014;5(2): 115–131.
28. C. Concrete. Reinforced Concrete Design in accordance with AS 3600—2009, 2009.
29. O. F. I. Standards. Plain and Reinforced Concrete – Code of Practice, no. July 2000.
30. Bellum RR, Muniraj K, Rama S, Madduru C. Empirical relationships on mechanical properties of class-F fly ash and GGBS based geopolymer concrete. Ann Chim – Sci Matér. 2019;43(3): 189–197.

Uncorrected Proofs

AUTHOR BIOGRAPHIES



Manvendra Verma,
Department of Civil Engineering,
Delhi Technological University,
Delhi, India.
Email: manvendraverma_2k16phdc-e05@dtu.ac.in



Nirendra Dev,
Department of Civil Engineering,
Delhi Technological University,
Delhi, India.
Email: nirendradev@dce.ac.in

How to cite this article: Verma M, Dev N. Sodium hydroxide effect on the mechanical properties of flyash-slag based geopolymer concrete. *Structural Concrete*. 2020;1–12. https://doi.org/10.1002/suco.202000068

Uncorrected Proofs

geosciences

MDPI

Article

Spatial Landslide Risk Assessment at Phuentsholing, Bhutan

Abhirup Dikshit [1], Raju Sarkar [2,3,*], Biswajeet Pradhan [1,4], Saroj Acharya [2] and Abdullah M. Alamri [5]

1 Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), University of Technology Sydney, Sydney, NSW 2007, Australia; abhirupdikshit@gmail.com (A.D.); biswajeet.pradhan@uts.edu.au (B.P.)
2 Center for Disaster Risk Reduction and Community Development Studies, Royal University of Bhutan, Phuentsholing 21101, Bhutan; sarojzacharya12@gmail.com
3 Department of Civil Engineering, Delhi Technological University, Bawana Road, Delhi 110042, India
4 Department of Energy and Mineral Resources Engineering, Sejong University, Choongmu-gwan, 209 Neungdong-ro, Gwangjin-gu, Seoul 05006, Korea
5 Department of Geology & Geophysics, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia; alamri.geo@gmail.com
* Correspondence: rajusarkar@dce.ac.in

Received: 13 February 2020; Accepted: 6 April 2020; Published: 7 April 2020



Abstract: Landslides are one of the most destructive and most recurring natural calamities in the Himalayan region. Their occurrence leads to immense damage to infrastructure and loss of land, human lives, and livestock. One of the most affected regions is the Bhutan Himalayas, where the majority of the landslides are rainfall-induced. The present study aims to determine the hazard and risk associated with rainfall-induced landslides for the Phuentsholing region located in the southwestern part of the Bhutan Himalayas. The work involves developing a landslide risk map using hazard and vulnerability maps utilizing landslide records from 2004 to 2014. The landslide hazard map was generated by determining spatial and temporal probabilities for the study region. The spatial probability was computed by analyzing the landslide contributing factors like geology, slope, elevation, rainfall, and vegetation based on comprehensive field study and expertise about the area. The contributing factors were divided into various classes and the percentage of landslide occurrence under each class was calculated to understand its contributing significance. Thereafter, a weighted linear combination approach was used in a GIS environment to develop the spatial probability map which was multiplied with temporal probabilities based on regional rainfall thresholds already determined for the region. Consequently, vulnerability assessment was conducted using key elements at risk (population, land use/land cover, proximity to road, proximity to stream) and the weights were provided based on expert judgment and comprehensive field study. Finally, risk was determined and the various regions in the study area were categorized as high, medium, and low risk. Such a study is necessary for low-economic countries like Bhutan which suffers from unavailability of extensive data and research. The study is conducted for a specific region but can be extended to other areas around the investigated area. The tool can serve as an indicator for the civil authorities to analyze the risk posed by landslides due to the rapid infrastructure development in the region.

Keywords: landslide risk; GIS; hazard; Bhutan

1. Introduction

Landslides are considered one of the most prominent and devastative natural disasters leading to immense loss of human lives and infrastructure [1]. The triggering factor for the majority of the

landslides is either rainfall or earthquake of which rainfall-triggered landslides can be categorized as very high and intense. The Himalayan region experiences very high rainfall and contributes over 70% of global fatal landslides [2]. Landslides in the Himalayas occur mainly due to monsoonal precipitation leading to huge financial loss and often disrupting the only livelihood of people [3,4]. The global fatal landslide database developed by [2] shows a total of 64 deaths in Bhutan due to nonseismic landslide events from 2004 to 2017, but the actual number of deaths is expected to be very high. The menace is expected to escalate because of the cutting of hills and deforestation emphasizing infrastructure development linked to a rise in population.

The studies on rainfall-induced landslides are quite varied and focus on several aspects like threshold modeling [5,6], monitoring systems [4,7], and computational techniques [8], along with GIS approach. Apart from landslide forecasting, landslide mitigation is important to understand the landslide hazard and the associated risk. Several works have been carried out to develop risk maps using the topography of the area susceptible to landslides [9–11]. Others have used ground conditions like the combined effect of rainfall and topography of the region to develop risk maps (South Italy [12], San Francisco Bay area [13]). Ho et al. [14] and Wong [15] used population density along with the natural contributing landslide factors to develop risk maps. Hadmoko et al. [16] utilized a vector-based GIS method to analyze several factors affecting landslides to develop hazard maps of the Central Java region. Pradhan and Abdulwahid [17] developed landslide risk maps using logistic regression and multihazard cases developed by considering precipitation data. Althuwaynee and Pradhan [11] developed a semi-quantitative landslide risk assessment using GIS-based analysis for Malaysia and were able to successfully show the predictive capability of the model. Liu and Miao [18] determined the landslide hazard, vulnerability, and risk in China by using certainty factor and a logistic regression model. Goorchi et al. [19] developed a new technique to analyze the risk linked to a rock slope, including the probability of failure and vulnerability rates. Saldivar-Sali and Einstein [20] proposed a regional landslide risk rating system in the Philippines. Aljeano et al. [21] developed a similar approach to analyze rock fall risk in quarries and the results depicted close correlation to the experimental work conducted for unstable slopes typically used for road construction. The effect of soil properties on slope instability has also been studied by several researchers using reliability-based optimization techniques, wherein the spatial variability related to shear parameters of soil is analyzed [22,23]. Cardarilli et al. [24] conducted an investigation to understand the territorial resilience towards landslide in the Southern Italy region by analyzing the correlations among the causative factors with an aim to reduce loss of lives and economic damage.

All these aforementioned studies have one thing in common, that is, the availability of sufficient data and prior research work. The present study has been performed for the Bhutan Himalayas, which lacks in both aspects. The research conducted in the past is so minimal that a Scopus search using the keywords “landslide*” and “Bhutan” resulted in only 36 documents. The publication of the latest landslide global disaster database has led several researchers to focus on this area and in the recent years, various research works have been conducted to better understand the landslides in the region. The majority of the landslide incidences are triggered due to rainfall, therefore we have focused solely on rainfall-induced landslides. The previous research in Bhutan has been towards forecasting, which led to the development of several thresholds using various statistical [3,25,26] and probabilistic [27] approaches. However, landslide mitigation and understanding the elements at risk are yet to be conducted for any part of Bhutan. Therefore, the present study is a preliminary step towards landslide risk assessment with an aim to perform a complete and comprehensive risk assessment in the future. The study was focused only on the Phuentsholing region, primarily due to two main reasons, the first being its significance as a major trade route and the second being the availability of sufficient data to conduct risk analysis. The work presented reflects a very simple but effective approach to determine landslide risk by computing hazard and vulnerability. Hazard comprises three different components (spatial, temporal, and magnitude) and these are multiplied directly, as an independent relationship is assumed between the components [28,29]. However, the unavailability of a detailed landslide inventory

map results in the exclusion of the magnitude component. In the present study, the spatial component was calculated using a weighted sum analysis based on landslide contributing factors which was multiplied with the temporal probabilities determined based on regional rainfall thresholds [26] to develop the landslide hazard map. The predisposing factors for spatial aspect were identified based on a comprehensive field study conducted by the authors and examining the historical landslide locations in the study area. Landslide vulnerability was computed using weighted overlay analysis on key elements at risk by providing weights based on expert opinion. Finally, a landslide risk map was developed by combining both the hazard and vulnerability maps. Such an approach is useful for a country like Bhutan which lacks a proper landslide mitigation taskforce along with an operational early warning landslide system [30]. The present study would give an opportunity to develop risk assessment tools to analyze the slide activity and provide recommendations for mitigation purposes.

2. Study Area

The study area is Phuentsholing, situated in southwestern Bhutan, and is located from 26°50′ N to 26°59′ N latitude and 89°15′ E to 89°28′ E longitude covering an area of around 131.7 km^2. The area is further divided into six subdivisions (known as chiwog), which are Lingdaen, Pachhu, Dzongkhag Tromde, Dophugchen Wangduegatshel, Deling Marpji, andChong Geykha Dophulakha (Figure 1).

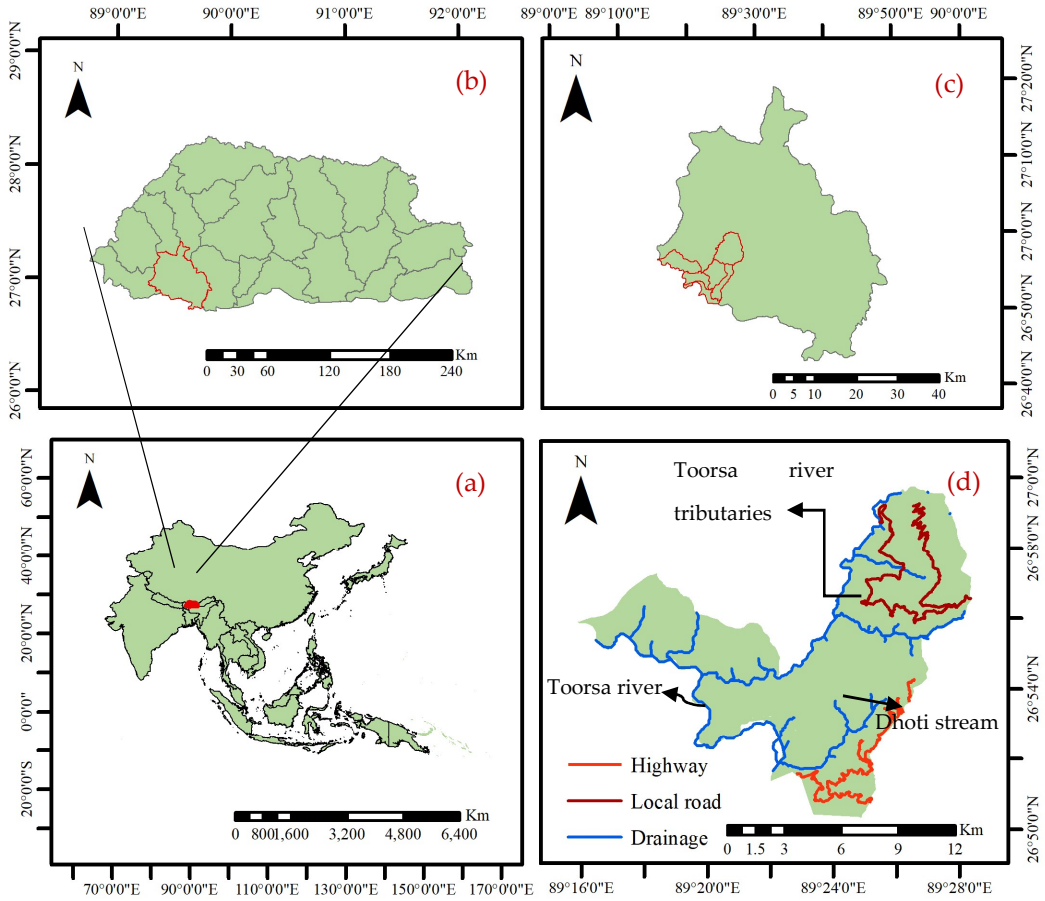


Figure 1. Location of the study region showing (**a**) Bhutan in Southeast Asia, (**b**) Bhutan highlighting Chukha Dzongkhag, (**c**) Chukha Dzongkhag highlighting Phuentsholing region, and (**d**) territorial boundaries of the study region.

Landslides in the study region are mainly due to intense monsoon rainfall, weak geology, and anthropogenic activities. The landslide typology in the study area can be characterized as rock fall, rockslide, debris flow, debris slide, and earth slide [24,25]. Gansser [31] described the regional

geological setting of the Bhutan Himalayas and characterized the majority of the country as part of the Himalayan Crystalline Complex. The areas that are most prone to landslides are heavily fractured and weathered rocks of phyllite, slate, and schist that contain high amounts of clay [26,27]. The study region has been identified as geologically weak with the presence of several shear zones, including weathered and folded phyllite rocks turning to clay [25].

The Phuentsholing-Thimphu highway, which is a part of the study region, is a vital lifeline as the majority of the import and export goods of the country are carried out through this route. The landslides lead to blockage of roads at several sections leading to a considerable amount of economic losses in repair and maintenance every year. The highway was the first motorable road of the country, constructed in 1961, and it crosses through areas susceptible to landslide, subsidence, erosion, and rockfalls. The instability in certain sections of the road has been as long as 20–50 years old. A 2005 report on the landslide investigations along the highway showed that a landslide at a 360 m section of the highway between the 17 km and 18 km point, known as Sorchen landslide, was so severe that the road alignment, which had four bends, eventually was left with only one bend over time. Figure 2 depicts two damage pictures taken during landslide events in Chukha Dzongkhag along the Phuentsholing-Thimphu Highway after the 2017 monsoon.



Figure 2. Landslide damage along Phuentsholing-Thimphu Highway (after 2017 monsoon).

3. Data Used

The understanding of various causative factors and preparation of thematic layers are necessary to develop hazard and vulnerability maps. In the present study, a risk map was developed by analyzing the hazard contributory factors (geology, slope, elevation, vegetation, rainfall) with vulnerability factors (land use, population, proximity to road and stream). Hazard has been defined as a probability of a specific danger likely to occur in a given time period whereas risk refers to hazard multiplied by the potential loss of life [20].

Geology is one of the most significant contributory factors to landslide initiation [32]. The region has four significant geological groups, which can be subcategorized into a total of seven types, which are: (i) Baxa Group (Phuentsholing Formation, Pangsari Formation), (ii) Daling–Shumar Group (Daling Formation, Orthogneiss unit, Shumar Formation), (iii) Lesser Himalayan Zone (Jaishidanda Formation), and iv) structurally lower Greater Himalayan section (lower metasedimentary unit) (Figure 3). The Baxa Group can be characterized as heavily fractured and weathered phyllites, slates, and schists [25]. The Daling–Shumar Group comprises tan weathering, cliff-forming quartzite and muscovite-biotite schist and phyllite [33]. The majority of the settlement area is dominated by the Baxa formation group wherein the thrust fault, strike, and dip of bedding and foliation are more concentrated. Similarly, such features were also delineated in the structurally lower Greater Himalayan Zone.

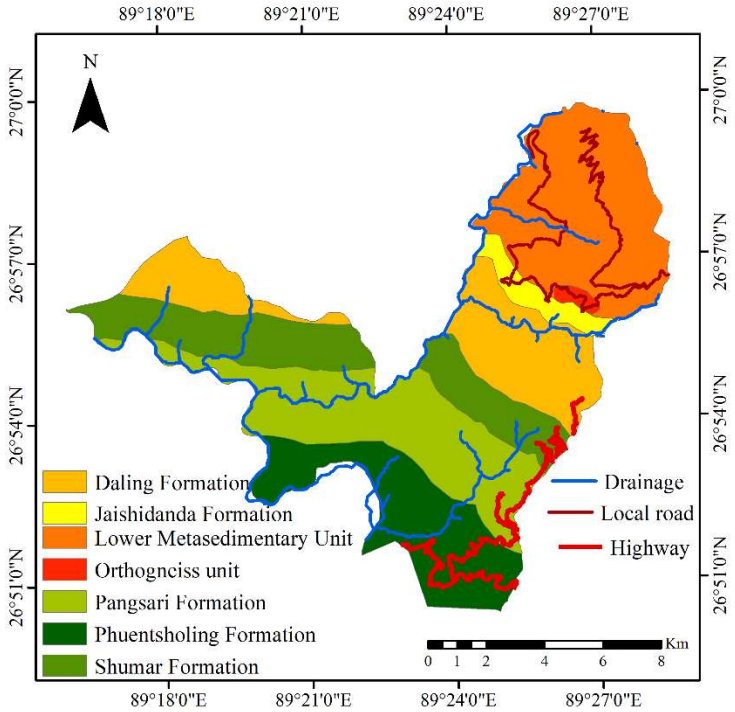


Figure 3. Geological map of the study area.

The other important factors are slope and elevation, which vary from 0°–75° and 183–2500 m, respectively. The majority of the study region (>80%) falls under 45° suggesting that the area is situated on a relatively flat plateau. More than half of the region has an elevation of less than 600 m and only 5% of the region is under high elevation zone (1500–2500 m).

The vegetation map was derived using the Normalized Difference Vegetation Index (NDVI) from the Landsat image (2017) and was classified into vegetated (86.76%) and nonvegetated areas (13.24%) (Figure 4c). NDVI always generates a value between −1 and +1 with high positive NDVI values (green) meaning high vegetation, while water usually has negative NDVI values (yellow and red), with urban features usually having values near zero. The maps were developed using the image analysis toolbar in ArcGIS 10.4.

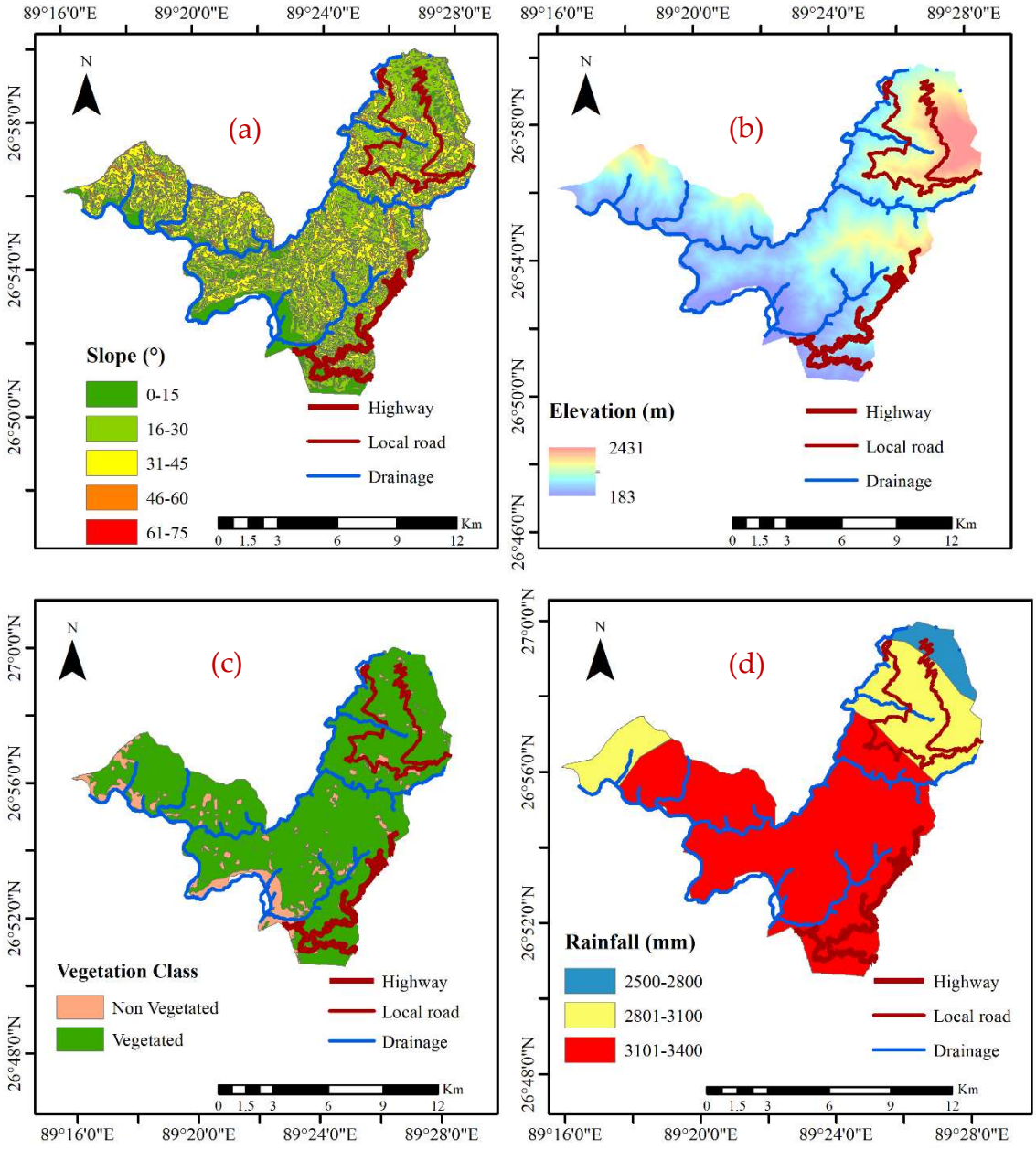


Figure 4. (**a**) Slope, (**b**) elevation, (**c**) vegetation, and (**d**) rainfall map of the study area.

The study area is located in the subtropical and temperate climatic zone, which receives very high annual rainfall up to 6000 mm. As per the Köppen climate classification [34], the climate is subtropical monsoonal receiving 10 times more rainfall during the wettest summer month as compared to dry winter months. The contribution of monsoonal rainfall (June–September) is 80% of the annual rainfall. However, pre- and post-monsoonal months can also suffer from heavy rainfall due to local convection and tropical cyclone activity over the Bay of Bengal [35], with January and February being relatively drier [25]. Figure 4d shows the rainfall map based on the annual rainfall measurements from 2004 to 2014 collected from the National Center for Hydrology and Meteorology of the Royal Government of Bhutan.

A total of 64 landslides occurred from 2004 to 2014, which was collected from Project Dantak, the Border Roads Organization, Government of India. The data included the coordinates of the landslide events which were collected from the scarp along with the dates of occurrences. The number of landslide events was counted as a single landslide event if multiple events occurred on the same

day and coordinates were selected where maximum damage or loss of land prevailed. Table 1 shows the number of landslides in each class of landslide contributing factors.

Table 1. Number of landslides under various factors.

Factor		Area (km^2)	Number of Landslides
Geology	Pangsari Formation	28.4	20
	Shumar Group	22.7	20
	Phuentsholing Formation	21.0	24
	Daling Formation	23.9	-
	Jaishidanda Formation	4.0	-
	Orthogneiss unit	0.9	-
	Lower metasedimentary unit	30.8	-
Slope (in °)	0–15	19.8	14
	16–30	62.4	31
	31–45	44.8	18
	46–60	4.6	1
	61–75	0.1	
Elevation (in m)	187–587	64.4	26
	588–994	45.5	38
	995–1499	17.7	-
	1500–2437	4.1	-
Rainfall (in mm)	2500–2800	5.8	-
	2801–3100	33.0	-
	3101–3400	92.9	64
Land Use/Land Cover Type	Forest	104.7	37
	Built up	1.4	7
	Agriculture	19.4	14
	Miscellaneous	6.2	6

4. Method

The method revolves around the determination of some important landslide risk factors (i.e., hazard, vulnerability, elements at risk, and potential damage). Landslide risk is defined as the maximum potential loss due to a landslide event for a definite area and during a certain period [18]. Hence, landslide risk can be calculated as:

$$\text{Risk} = \text{Hazard} \times \text{Vulnerability} \times \text{Elements at risk} \quad (1)$$

A complete quantitative risk analysis includes explicitly the economical exposure of elements at risk like buildings, infrastructures, and assets. However, due to the unavailability of such factors, population is used as a proxy to obtain a relative estimation of exposure and the final output is a map of qualitative risk classes.

The first factor is "hazard" which has been often erroneously used synonymously with "susceptibility" [36], but there are major differences between them. Landslide hazard is the probability of a landslide event of a certain magnitude within spatial and temporal limits [11], whereas landslide susceptibility is the likelihood of landslide occurrence depending on local conditions [37].

The present study determines the landslide hazard by considering spatial (P_s) and temporal (P_T) components [33,34]. The spatial aspect was extracted by analyzing all the landslide contributing factors and applying a weighted linear combination (WLC) method. WLC is quite often used by landslide researchers, but an accurate method of determining weights is not defined and is based on expert opinion [38]. It is based on determining weighted averages of landslide causative parameters pertinent to the study area and available data. The parameters are classified and multiplied with the provided weight in a GIS overlay environment, and the final output is determined by adding weighted averages.

The temporal probability was calculated using the rainfall threshold determined by Dikshit et al. [26] using the Poisson probability model. Readers are referred to [3,39,40] for the detailed explanation on determining the temporal landslide probability.

$$\text{Hazard (H)} = P_s \times P_T, \quad (2)$$

Landslide vulnerability is the degree of loss of a particular element at risk due to the occurrence of a certain magnitude of the event [41,42]. Glade and Crozier [43] highlighted that there is no fixed technique for quantitative vulnerability assessment of elements at risk with respect to different landslide magnitudes and types. Therefore, vulnerability is usually determined by either assigning a fixed value to elements at risk (0 for no damage and 1 for complete damage), or assigning values based on expert opinion and knowledge about the area [44]. Thereafter, the potential damage is calculated by combining the vulnerability and elements at risk. In the present study, the potential damage map was developed using the WLC method by providing the exposure of the weights to causative factors expressing both elements of vulnerability and exposure. The factors considered for vulnerability were land use/land cover (LULC), proximity to road, and proximity to stream, and for exposure was population.

The classification of derived thematic layers can be conducted using several techniques like equal intervals, natural breaks, and quantile and statistical techniques which are default GIS processes [45]. In the present study, the natural breaks method was used to classify values as it is quite popular among the researchers [45–47].

5. Results and Discussions

5.1. Landslide Hazard Assesment

As described in the methodology, the spatial component (susceptibility) was based on the WLC method and the temporal probability was determined based on the thresholds obtained using an empirical-based approach. The susceptibility map was based on landslide causative factors and the weights were provided as per Equation (3).

$$\text{LS} = \text{Rainfall} \times 0.3 + \text{Geology} \times 0.3 + \text{Elevation} \times 0.2 + \text{Slope} \times 0.1 + \text{Vegetation} \times 0.1 \quad (3)$$

where LS is landslide susceptibility. The weights are assigned using GIS overlay capability, based on its influence on landslide in comparison to other factors and performing a weighted sum to combine the factors [38]. The fault data were not added as the majority of the landslides are rainfall-triggered and can be attributed as a low-seismic region [48]. The temporal probability is calculated by multiplying annual exceedance probability and the probability of landslide occurrence. The annual exceedance probability is the probability of threshold being exceeded in a specific year, which is determined using the Poisson probability model. The maps derived using both the components resulted in the hazard map. Figure 5 represents the landslide hazard map, which has been divided into five classes: very low (0–0.33), low (0.34–0.44), moderate (0.45–0.58), high (0.59–0.72), and very high (0.73–1).

The results show that the central-eastern parts of the area are more highly hazardous than the 4 km radius of central Phuentsholing, which comes under high hazard zone. The eastern and western parts of the region fall under very low to medium hazard zone. This could prove to be very helpful in landslide prevention mapping and could be used for land use planning and in inhibitory works by governmental agencies.

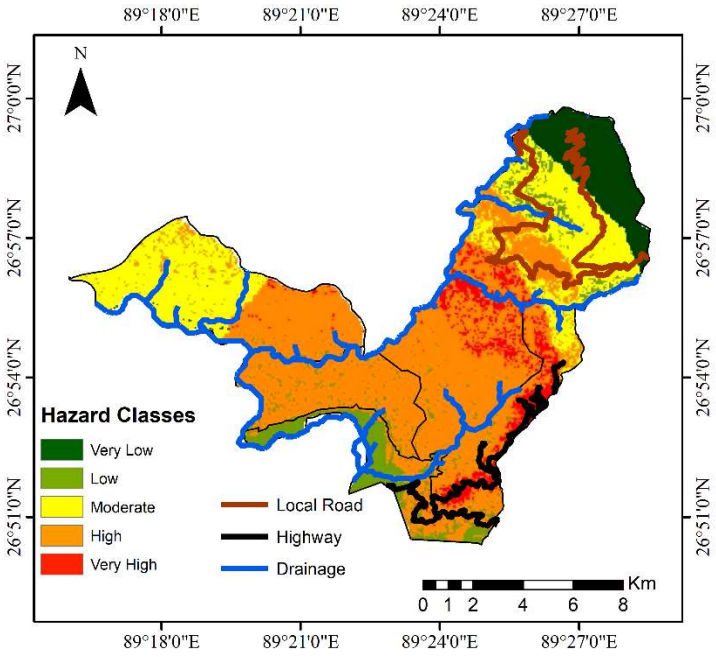


Figure 5. Landslide hazard map of the study region.

5.2. Landslide Vulnerability Assessment

The vulnerability assessment is defined as the expected degree of loss for elements at risk due to a certain event, with values ranging between 0 (no damage) and 1 (complete damage). Its assessment is based on estimation of vulnerability of life and property. A detailed risk assessment would involve considering several elements at risk, like population density, residential and commercial buildings, and their relationship with the type of landslides. However, due to the unavailability of such detailed data, the component of elements at risk could not be considered. The potential damage map was developed using four layers of elements related to exposure and vulnerability, which were LULC, population, proximity to road, and proximity to stream Figure 6a–d and a similar approach to susceptibility mapping was used. Apart from expert opinion, the weights (Equation (4)) provided indicate a logical progression. For example, the built-up area would have more loss due to a higher population, thereby providing them with a higher multiplier. Such regions are specifically used for either commercial or residential purposes or both. These areas can be dangerous on the hills as most of the houses are built on slopes and the roads are constructed by hill cutting. Agriculture covers a significant portion of the study area as it is a means of livelihood for the people staying there. Miscellaneous region includes water bodies and mining sites, and any effect on such sites due to landslides can lead to economic and ecological impact. Similarly, the road network is a lifeline to the people and to the economy of the country, therefore a higher multiplier was provided. This makes the proximity to road factor critical as road cuts are quite common in this region and the risk decreases as distance increases away from the road network. The relevance to the proximity to stream factor is due to the presence of stream channels which could suffer from erosion and hill cutting especially during heavy monsoons. The cutting of hills increases slope instability, making the area more vulnerable to landslides, and it usually increases in regions of higher drainage density and higher elevation. The total population in the study area is 33,444 with 18,043 males and 15,401 females as per census in Bhutan, 2017 [49].

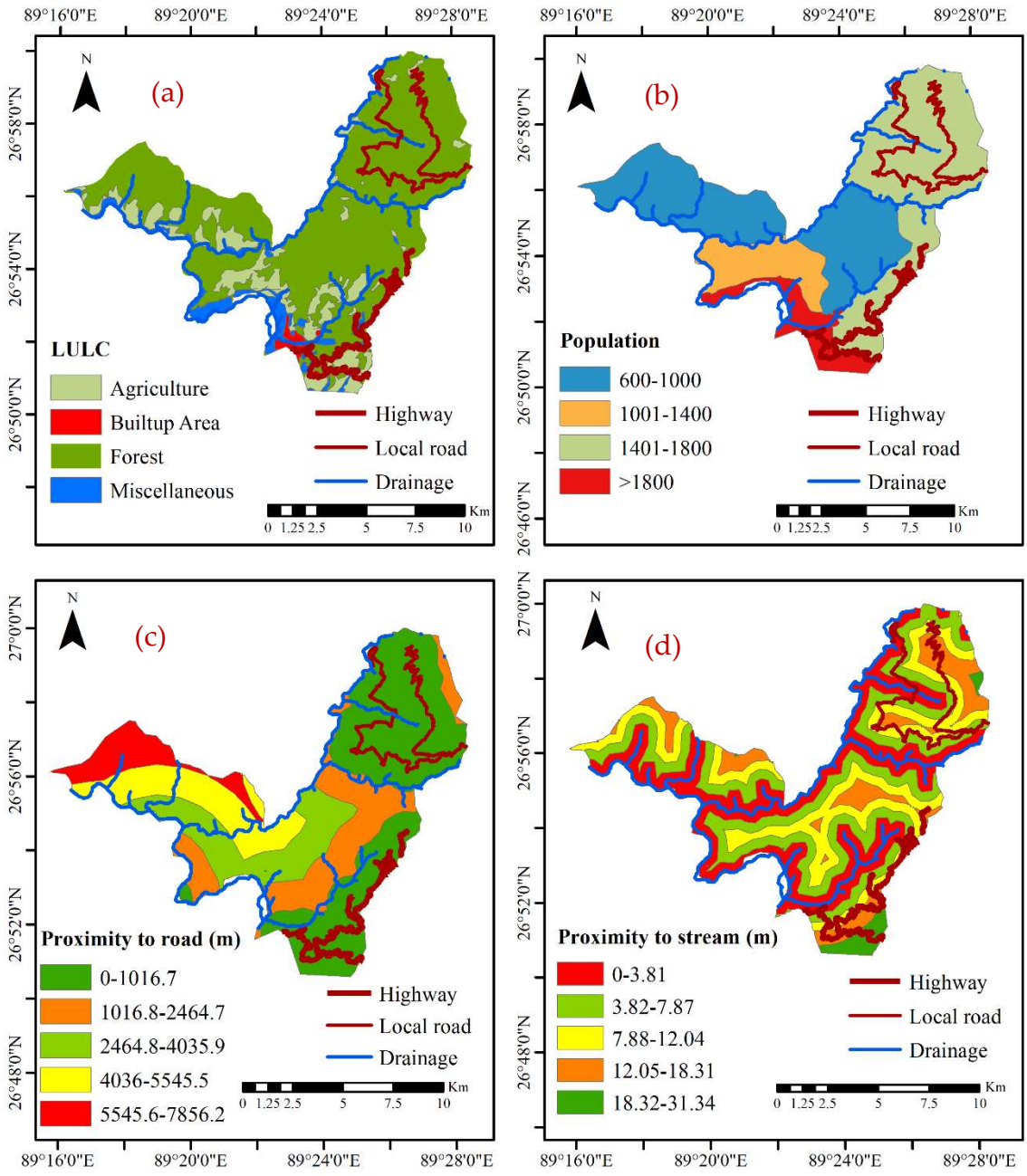


Figure 6. Landslide vulnerability factor maps (**a**) LULC, (**b**) population, (**c**) proximity to road, (**d**) proximity to stream.

The weights to the layers were provided based on Equation (4).

$$LV = \text{Population} \times 0.3 + \text{LULC} \times 0.2 + \text{Proximity to road} \times 0.3 + \text{Proximity to stream} \times 0.2 \quad (4)$$

where LV is landslide vulnerability map. The vulnerability map was developed and classified into five classes based on the natural breaks classification technique. Figure 7 represents the landslide vulnerability map and has been divided into five categories, namely, very low (0–0.36), low (0.37–0.48), moderate (0.49–0.64), high (0.65–0.76), and very high (0.77–1). The map depicts that very high and high vulnerable regions can be found in the NE and SW part of the location covering the populated areas. The developed vulnerability map can serve as a guideline for local disaster management agencies for development activities.

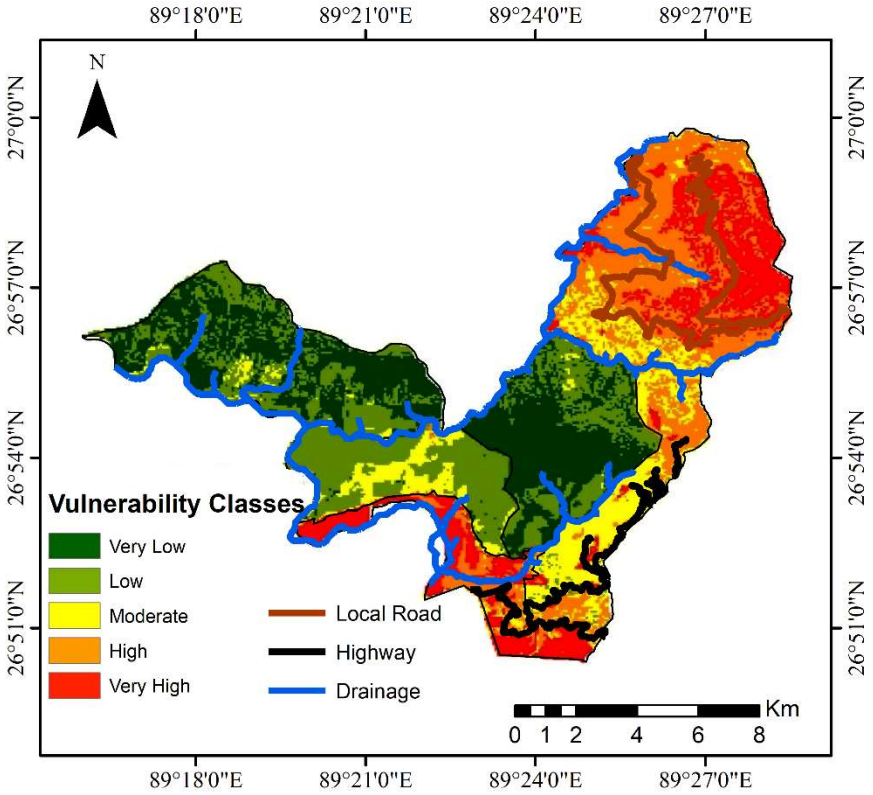


Figure 7. Landslide vulnerability map for Phuentsholing.

5.3. Landslide Risk Assessment

The type of damage linked to landslide risk in the present study has been confined to loss of life and/or property along with a road network. The risk map was developed by multiplying the landslide hazard and vulnerability map. Such a study would also help in estimating the vulnerable area and the impact from any landslide event in the foreseeable future. In order to understand the vulnerability to human lives, the risk to population was estimated and the high-risk zones were identified (Figure 8). The obtained map was classified into five classes (very low (0–0.12), low (0.13–0.21), moderate (0.22–0.37), high (0.38–0.55), and very high (0.56–1)) based on the natural break classification technique [47]. Table 2 depicts the area and affected population under various risk zones whereas Table 3 determines the various landslide risk zones for all six subdivisions (chiwogs). It can be inferred that 20.2% of the population with an area of 15.13 km^2 lives in high to high-risk landslide zones.

Table 2. Area and population distribution with respect to risk zones.

Risk Zone	Area (%)	Area (km^2)	Population
Very Low	44.70	58.7	19,148
Low	27.75	36.7	2030
Moderate	15.85	21.2	5496
High	8.50	11.3	6341
Very High	3.20	3.8	429

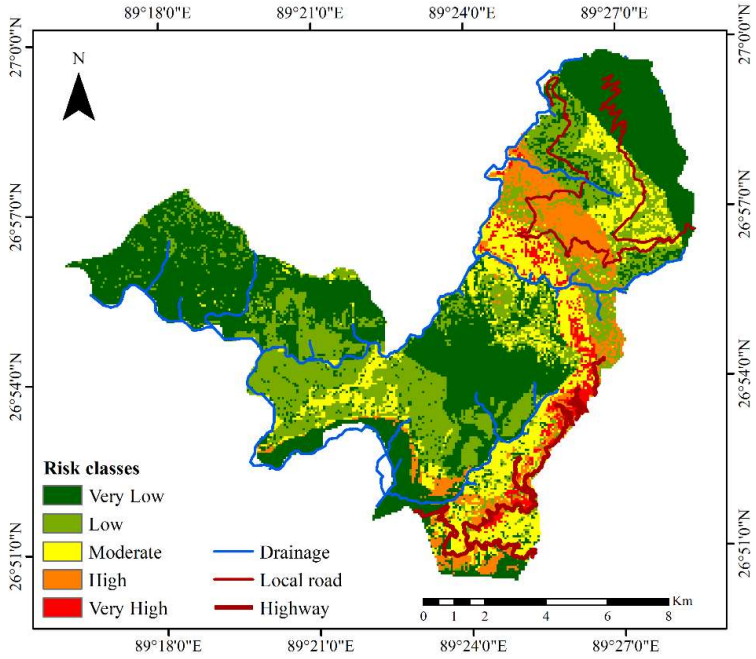


Figure 8. Landslide risk map of the study region.

Table 3. Area affected by various risk zones for all six chiwogs.

Chiwogs	Risk Zone	Area (km^2)
Lingdaen	Very Low	21.71
	Low	5.60
	Moderate	0.20
	High	0.00
	Very High	0.00
Pachhu	Very Low	1.30
	Low	11.40
	Moderate	2.80
	High	0.07
	Very High	0.00
Dzongkhag Thromde	Very Low	7.02
	Low	0.09
	Moderate	1.70
	High	2.35
	Very High	0.03
Dophugchen Wangduegatshel	Very Low	0.50
	Low	2.71
	Moderate	6.90
	High	2.82
	Very High	2.71
Deling Marpji	Very Low	14.20
	Low	6.00
	Moderate	2.60
	High	0.14
	Very High	0.01
Chong Geykha Dophulakha	Very Low	13.76
	Low	11.10
	Moderate	6.96
	High	5.95
	Very High	1.07

6. Conclusions

The occurrence of landslides and their growing incidence is a major threat to population and developments to the mountainous region, especially in developing countries. This has led to a need for comprehensive study on various aspects of landslides in such regions. However, such studies in the Bhutan Himalayas is very scarce and studies need to be conducted to understand the landslide effects in the region. Therefore, a landslide risk map has been developed for the Phuentsholing region of southwestern Bhutan. The study area has been selected since the major trade route, Phuentsholing-Thimphu Highway, which connects Bhutan with neighboring countries, passes through it and a landslide in this area can hamper people's lives and lead to economic losses. The method involved assessing the risk by multiplying the hazard and vulnerability maps. The landslide hazard map was developed by determining the spatial and temporal probabilities of landslides. The spatial probability was determined by performing a weighted sum analysis in an overlay environment for landslide causative factors (geology, slope, rainfall, vegetation, elevation), whereas temporal probability was determined using Poisson's probability model based on the already calculated regional thresholds. Similarly, landslide vulnerability was determined by providing weights based on expert knowledge gathered after carrying out extensive field study to various elements at risk involved. Finally, the landslide risk map was developed by combining the landslide hazard and vulnerability maps and was divided into five categories depending on various levels of risk associated. The analysis showed that 11.7% of the area comes under the high to very high category of risk, whereas 15.85% falls under moderate risk. One of the major problems for determining landslide risk in such a region is the unavailability of longer and precise rainfall and landslide records, with often under-reporting of the landslide damages. However, the present study can be considered as a preliminary step towards achieving a deep understanding of landslide risk at a local scale. The maps can be further modified and improved with the availability of more detailed data like the distribution of population in terms of gender, and the landslide type. The areas along the highway are expected to be at considerable risk in the near future as more hills are cut for residential and economic development purposes. Such an analysis may be used as a tool to develop a building code based on landslide risk, which Bhutan is presently lacking. This analysis would help the concerned authorities to understand the potential risk in certain sections and could be used for landslide mitigation purposes as there is a rise of infrastructure development in the region.

Author Contributions: Conceptualization, A.D. and R.S.; methodology, A.D. and S.A.; data curation, S.A.; writing—original draft preparation, A.D. and R.S.; writing—review and editing, B.P. and A.M.A.; supervision, B.P. and R.S. All authors have read and agreed to the published version of the manuscript.

Funding: The research was funded from BRACE project (NERC/GCRF NE/P016219/1) granted to Raju Sarkar. This research was supported by the Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS) in the University of Technology Sydney (UTS) under Grants 321740.2232335 and 321740.2232357; Grant 321740.2232424, and Grant 321740.2232452. This research is also supported by Researchers Supporting Project number RSP-2019/14, King Saud University, Riyadh, Saudi Arabia.

Acknowledgments: The authors are thankful to the National Center for Hydrology and Meteorology, Royal Government of Bhutan for providing rainfall data and the Border Roads Organization (Project DANTAK), Government of India for providing landslide data. The authors acknowledge the three anonymous reviewers and the editors for their useful comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Petley, D. Global patterns of loss of life from landslides. *Geology* **2012**, *40*, 927–930. [CrossRef]
2. Froude, M.J.; Petley, D.N. Global fatal landslide occurrence from 2004 to 2016. *Nat. Hazards Earth Syst. Sci.* **2018**, *18*, 2161–2181. [CrossRef]
3. Dikshit, A.; Sarkar, R.; Pradhan, B.; Jena, R.; Drukpa, D.; Alamri, A. Temporal Probability Assessment and Its Use in Landslide Susceptibility Mapping for Eastern Bhutan. *Water* **2020**, *12*, 267. [CrossRef]

4. Dikshit, A.; Satyam, N. Probabilistic rainfall thresholds in Chibo, India: Estimation and validation using monitoring system. *J. Mt. Sci.* **2019**, *16*, 870–883. [CrossRef]
5. Segoni, S.; Piciullo, L.; Gariano, S.L. A review of the recent literature on rainfall thresholds for landslide occurrence. *Landslides* **2018**, *15*, 1483–1501. [CrossRef]
6. Dikshit, A.; Sarkar, R.; Satyam, N. Probabilistic approach toward Darjeeling Himalayas landslides-A case study. *Cogent Eng.* **2018**, *5*, 1–11. [CrossRef]
7. Yin, Y.; Zheng, W.; Liu, Y.; Zhang, J.; Li, X. Integration of GPS with InSAR to monitoring of the Jiaju landslide in Sichuan, China. *Landslides* **2010**, *7*, 359–365. [CrossRef]
8. Bui, D.T.; Shahabi, H.; Omidvar, E.; Shirzadi, A.; Geertsema, M.; Clague, J.J.; Khosravi, K.; Pradhan, B.; Pham, B.T.; Chapi, K.; et al. Shallow Landslide Prediction Using a Novel Hybrid Functional Machine Learning Algorithm. *Remote. Sens.* **2019**, *11*, 931.
9. Vranken, L.; Vantilt, G.; Eeckhaut, M.V.D.; Vandekerckhove, L.; Poesen, J. Landslide risk assessment in a densely populated hilly area. *Landslides* **2014**, *12*, 787–798. [CrossRef]
10. Einstein, H.H. Landslide risk–Systematic approaches to assessment and management. *Landslide Risk Assess.* **2018**, 25–50. [CrossRef]
11. Althuwaynee, O.F.; Pradhan, B. Semi-quantitative landslide risk assessment using GIS-based exposure analysis in Kuala Lumpur City. *Geomat. Nat. Hazards Risk* **2016**, *8*, 706–732. [CrossRef]
12. Carrara, A.; Publiese-Carratelli, E.; Merenda, L. Computer-based data bank and statistical analysis of slope instability phenomena. *Z. fur Geomorphol. N.F.* **1977**, *21*, 187–222.
13. Nilsen, T.H.; Wright, R.; Vlasic, T.; Spangle, W. Relative slope stability and land-use planning in the San Francisco Bay region, California. *Prof. Pap.* **1979**, *944*. [CrossRef]
14. Lovell, D.P.; Thomas, G. Quantitative risk assessment. In Proceedings of the Food Chemical Risk Analysis; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 1997; pp. 57–86.
15. Wong, H. Landslide risk assessment for individual facilities. In *Landslide Risk Management*; CRC Press: Boca Raton, FL, USA, 2005; pp. 247–306.
16. Hadmoko, D.S.; Lavigne, F.; Sartohadi, J.; Hadi, P. Winaryo Landslide hazard and risk assessment and their application in risk management and landuse planning in eastern flank of Menoreh Mountains, Yogyakarta Province, Indonesia. *Nat. Hazards* **2010**, *54*, 623–642. [CrossRef]
17. Pradhan, B.; Abdulwahid, W.M. Landslide Risk Assessment Using Multi-hazard Scenario Produced by Logistic Regression and LiDAR-Based DEM. In *Laser Scanning Applications in Landslide Assessment*; Springer Science and Business Media LLC: Berlin/Heidelberg, German, 2017; Volime 9, pp. 253–275.
18. Liu, X.; Miao, C. Large-scale assessment of landslide hazard, vulnerability and risk in China. *Geomat. Nat. Hazards Risk* **2018**, *9*, 1037–1052. [CrossRef]
19. Goorchi, R.N.; Amini, M.; Memarian, H. A new rating system approach for risk analysis of rock slopes. *Nat. Hazards* **2017**, *91*, 75–102.
20. Saldivar-Sali, A.; Einstein, H.H. A Landslide Risk Rating System for Baguio, Philippines. *Eng. Geol.* **2007**, *91*, 85–99. [CrossRef]
21. Alejano, L.R.; Stockhausen, H.; Alonso, E.; Bastante, F.G.; Oyanguren, P.R. ROFRAQ: A statistics-based empirical method for assessing accident risk from rockfalls in quarries. *Int. J. Rock Mech. Min. Sci.* **2008**, *45*, 1252–1272. [CrossRef]
22. Lombardi, M.; Cardarilli, M.; Raspa, G. Spatial variability analysis of soil strength to slope stability assessment. *Géoméch. Eng.* **2017**, *12*, 483–503. [CrossRef]
23. Tofani, V.; Bicocchi, G.; Rossi, G.; Segoni, S.; D'Ambrosio, M.; Casagli, N.; Catani, F. Soil characterization for shallow landslides modeling: A case study in the Northern Apennines (Central Italy). *Landslides* **2017**, *14*, 755–770. [CrossRef]
24. Cardarilli, M.; Lombardi, M.; Corazza, A. Landslide risk management through spatial analysis and stochastic prediction for territorial resilience evaluation. *Int. J. Saf. Secur. Eng.* **2019**, *9*, 109–120. [CrossRef]
25. Gariano, S.L.; Sarkar, R.; Dikshit, A.; Dorji, K.; Brunetti, M.T.; Peruccacci, S.; Melillo, M. Automatic calculation of rainfall thresholds for landslide occurrence in Chukha Dzongkhag, Bhutan. *Bull. Int. Assoc. Eng. Geol.* **2018**, *78*, 4325–4332. [CrossRef]

26. Dikshit, A.; Sarkar, R.; Pradhan, B.; Acharya, S.; Dorji, K. Estimating Rainfall Thresholds for Landslide Occurrence in the Bhutan Himalayas. *Water* **2019**, *11*, 1616. [CrossRef]
27. Sarkar, R.; Dorji, K. Determination of the Probabilities of Landslide Events—A Case Study of Bhutan. *Hydrology* **2019**, *6*, 52. [CrossRef]
28. Erener, A.; Duzgun, H.S.B. A regional scale quantitative risk assessment for landslides: Case of Kumluca watershed in Bartin, Turkey. *Landslides* **2012**, *10*, 55–73. [CrossRef]
29. Luca, F.; D'Ambrosio, D.; Robustelli, G.; Rongo, R.; Spataro, W. Integrating geomorphology, statistic and numerical simulations for landslide invasion hazard scenarios mapping: An example in the Sorrento Peninsula (Italy). *Comput. Geosci.* **2014**, *67*, 163–172. [CrossRef]
30. Piciullo, L.; Calvello, M.; Cepeda, J.M. Territorial early warning systems for rainfall-induced landslides. *Earth-Sci. Rev.* **2018**, *179*, 228–247. [CrossRef]
31. Gansser, A. Geology of the Bhutan Himalaya. In *Denkschriften der Schweizerischen Naturforschenden Geselschaft*; Birkhäuser Verlag: Basel, Switzerland, 1983; p. 181.
32. Segoni, S.; Pappafico, G.; Luti, T.; Catani, F. Landslide susceptibility assessment in complex geological settings: Sensitivity to geological information and insights on its parameterization. *Landslides* **2020**, 1–11. [CrossRef]
33. Long, S.; McQuarrie, N.; Tobgay, T.; Grujic, D. Geometry and crustal shortening of the Himalayan fold-thrust belt, eastern and central Bhutan. *GSA Bull.* **2011**, *123*, 1427–1447. [CrossRef]
34. Lohmann, U.; Sausen, R.; Bengtsson, L.; Cubasch, U.; Perlwitz, J.; Roeckner, E. The Köppen climate classification as a diagnostic tool for general circulation models. *Clim. Res.* **1993**, *3*, 177–193. [CrossRef]
35. Prokop, P.; Walanus, A. Impact of the Darjeeling–Bhutan Himalayan front on rainfall hazard pattern. *Nat. Hazards* **2017**, *89*, 387–404. [CrossRef]
36. Hervás, J.; Bobrowsky, P. Mapping: Inventories, Susceptibility, Hazard and Risk. In *Landslides–Disaster Risk Reduction*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2008; pp. 321–349.
37. Reichenbach, P.; Rossi, M.; Malamud, B.; Mihir, M.; Guzzetti, F. A review of statistically-based landslide susceptibility models. *Earth-Sci. Rev.* **2018**, *180*, 60–91. [CrossRef]
38. Michael, E.A.; Samanta, S. Landslide vulnerability mapping (LVM) using weighted linear combination (WLC) model through remote sensing and GIS techniques. *Model. Earth Syst. Environ.* **2016**, *2*, 88. [CrossRef]
39. Jaiswal, P.; Van Westen, C. Estimating temporal probability for landslide initiation along transportation routes based on rainfall thresholds. *Geomorphology* **2009**, *112*, 96–105. [CrossRef]
40. Guzzetti, F.; Reichenbach, P.; Cardinali, M.; Galli, M.; Ardizzone, F. Probabilistic landslide hazard assessment at the basin scale. *Geomorphology* **2005**, *72*, 272–299. [CrossRef]
41. Lin, Q.; Wang, Y.; Liu, T.; Zhu, Y.; Sui, Q. The Vulnerability of People to Landslides: A Case Study on the Relationship between the Casualties and Volume of Landslides in China. *Int. J. Environ. Res. Public Heal.* **2017**, *14*, 212. [CrossRef]
42. Baiocchi, V.; Dominici, D.; Guarascio, M.; Lombardi, M.; Vatore, F. Mapping seismic vulnerability in buildings by means of open source tools and open data. *Geoing. Ambient. e Min.* **2017**, *151*, 27–32.
43. Glade, T.; Crozier, M.J. *The Nature of Landslide Hazard Impact*; Landslide Hazard and Risk; Wiley: Chichester, UK, 2005; pp. 43–74.
44. Van Westen, C.; Alkema, D.; Damen, M.; Kerle, N.; Kingma, N. *Multi-Hazard Risk Assessment: Distance Education Course Guide Book*; United Nations University—ITC School on Disaster Geoinformation Management: Tokyo, Japan, 2009.
45. Jena, R.; Pradhan, B.; Beydoun, G. Earthquake vulnerability assessment in Northern Sumatra province by using a multi-criteria decision-making model. *Int. J. Disaster Risk Reduct.* **2020**, *46*, 101518. [CrossRef]
46. Pourghasemi, H.R.; Mohammady, M.; Pradhan, B. Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran. *Catena* **2012**, *97*, 71–84. [CrossRef]
47. Park, S.-J.; Lee, C.-W.; Lee, S.; Lee, M.-J. Landslide Susceptibility Mapping and Comparison Using Decision Tree Models: A Case Study of Jumunjin Area, Korea. *Remote. Sens.* **2018**, *10*, 1545. [CrossRef]

48. Kayal, J.R.; Arefiev, S.S.; Baruah, S.; Tatevossian, R.; Gogoi, N.; Sanoujam, M.; Gautam, J.L.; Hazarika, D.; Borah, D. The 2009 Bhutan and Assam felt earthquakes (Mw6.3 and 5.1) at the Kopili fault in the northeast Himalaya region. *Geomat. Nat. Hazards Risk* **2010**, *1*, 273–281. [CrossRef]
49. Bhutan, N. *Population and Housing Census of Bhutan*; Royal Government of Bhutan: Thimpu, Bhutan, 2017.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

arXiv:2007.07816v1 [math.CV] 15 Jul 2020

SUBORDINATION AND RADIUS PROBLEMS FOR CERTAIN STARLIKE FUNCTIONS

S. SIVAPRASAD KUMAR AND KAMALJEET GANGANIA

Abstract. We study the following class of starlike functions

$$\mathcal{S}^*_\wp := \{f \in \mathcal{A} : zf'(z)/f(z) \prec 1 + ze^z =: \wp(z)\},$$

that are associated with the cardioid domain $\wp(\mathbb{D})$, by deriving certain convolution results, radius problems, majorization result, radius problems in terms of coefficients and differential subordination implications. Consequently, we establish some interesting generalizations of our results for the Ma-Minda class of starlike functions $\mathcal{S}^*(\psi)$. We also provide, the set of extremal functions maximizing $\Re\Phi\left(\log\left(f(z)/z\right)\right)$ or $|\Phi\left(\log\left(f(z)/z\right)\right)|$ for functions in $\mathcal{S}^*(\psi)$, where Φ is a non-constant entire function. Further T. H. MacGregor's result for the class $\mathcal{S}^*(\alpha)$ and $\mathcal{S}^*_\wp$ are obtained as special case to our result.

2010 AMS Subject Classification. Primary 30C45, Secondary 30C50, 30C80.
Keywords and Phrases. Starlike functions; Convolution; Radius problem; Extremal problem; Subordination; Majorization; Special functions.

1. Introduction

Let $\mathcal{A}_0$ be the collection of analytic functions of the form $p(z) = 1+p_1z+p_2z+\cdots$ and $\mathcal{A}$ consists of analytic functions, f normalized by the conditions $f(0) = 0$ and $f'(0) = 1$ defined in the unit disk $\mathbb{D} := \mathbb{D}_1$, where $\mathbb{D}_r := \{z : |z| < r\}$. The Carathéodory class, $\mathcal{P}$ consists of functions $p \in \mathcal{A}_0$ with $\Re p(z) > 0$. Let us denote the class of normalized univalent functions by $\mathcal{S}$. For two analytic functions f and g, we say f subordinate to g, written as $f \prec g$, if there exists a Schwarz function w such that $f(z) = g(w(z))$. If g is univalent, then $f \prec g$ if and only if $f(\mathbb{D}) \subseteq g(\mathbb{D})$ and $f(0) = g(0)$. Many subclasses of $\mathcal{S}$ were introduced and studied by several authors, the chief amongst them is $\mathcal{S}^*$ (or $\mathcal{C}$), the class of normalized starlike (or convex) functions which is characterized by the quantity $zf'(z)/f(z)$ (or $1 + zf''(z)/f'(z)$), assuming values in the right half plane. In 1992, using subordination, Ma and Minda [19] defined the classes (1.1) and (1.2), which unify many subclasses of $\mathcal{S}$ and they also studied some of it's common properties.

$$\mathcal{S}^*(\psi) := \left\{f \in \mathcal{A} : \frac{zf'(z)}{f(z)} \prec \psi(z)\right\} \tag{1.1}$$

and

$$\mathcal{C}(\psi) := \left\{f \in \mathcal{A} : 1 + \frac{zf''(z)}{f'(z)} \prec \psi(z)\right\}, \tag{1.2}$$

where $\psi \in \mathcal{P}$ is univalent and $\psi(\mathbb{D})$ starlike with respect to $\psi(0) = 1$. Note that $\mathcal{S}^*(\frac{1+z}{1-z})$ reduces to the class $\mathcal{S}^*$. Now a good amount of literature exists for different

The work of the second author is supported by University Grant Commission, New-Delhi, India under UGC-Ref. No.:1051/(CSIR-UGC NET JUNE 2017).

choices of ψ in (1.1). For example, the class $\mathcal{S}(\alpha,\beta) := \mathcal{S}^*(p_{\alpha,\beta}(z))$, where

$$p_{\alpha,\beta}(z) := 1 + \frac{\beta-\alpha}{\pi} i \log \frac{1-e^{2\pi i \frac{1-\alpha}{\beta-\alpha}} z}{1-z}, \tag{1.3}$$

$\alpha < 1$, $\beta > 1$ and $p_{\alpha,\beta}$ maps $\mathbb{D}$ onto the convex domain $\{w \in \mathbb{C} : \alpha < \Re w < \beta\}$, was introduced by Kuroki and Owa [18]. Sokół and Stankiewicz [27] introduced the class associated with lemniscate of Bernoulli $\mathcal{SL}^* := \mathcal{S}^*(\sqrt{1+z})$. The class $\mathcal{S}_e^* := \mathcal{S}^*(e^z)$ of exponential starlike functions was introduced by Mendiratta et al. [22] and the class $\mathcal{S}_b^* := \mathcal{S}^*(e^{e^z-1})$ of starlike functions related with Bell numbers was introduced by Kumar et at. [15]. For some recent classes see [24, 17, 26, 10, 12]. Evidently the functions ψ under discussion above are all convex and hence satisfy the following result:

Theorem 1.1. [19] *Let $\psi(\mathbb{D})$ be convex, $g \in \mathcal{C}$ and $f \in \mathcal{S}^*(\psi)$. Then $f * g \in \mathcal{S}^*(\psi)$.*

But the above result is not applicable when $\psi(z) = 1 + ze^z$, $z + \sqrt{1+z^2}$, e^{e^z-1} and $1 + 4z/3 + 2z^2/3$ etc. due to the fact that these are not convex. We deal with Theorem 1.1 for the case when ψ starlike and derive some radius related problems.

In Geometric function theory, we usually find the sufficient conditions in terms of it's coefficients for the normalized functions f in $\mathcal{A}$ to be in a desired class. In recent times, finding the sufficient conditions on functions to be in a desired class using the technique of differential subordination, admissibility conditions and subordination chain has attracted many researcher's attention. See [4, 12, 23].

In the present work, we consider the following first-order differential subordination implications:

$$1 + \beta \frac{zp'(z)}{p^n(z)} \prec h(z) \Rightarrow p(z) \prec q(z), \tag{1.4}$$

where $n \in \{1,2,3\}$. We find the condition on β so that (1.4) holds whenever the functions h and q are already known. Note that if p and q are two fixed analytic and univalent functions in $\mathbb{D}$ and q does not have the nice properties like, explicit inverse representation and convexity, then proving $p \prec q$ is not so easy. See that if $p \prec q$, then geometrically the boundary curve of $p(\mathbb{D})$ is contained in $q(\mathbb{D})$, which can be ensured by the technique of radial distances. Further, if the images $p(\mathbb{D})$ and $q(\mathbb{D})$ are symmetric about the real axis and both the functions p and q have the same orientation, then we can use the maximum arguments of $p(z)$ and $q(z)$ for $|z| = 1$ to prove that $p \prec q$.

Recall that for the subfamilies $\mathcal{G}_1$ and $\mathcal{G}_2$ of $\mathcal{A}$, we say that r_0 is the $\mathcal{G}_1$-radius of the class $\mathcal{G}_2$, if $r_0 \in (0,1)$ is largest number such that $r^{-1}f(rz) \in \mathcal{G}_1$, $0 < r \leq r_0$ for all $f \in \mathcal{G}_2$. Recently, the radii of starlikeness and convexity of some normalized special functions were studied as they can be represented as Hadamard factorization under certain conditions, a few such special functions are Bessel functions [1], Struve functions [5, 1], Wright functions [6], Lommel functions [5, 1] and Legendre polynomials of odd degree [8]. If f and g be analytic in $\mathbb{D}$, then $g(z)$ is said to be majorized by $f(z)$, if there exists an analytic function $\Phi(z)$ in $\mathbb{D}$ satisfying $|\Phi(z)| \leq 1$ and $g(z) = \Phi(z)f(z)$ for all $z \in \mathbb{D}$. MacGregor [20] found the largest radius r_0 so that $|g'(z)| \leq |f'(z)|$ holds in $|z| \leq r_0$ and he also proved that $r_0 = 2 - \sqrt{3}$, if f is univalent. Recently, Tang and Deng [28] obtained this majorization-radius when f belongs to some class $\mathcal{S}^*(\psi)$.

Consider the class of cardioid starlike functions:

$$\mathcal{S}^*_{\wp} := \left\{ f \in \mathcal{A} : \frac{zf'(z)}{f(z)} \prec \wp(z) := 1 + ze^z \right\}. \tag{1.5}$$

Let $\mathbb{E}_{\alpha,\beta}$ be the normalized form of the Mittag-Leffler function:

$$\mathbb{E}_{\alpha,\beta}(z) = z + \sum_{n\geq 2} \frac{\Gamma(\beta)}{\Gamma(\alpha(n-1)+\beta)} z^n, \quad (z, \alpha, \beta \in \mathbb{C}; \Re(\alpha) > 0, \beta \neq 0, -1, \cdots).$$

Then clearly $\wp(z) = 1 + \mathbb{E}_{1,1}(z)$. In [16], $\mathcal{S}^*_{\wp}$ was extensively studied with inclusion properties, radius problems, sharp coefficient's estimations including hankel determinants. Where in, it is proved that $r_c := (3-\sqrt{5})/2$ is the radius of convexity for $\wp$ however in [3], $\mathbb{E}_{\alpha,\beta}$ is shown to be convex for $\alpha \geq 1$ and $\beta \geq (3+\sqrt{17})/2$. So, $\mathbb{E}_{1,1}$ has the radius of convexity r_c.

In this paper, we continue to study further in finding the largest radius for some normalized special functions to be in $\mathcal{S}^*_{\wp}$. We also obtain largest radius r such that $f \in \mathcal{S}^*_{\wp}$ in $|z| < r$, whenever f belongs to the class $\mathcal{S}(\alpha, \beta)$ and the class $\mathcal{S}_\lambda$ defined in (1.3) and (2.5), respectively. We find that special function's radii of starlikness results obtained here naturally generalize to $\mathcal{S}^*(\psi)$-radii. Further, radius problem related to the majorization and some sufficient conditions in terms of coefficients for the normalized functions to be in the class $\mathcal{S}^*_{\wp}$ are established. We discuss the modified form of Theorem 1.1 and as an application, we find the largest radius so that Theorem 1.1 holds for $\mathcal{S}^*_{\wp}$ and the convolution of two starlike functions to be in $\mathcal{S}^*_{\wp}$. In the last section, we obtain the condition on β so that the differential subordination implication (1.4) holds, whenever $h(z) = \wp(z)$ or $q(z) = \wp(z)$.

2. Radius problems

We need the following lemma to prove the subsequent results:

Lemma 2.1. [16] *Let $\wp(z) = 1 + ze^z$. Then we have $\{w : |w-a| < R_a\} \subset \wp(\mathbb{D})$, where*

$$R_a = \begin{cases} (a-1) + 1/e, & 1 - 1/e < a \leq 1 + (e - e^{-1})/2; \\ e - (a-1), & 1 + (e-e^{-1})/2 \leq a < 1 + e. \end{cases}$$

Theorem 2.1. *Let $f \in \mathcal{S}(\alpha, \beta)$. Then $f \in \mathcal{S}^*_{\wp}$ in $\mathbb{D}_{r_0}$, where r_0 is the least positive root of the equation*

$$\frac{\beta - \alpha}{\pi} \left(\log \frac{1 + \sqrt{2(1 + \cos(2\pi \frac{1-\alpha}{\beta-\alpha}))} r + r^2}{1 - r^2} + 2 \arctan \frac{r}{1-r} \right) - \frac{1}{e} = 0. \tag{2.1}$$

Proof. Consider the analytic function

$$p_{\alpha,\beta}(z) := 1 + \frac{\beta - \alpha}{\pi} i \log q(z),$$

where

$$q(z) = \frac{1 - cz}{1 - z} \quad \text{and} \quad c = \exp\left(2\pi i \frac{1-\alpha}{\beta - \alpha} \right).$$

Note that $q(z)$ is a bilinear transformation, maps $\mathbb{D}_r$ onto the disk:

$$\left| q(z) - \frac{1 + cr^2}{1 - r^2} \right| \leq \frac{|1 + c| r}{1 - r^2},$$

which implies

$$|q(z)| \leq \frac{1+|1+c|r+r^2}{1-r^2}$$

and therefore

$$\log|q(z)| \leq \log\left(\frac{1+|1+c|r+r^2}{1-r^2}\right). \tag{2.2}$$

For any $\delta \in \mathbb{C}$ with $|\delta| = 1$, we have $1+\delta z \prec 1+z$. So to maximize $|\arg(1+\delta z)|$, it suffices to consider $|\arg(1+z)|$. Now for $|z| = r$, we have

$$|\arg(1+z)| \leq \arctan\frac{r}{1-r}. \tag{2.3}$$

Hence to apply Lemma 2.1, we need to maximize $|p_{\alpha,\beta}(z)-1|$, that is,

$$|p_{\alpha,\beta}-1| = \frac{\beta-\alpha}{\pi}\left|\log|q(z)| + i\arg\frac{1-cz}{1-z}\right|. \tag{2.4}$$

Using (2.2) and (2.3) in (2.4), we see that

$$|p_{\alpha,\beta}-1| \leq \frac{\beta-\alpha}{\pi}\left(\log\frac{1+|1+c|r+r^2}{1-r^2} + 2\arctan\frac{r}{1-r}\right) \leq \frac{1}{e}$$

holds in $|z| < r_0$ whenever r_0 is the smallest positive root of the equation (2.1). This ends the proof. □

Note that if we choose $\alpha = 1+\frac{\delta-\pi}{2\sin\delta}$ and $\beta = 1+\frac{\delta}{2\sin\delta}$, where $\pi/2 \leq \delta < \pi$, then the class $\mathcal{S}(\alpha,\beta)$ reduces to the class $\mathcal{V}(\delta)$ introduced by Kargar et al. [14].

Corollary 2.2. *Let $f \in \mathcal{V}(\delta)$. Then $f \in \mathcal{S}^*_\wp$ in $\mathbb{D}_{r_\delta}$, where r_δ is the least positive root of the equation*

$$\frac{1}{2\sin\delta}\left(\log\frac{1+\sqrt{2(1+\cos(2(\pi-\delta)))}r+r^2}{1-r^2} + 2\arctan\frac{r}{1-r}\right) - \frac{1}{e} = 0.$$

Now we consider the following class introduced in [9]:

$$\mathcal{S}_\lambda := \left\{f \in \mathcal{A} : \frac{f(z)}{z} \in P_\lambda\right\}, \tag{2.5}$$

where

$$P_\lambda := \{p \in \mathcal{A}_0 : \Re(e^{i\lambda}p(z)) > 0, \quad -\pi/2 \leq \lambda \leq \pi/2\}$$

denotes the class of tilted Carathéodory functions [29]. Note that P_0 reduces to $\mathcal{P}$, the class of Carathéodory functions. For the function $p \in P_\lambda$, upper bound on the quantity $zp'(z)/p(z)$ is given by the following lemma that will be used for our next result:

Lemma 2.2. [29] *If $p \in P_\lambda$, then $|zp'(z)/p(z)| \leq M(\lambda,r)$, where*

$$M(\lambda,r) = \begin{cases} \frac{2r\cos\lambda}{r^2-2r|\sin\lambda|+1} & \text{for} \quad r < |\tan\frac{\lambda}{2}|; \\ \frac{2r}{1-r^2} & \text{for} \quad r \geq |\tan\frac{\lambda}{2}|. \end{cases}$$

The equality holds for some point $z = re^{i\theta}$, $r \in (0,1)$ if and only if $p(z) = p_\lambda(yz)$, where $p_\lambda(z) = \frac{1+e^{-2i\lambda}z}{1-z}$ and $y = e^{i(\theta_0-\theta)}$ with

$$\theta_0 = \begin{cases} \frac{\pi}{2}+\lambda & \text{for} \quad r < -\tan\frac{\lambda}{2}; \\ -\frac{\pi}{2}+\lambda & \text{for} \quad r < \tan\frac{\lambda}{2}; \\ \arcsin\left(\frac{1+r^2}{r^2-1}\right)+\lambda & \text{for} \quad r \geq |\tan\frac{\lambda}{2}|. \end{cases}$$

We determine the largest radius r such that the function $F(z) := f(z)g(z)/z \in \mathcal{S}^*_\wp$ in $|z| < r$, whenever $f, g \in \mathcal{S}_\lambda$.

Theorem 2.3. *Let $c_\lambda = \cos\lambda$, $s_\lambda = \sin\lambda$ and $t_\lambda = |\tan(\lambda/2)|$. If $f, g \in \mathcal{S}_\lambda$, then $F \in \mathcal{S}^*_\wp$ in $\mathbb{D}_{r_0}$, where*

$$r_0 := \begin{cases} 2ec_\lambda + |s_\lambda| + \sqrt{((4e^2-1)c_\lambda + 4e|s_\lambda|)c_\lambda}, & \text{if } \ r < t_\lambda; \\ \sqrt{4e^2+1} - 2e, & \text{if } \ r \geq t_\lambda. \end{cases}$$

Proof. Since $f, g \in \mathcal{S}_\lambda$, it follows that the functions $p(z) = f(z)/z$ and $q(z) = g(z)/z$ belong to the class P_λ such that $F(z) = zp(z)q(z)$. Thus

$$\frac{zF'(z)}{F(z)} - 1 = \frac{zp'(z)}{p(z)} + \frac{zq'(z)}{q(z)}.$$

Now from Lemma 2.2, we obtain

$$\left|\frac{zF'(z)}{F(z)} - 1\right| \leq 2M(\lambda, r).$$

Therefore, using Lemma 2.1, we conclude that if $2M(\lambda, r) \leq 1/e$, then $F \in \mathcal{S}^*_\wp$. Since $2M(\lambda, r) \leq 1/e$ holds whenever

$$\frac{2rc_\lambda}{r^2 - 2|s_\lambda|r + 1} \leq \frac{1}{2e}, \quad \text{if} \quad r < t_\lambda$$

and

$$\frac{2r}{1-r^2} \leq \frac{1}{2e}, \quad \text{if} \quad r \geq t_\lambda;$$

or equivalently

$$r^2 - 2(|s_\lambda| + 2ec_\lambda)r + 1 \geq 0, \quad \text{if} \quad r < t_\lambda$$

and

$$r^2 + 4er - 1 \leq 0, \quad \text{if} \quad r \geq t_\lambda,$$

respectively. Hence the result follows with r_0 as given in the hypothesis. Further, for the functions

$$f(z) = g(z) = \frac{z(1 + e^{-2i\lambda}yz)}{1 - yz},$$

sharpness hold in view of Lemma 2.2. □

2.1. **Cardioid starlikeness radius for some special functions and generalizations.** We begin with the following: The Bessel function $\mathcal{J}_\beta$ of first kind of order $\beta \in \mathbb{C}$ is a particular solution of the homogeneous Bessel differential equation

$$z^2w''(z) + zw'(z) + (z^2 - \beta^2)w(z) = 0$$

and have the following series expansion:

$$\mathcal{J}_\beta(z) := \sum_{n\geq 1} \frac{(-1)^n}{n!\Gamma(n+\beta+1)} \left(\frac{z}{2}\right)^{2n+\beta},$$

where $z \in \mathbb{C}$ and $\beta \notin \mathbb{Z}^-$. The following three normalized functions expressed in terms of $\mathcal{J}_\beta(z)$ have been studied extensively by many authors:

$$f_\beta(z) = (2^\beta\Gamma(\beta+1)\mathcal{J}_\beta(z))^{1/\beta} = z - \frac{1}{4\beta(\beta+1)}z^3 + \cdots, \quad \beta \neq 0,$$

$$g_\beta(z) = 2^\beta\Gamma(\beta+1)z^{1-\beta}\mathcal{J}_\beta(z) = z - \frac{1}{4(\beta+1)}z^3 + \cdots,$$

$$h_\beta(z) = 2^\beta \Gamma(\beta+1) z^{1-\beta/2} \mathcal{J}_\beta(\sqrt{z}) = z - \frac{1}{4(\beta+1)} z^2 + \cdots . \tag{2.6}$$

Since the zeros of $\mathcal{J}_\beta$ are real if $\beta > 0$, therefore using Weierstrass decomposition, we have for $\beta > 0$:

$$\mathcal{J}_\beta(z) := \frac{z^\beta}{2^\beta \Gamma(\beta+1)} \prod_{n\geq 1} \left(1 - \frac{z^2}{j_{\beta,n}^2}\right),$$

where $j_{\beta,n}$ is the n-th positive zero of $\mathcal{J}_\beta$ and satisfies $j_{\beta,n} < j_{\beta,n+1}$ for $n \in \mathbb{N}$. Thus we have

$$\frac{z\mathcal{J}_\beta'(z)}{\mathcal{J}_\beta(z)} = \beta - \sum_{n\geq 1} \frac{2z^2}{j_{\beta,n}^2 - z^2}. \tag{2.7}$$

Now using the above representation, we obtain the radii of cardioid-starlikeness of the functions f_β, g_β and h_β for the case $\beta > 0$.

Theorem 2.4. *Let $\beta > 0$. Then the radii of cardioid-starlikeness $r_\wp(f_\beta)$, $r_\wp(g_\beta)$ and $r_\wp(h_\beta)$ of the functions f_β, g_β and h_β are the smallest positive roots of the following equations, respectively:*

(i) $er\mathcal{J}_\beta'(r) + \beta(e-1)\mathcal{J}_\beta(r) = 0$;
(ii) $er\mathcal{J}_\beta'(r) - (e\beta - 1)\mathcal{J}_\beta(r) = 0$;
(iii) $e\sqrt{r}\mathcal{J}_\beta'(\sqrt{r}) - (e\beta - 2)\mathcal{J}_\beta(\sqrt{r}) = 0$.

Proof. We now prove the first part. Using the representation (2.7) and equation (2.6), we get

$$\frac{zf_\beta'(z)}{f_\beta(z)} = \frac{z\mathcal{J}_\beta'(z)}{\beta\mathcal{J}_\beta(z)} = 1 - \frac{1}{\beta}\sum_{n\geq 1} \frac{2z^2}{j_{\beta,n}^2 - z^2}. \tag{2.8}$$

Further using a result [11, Lemma 3.2, p. 10] and from (2.8), $|z| = r < j_{\beta,1}$ we obtain

$$\left|\frac{zf_\beta'(z)}{f_\beta(z)} - a\right| \leq \frac{2}{\beta}\sum_{n\geq 1} \frac{j_{\beta,n}^2 r^2}{j_{\beta,n}^4 - r^4}, \tag{2.9}$$

where $a := 1 - \frac{2}{\beta}\sum_{n\geq 1} \frac{r^4}{j_{\beta,n}^4 - r^4}$ and $j_{\beta,n}$ denotes the n-th positive zero of the Bessel function $\mathcal{J}_\beta$. Also a simple calculation shows that $a \leq 1$. Thus for the disk (2.9) to lie inside $\wp(\mathbb{D})$, we need only to consider that $1 - \frac{1}{e} < a < 1 + \frac{e-e^{-1}}{2}$, and so by Lemma 2.1, we have

$$\frac{2}{\beta}\sum_{n\geq 1} \frac{j_{\beta,n}^2 r^2}{j_{\beta,n}^4 - r^4} \leq a - 1 + \frac{1}{e} = \frac{1}{e} - \frac{2}{\beta}\sum_{n\geq 1} \frac{r^4}{j_{\beta,n}^4 - r^4},$$

or equivalently,

$$\frac{2}{\beta}\sum_{n\geq 1} \frac{r^2}{j_{\beta,n}^2 - r^2} - \frac{1}{e} \leq 0. \tag{2.10}$$

Also using (2.8), (2.10) can be written as $\frac{er\mathcal{J}_\beta'(r)}{\beta\mathcal{J}_\beta(r)} + 1 - e \geq 0$. Note that in view of Lemma 2.1, we can also obtain (2.10) directly from (2.8). Now let us consider the strictly decreasing continuous function

$$\Psi(r) := \frac{1}{e} - \frac{2}{\beta}\sum_{n\geq 1} \frac{r^2}{j_{\beta,n}^2 - r^2}, \quad r \in (0, j_{\beta,1}).$$

Then $\lim_{r\to 0}\Psi(r) = 1/e > 0$ and $\lim_{r\to j_{\beta,1}}\Psi(r) = -\infty$. Also $\Psi'(r) < 0$, since $r < j_{\beta,1}$. So we may assume $r_{\wp}(f_\beta)$ be the unique root of $\Psi(r) = 0$ in $(0, j_{\beta,1})$ such that f_β in $\mathcal{S}^*_{\wp}$ in $|z| < r_{\wp}(f_\beta)$. Proof of other parts follows similarly. □

The Struve function $\mathbf{H}_\beta$ of first kind is a particular solution of the second-order inhomogeneous Bessel differential equation

$$z^2w''(z) + zw'(z) + (z^2 - \beta^2)w(z) = \frac{4(\frac{z}{2})^{\beta+1}}{\sqrt{\pi}\Gamma(\beta + \frac{1}{2})}$$

and have the following form:

$$\mathbf{H}_\beta(z) := \frac{(\frac{z}{2})^{\beta+1}}{\sqrt{\frac{\pi}{4}}\Gamma(\beta + \frac{1}{2})}{}_1F_2\left(1; \frac{3}{2}, \beta + \frac{3}{2}; -\frac{z^2}{4}\right),$$

where $-\beta - \frac{3}{2} \notin \mathbb{N}$ and ${}_1F_2$ is a hypergeometric function. Since it is not normalized, so we consider the following three normalized functions involving $\mathbf{H}_\beta$:

$$\begin{aligned} U_\beta(z) &= \left(\sqrt{\pi}2^\beta(\beta + \frac{3}{2})\mathbf{H}_\beta(z)\right)^{\frac{1}{\beta+1}}, \\ V_\beta(z) &= \sqrt{\pi}2^\beta z^{-\beta}\Gamma(\beta + \frac{3}{2})\mathbf{H}_\beta(z), \\ W_\beta(z) &= \sqrt{\pi}2^\beta z^{\frac{1-\beta}{2}}\Gamma(\beta + \frac{3}{2})\mathbf{H}_\beta(\sqrt{z}). \end{aligned} \tag{2.11}$$

Moreover, for $|\beta| \le \frac{1}{2}$, it has the Hadamard factorization given by

$$\mathbf{H}_\beta(z) = \frac{z^{\beta+1}}{\sqrt{\pi}2^\beta\Gamma(\beta + \frac{3}{2})}\prod_{n\ge 1}\left(1 - \frac{z^2}{z^2_{\beta,n}}\right), \tag{2.12}$$

where $z_{\beta,n}$ is the n-th positive root of $\mathbf{H}_\beta$ such that $z_{\beta,n+1} > z_{\beta,n}$ and $z_{\beta,1} > 1$ and also from (2.12), we obtain

$$\frac{z\mathbf{H}'_\beta(z)}{\mathbf{H}_\beta(z)} = (\beta + 1) - \sum_{n\ge 1}\frac{2z^2}{z^2_{\beta,n} - z^2}. \tag{2.13}$$

Now using the above representation, we can obtain the radii of cardioid-starlikeness of the functions U_β, V_β and W_β for the case $|\beta| \le 1/2$.

Theorem 2.5. *Let $|\beta| \le 1/2$. Then the radii of cardioid-starlikeness $r_{\wp}(U_\beta)$, $r_{\wp}(V_\beta)$ and $r_{\wp}(W_\beta)$ of the functions U_β, V_β and W_β are the smallest positive roots of the following equations, respectively:*

(i) $r\mathbf{H}'_\beta(r) - (1 - \frac{1}{e})(\beta + 1)\mathbf{H}_\beta(r) = 0$;
(ii) $r\mathbf{H}'_\beta(r) - ((1 + \beta) - \frac{1}{e})\mathbf{H}'_\beta(r) = 0$;
(iii) $\sqrt{r}\mathbf{H}'_\beta(\sqrt{r}) - (1 + \beta - \frac{2}{e})\mathbf{H}_\beta(\sqrt{r}) = 0$.

Proof. From (2.11) and (2.13), by logarithmic differentiation, we get

$$\begin{aligned} \frac{zU'_\beta(z)}{U_\beta(z)} &= \frac{1}{\beta + 1}\frac{z\mathbf{H}'_\beta(z)}{\mathbf{H}_\beta(z)} = 1 - \frac{1}{\beta + 1}\sum_{n\ge 1}\frac{2z^2}{z^2_{\beta,n} - z^2}, \\ \frac{zV'_\beta(z)}{V_\beta(z)} &= -\beta + \frac{z\mathbf{H}'_\beta(z)}{\mathbf{H}_\beta(z)} = 1 - \sum_{n\ge 1}\frac{2z^2}{z^2_{\beta,n} - z^2}, \end{aligned}$$

$$\frac{zW'_\beta(z)}{W_\beta(z)} = \frac{1-\beta}{2} + \frac{\sqrt{z}\mathbf{H}'_\beta(\sqrt{z})}{2\mathbf{H}_\beta(\sqrt{z})} = 1 - \sum_{n\geq 1} \frac{z}{z^2_{\beta,n} - z}. \tag{2.14}$$

Now applying the triangle inequality $||x| - |y|| \leq |x - y|$ and Lemma 2.1 in (2.14), we see that U_β, V_β and W_β belongs to $\mathcal{S}^*_\wp$, respectively whenever

$$\begin{aligned}
\left|\frac{zU'_\beta(z)}{U_\beta(z)} - 1\right| &\leq \frac{1}{\beta+1}\sum_{n\geq 1} \frac{2r^2}{z^2_{\beta,n} - r^2} \leq \frac{1}{e},\\
\left|\frac{zV'_\beta(z)}{V_\beta(z)} - 1\right| &\leq \sum_{n\geq 1} \frac{2r^2}{z^2_{\beta,n} - r^2} \leq \frac{1}{e},\\
\left|\frac{zW'_\beta(z)}{W_\beta(z)} - 1\right| &\leq \sum_{n\geq 1} \frac{r}{z^2_{\beta,n} - r} \leq \frac{1}{e}
\end{aligned} \tag{2.15}$$

holds, where $|z| = r < z_{\beta,1}$. Now to find the largest positive radius for which (2.15) holds. Let us consider the strictly increasing continuous functions

$$\begin{aligned}
\Psi_1(r) &:= \frac{1}{\beta+1}\sum_{n\geq 1} \frac{2r^2}{z^2_{\beta,n} - r^2} - \frac{1}{e},\\
\Psi_2(r) &:= \sum_{n\geq 1} \frac{2r^2}{z^2_{\beta,n} - r^2} - \frac{1}{e},\\
\Psi_3(r) &:= \sum_{n\geq 1} \frac{r}{z^2_{\beta,n} - r} - \frac{1}{e}.
\end{aligned}$$

Since $\lim_{r\to 0}\Psi_i(r) < 0$, $\Psi'_i(r) > 0$ for $i = 1$ to 3, $\lim_{r\to z_{\beta,1}}\Psi_i(r) > 0$ for $i = 1, 2$ and $\lim_{r\to z^2_{\beta,1}}\Psi_3(r) > 0$, there exist the unique positive roots, $r_\wp(U_\beta), r_\wp(V_\beta) \in (0, z_{\beta,1})$ and $r_\wp(W_\beta) \in (0, z^2_{\beta,1})$ for Ψ_i, respectively so that the inequalities in (2.15) holds in $|z| < r_\wp(U_\beta)$, $|z| < r_\wp(V_\beta)$ and $|z| < r_\wp(W_\beta)$, respectively. Further using (2.14) in $\Psi_i(r) = 0$, respectively, we obtain the desired equations. This completes the proof. □

The Lommel function $\mathcal{L}_{u,v}$ of first kind is a particular solution of the second-order inhomogeneous Bessel differential equation

$$z^2w''(z) + zw'(z) + (z^2 - v^2)w(z) = z^{u+1},$$

where $u \pm v \notin \mathbb{Z}^-$ and is given by

$$\mathcal{L}_{u,v} = \frac{z^{u+1}}{(u-v+1)(u+v+1)}{}_1F_2\left(1; \frac{u-v+3}{2}, \frac{u+v+3}{2}; -\frac{z^2}{4}\right),$$

where $\frac{1}{2}(-u \pm v - 3) \notin \mathbb{N}$ and ${}_1F_2$ is a hypergeometric function. Since it is not normalized, so we consider the following three normalized functions involving $\mathcal{L}_{u,v}$:

$$\begin{aligned}
f_{u,v}(z) &= ((u-v+1)(u+v+1)\mathcal{L}_{u,v}(z))^{\frac{1}{u+1}},\\
g_{u,v}(z) &= (u-v+1)(u+v+1)z^{-u}\mathcal{L}_{u,v}(z),\\
h_{u,v}(z) &= (u-v+1)(u+v+1)z^{\frac{1-u}{2}}\mathcal{L}_{u,v}(\sqrt{z}).
\end{aligned} \tag{2.16}$$

Authors in [5, 1] proved the radius of starlikeness for the normalized functions expressed in terms of $\mathcal{L}_{u,v}$:

$$f_{u-\frac{1}{2},\frac{1}{2}}(z), \quad g_{u-\frac{1}{2},\frac{1}{2}}(z) \quad \text{and} \quad h_{u-\frac{1}{2},\frac{1}{2}}(z), \tag{2.17}$$

where $0 \neq u \in (-1,1)$. Now we find the radii of cardioid-starlikeness of the functions defined in (2.17). For simplicity, we write these as f_u, g_u and h_u, respectively.

Theorem 2.6. *Let $0 \neq u \in (-1,1)$ and write $\mathcal{L}_{u-\frac{1}{2},\frac{1}{2}}(z) =: \mathcal{L}_u(z)$. Then the radii of cardioid-starlikeness $r_\wp(f_u)$, $r_\wp(g_u)$ and $r_\wp(h_u)$ of the functions f_u, g_u and h_u are the smallest positive roots of the following equations, respectively:*

(i) $\begin{cases} 2er\mathcal{L}'_u(r) - (2u+1)(e-1)\mathcal{L}_u(r) = 0, & for \quad u \in (-\frac{1}{2}, 1) \\ 2er\mathcal{L}'_u(r) - (2u+1)(e+1)\mathcal{L}_u(r) = 0, & for \quad u \in (-1, -\frac{1}{2}); \end{cases}$

(ii) $2er\mathcal{L}'_u(r) - (2eu+e-2)\mathcal{L}_u(r) = 0;$

(iii) $2e\sqrt{r}\mathcal{L}'_u(\sqrt{r}) - (2eu+e-4)\mathcal{L}_u(\sqrt{r}) = 0.$

Proof. We prove the first part. Let $0 \neq u \in (0,1)$. Then using a result from [7] (also see [5, Lemma 1, p. 3358]), we can write the Lommel function $\mathcal{L}_{u-\frac{1}{2},\frac{1}{2}}$ as follows:

$$\mathcal{L}_{u-\frac{1}{2},\frac{1}{2}}(z) = \frac{z^{u+\frac{1}{2}}}{u(u+1)}{}_1F_2\left(1; \frac{u+2}{2}, \frac{u+3}{2}; -\frac{z^2}{4}\right) = \frac{z^{u+\frac{1}{2}}}{u(u+1)}\phi_0(z), \tag{2.18}$$

where

$$\phi_0(z) = \prod_{n\geq 1}\left(1 - \frac{z^2}{z^2_{u,0,n}}\right),$$

and $z_{u,0,n}$ is the simple and real n-th positive root of ϕ_0. Also $z_{u,0,n} \in (n\pi, (n+1)\pi)$ which ensures $z_{u,0,n} > z_{u,0,1} > \pi > 1$. Now with this representation, after logarithmic differentiation, from (2.16) we get

$$\frac{zf'_u(z)}{f_u(z)} = \frac{z\mathcal{L}'_{u-\frac{1}{2},\frac{1}{2}}(z)}{(u+\frac{1}{2})\mathcal{L}_{u-\frac{1}{2},\frac{1}{2}}(z)} = 1 - \frac{1}{u+\frac{1}{2}}\sum_{n\geq 1}\frac{2z^2}{z^2_{u,0,n} - z^2}.$$

Using the triangle inequality and Lemma 2.1, we have $f_u \in \mathcal{S}^*_\wp$ provided

$$T(r) := \frac{1}{u+\frac{1}{2}}\sum_{n\geq 1}\frac{2r^2}{z^2_{u,0,n} - r^2} - \frac{1}{e} \leq 0$$

holds for $|z| = r < z_{u,0,1}$, where $T(r)$ is a strictly increasing continuous function in $(0, z_{u,0,1})$. Since $\lim_{r\to 0} T(r) < 0$, $\lim_{r\to z_{u,0,1}} T(r) > 0$ and $T'(r) > 0$, there exists a root $r_\wp(f_u) \in (0, z_{u,0,1})$ so that $f_u \in \mathcal{S}^*_\wp$ in $|z| < r_\wp(f_u)$. Now for the case $u \in (-1,0)$, we proceed as in the case when $u \in (0,1)$, just replacing u by $u+1$ and ϕ_0 by ϕ_1, where

$$\phi_1(z) = {}_1F_2\left(1; \frac{u+1}{2}, \frac{u+2}{2}; -\frac{z^2}{4}\right) = \prod_{n\geq 1}\left(1 - \frac{z^2}{z^2_{u,1,n}}\right)$$

and $z_{u,1,n}$ be the n-th positive root of ϕ_1.

Proof for the part (ii) and (iii) follows in a similar fashion as in the Theorem 2.5 by applying Lemma 2.1 on the following two equations, respectively together with the triangle inequality $||x| - |y|| \leq |x - y|$:

$$\frac{zg'_u(z)}{g_u(z)} = -u + \frac{1}{2} + \frac{z\mathcal{L}'_{u-\frac{1}{2},\frac{1}{2}}(z)}{(u+\frac{1}{2})\mathcal{L}_{u-\frac{1}{2},\frac{1}{2}}(z)} = 1 - \sum_{n\geq 1}\frac{2z^2}{z^2_{u,0,n} - z^2}$$

and

$$\frac{zh_u'(z)}{h_u(z)} = \frac{3-2u}{4} + \frac{\sqrt{z}\mathcal{L}'_{u-\frac{1}{2},\frac{1}{2}}(\sqrt{z})}{2\mathcal{L}_{u-\frac{1}{2},\frac{1}{2}}(\sqrt{z})} = 1 - \sum_{n\geq 1} \frac{z}{z_{u,0,n}^2 - z},$$

where $z_{u,0,n}$ is the n-th positive root of the function ϕ_0. □

The Legendre polynomials P_n are the solutions of the Legendre differential equation:

$$((1-z^2)P_n'(z))' + n(n+1)P_n(z) = 0,$$

where $n \in \mathbb{Z}^+$ and using Rodrigues′ formula, P_n can be represented in the form:

$$P_n(z) = \frac{1}{2^n n!}\frac{d^n(z^2-1)^n}{dz^n}$$

and it also satisfies the geometric condition $P_n(-z) = (-1)^n P_n(z)$. Moreover, the odd degree Legendre polynomials $P_{2n-1}(z)$ have only real roots which satisfy

$$0 = z_0 < z_1 < \cdots < z_{n-1} \quad \text{and} \quad -z_1 < \cdots < -z_{n-1}. \tag{2.19}$$

Thus the normalized form is as follows:

$$\mathcal{P}_{2n-1}(z) := \frac{P_{2n-1}(z)}{P_{2n-1}'(0)} = z + \sum_{k=2}^{2n-1} a_k z^k = a_{2n-1} z \prod_{k=1}^{n-1}(z^2 - z_k^2). \tag{2.20}$$

Theorem 2.7. *The radii of cardioid-starlikeness $r_\wp(\mathcal{P}_{2n-1}) \in (0, z_1)$ of the normalized odd degree Legendre polynomial is the smallest positive root of the following equation:*

$$er\mathcal{P}_{2n-1}'(r) - (e-1)\mathcal{P}_{2n-1}(r) = 0.$$

Proof. From (2.20), after logarithmic differentiation, we obtain

$$\frac{z\mathcal{P}_{2n-1}'(z)}{\mathcal{P}_{2n-1}(z)} = 1 - \sum_{k=1}^{n-1} \frac{2z^2}{z_k^2 - z^2}. \tag{2.21}$$

Now applying Lemma 2.1 on (2.21), we have $\mathcal{P}_{2n-1} \in \mathcal{S}_\wp^*$ whenever

$$\left|\frac{z\mathcal{P}_{2n-1}'(z)}{\mathcal{P}_{2n-1}(z)} - 1\right| \leq \sum_{k=1}^{n-1} \frac{2r^2}{z_k^2 - r^2} \leq \frac{1}{e}, \tag{2.22}$$

where $|z| = r < z_1$ and z_k satisfies the condition given in (2.19). Now let us consider the strictly increasing continuous function

$$T(r) := \sum_{k=1}^{n-1} \frac{2r^2}{z_k^2 - r^2} - \frac{1}{e}, \quad r \in (0, z_1).$$

We have to show that $T(r) \leq 0$ in $|z| \leq r < z_1$ so that (2.22) holds. Since $\lim_{r\to 0} T(r) < 0$, $\lim_{r\to z_1} T(r) > 0$ and $T'(r) > 0$, there exists a unique positive root $r_\wp(\mathcal{P}_{2n-1}) \in (0, z_1)$ of $T(r)$ such that $\mathcal{P}_{2n-1} \in \mathcal{S}_\wp^*$ in $|z| < r_\wp(\mathcal{P}_{2n-1})$. □

Remark 2.1 (**Generalization for $\mathcal{S}^*(\psi)$**). Consider the Ma-Minda function ψ as defined in (1.1). Let $a \in \psi(\mathbb{D}) \cap \mathbb{R}$, r_a is the radius depending on a and assume the maximal disk $|w-a| < r_a$ such that

$$\{w : |w-a| < r_a\} \subset \psi(\mathbb{D}).$$

Then we can easily extend Theorem 2.1 ($\psi(z) \neq (1+z)/(1-z)$), Theorem 2.3, Theorem 2.4, Theorem 2.5, Theorem 2.6 and Theorem 2.7 for $\mathcal{S}^*(\psi)$ using the corresponding radius r_1.

Here we mention the $\mathcal{S}^*(\psi)$-radius for the special functions without proof using Remark (2.1). In particular, we show that there exists an $\alpha \in (0,1]$ such that $R[\mathcal{S}^*(\psi)] \geq R[\mathcal{S}^*(1+\alpha z)]$, where we denote $\mathcal{S}^*(\psi)$-radius by $R[\mathcal{S}^*(\psi)]$.

In the following theorems we mean $R[\mathcal{S}^*(\psi)] = R[\mathcal{S}^*(1+r_1 z)]$.

Theorem 2.8 (Bessel function $\mathcal{J}_\beta$)**.** *Let $\beta > 0$. Then the $\mathcal{S}^*(\psi)$-radii $r_\wp(f_\beta)$, $r_\wp(g_\beta)$ and $r_\wp(h_\beta)$ of the functions f_β, g_β and h_β are the smallest positive root of the following equations, respectively:*

(i) $r\mathcal{J}'_\beta(r) + \beta(1-r_1)\mathcal{J}_\beta(r) = 0$;
(ii) $r\mathcal{J}'_\beta(r) - (\beta - r_1)\mathcal{J}_\beta(r) = 0$;
(iii) $\sqrt{r}\mathcal{J}'_\beta(\sqrt{r}) - (\beta - 2r_1)\mathcal{J}_\beta(\sqrt{r}) = 0$.

Theorem 2.9 (Struve function $\mathbf{H}_\beta$)**.** *Let $|\beta| \leq 1/2$. Then the $\mathcal{S}^*(\psi)$-radii $r_\wp(U_\beta)$, $r_\wp(V_\beta)$ and $r_\wp(W_\beta)$ of the functions U_β, V_β and W_β are the smallest positive root of the following equations, respectively:*

(i) $r\mathbf{H}'_\beta(r) - (1-r_1)(\beta+1)\mathbf{H}_\beta(r) = 0$;
(ii) $r\mathbf{H}'_\beta(r) - ((1+\beta) - r_1)\mathbf{H}'_\beta(r) = 0$;
(iii) $\sqrt{r}\mathbf{H}'_\beta(\sqrt{r}) - (1+\beta-2r_1)\mathbf{H}_\beta(\sqrt{r}) = 0$.

Theorem 2.10 (Lommel function $\mathcal{L}_{u,v}$)**.** *Let $0 \neq u \in (-1,1)$ and write $\mathcal{L}_{u-\frac{1}{2},\frac{1}{2}}(z) =: \mathcal{L}_u(z)$. Then the $\mathcal{S}^*(\psi)$-radii $r_\wp(f_u)$, $r_\wp(g_u)$ and $r_\wp(h_u)$ of the functions f_u, g_u and h_u are the smallest positive root of the following equations, respectively:*

(i) $\begin{cases} 2r\mathcal{L}'_u(r) - (2u+1)(1-r_1)\mathcal{L}_u(r) = 0, & for \quad u \in (-\frac{1}{2}, 1) \\ 2r\mathcal{L}'_u(r) - (2u+1)(1+r_1)\mathcal{L}_u(r) = 0, & for \quad u \in (-1, -\frac{1}{2}); \end{cases}$
(ii) $2r\mathcal{L}'_u(r) - (2u+1-2r_1)\mathcal{L}_u(r) = 0$;
(iii) $2\sqrt{r}\mathcal{L}'_u(\sqrt{r}) - (2u+1-4r_1)\mathcal{L}_u(\sqrt{r}) = 0$.

Theorem 2.11 (Legendre polynomials P_n)**.** *The $\mathcal{S}^*(\psi)$-radius $r_\wp(\mathcal{P}_{2n-1}) \in (0, z_1)$ of the normalized odd degree Legendre polynomial is the smallest positive root of the following equation:*

$$r\mathcal{P}'_{2n-1}(r) - (1-r_1)\mathcal{P}(r) = 0.$$

Corollary 2.12. *If $\alpha = r_1$ be the radius of the lagest disk $\{w : |w-1| < \alpha\}$ inside $\psi(\mathbb{D})$, where*

(i) $\alpha = \min\left\{\left|1 - \frac{1+A}{1+B}\right|, \left|1 - \frac{1-A}{1-B}\right|\right\} = \frac{A-B}{1+|B|}$ *when* $\psi(z) = \frac{1+Az}{1+Bz}$, *where* $-1 \leq B < A \leq 1$;
(ii) $\alpha = \sqrt{2 - 2\sqrt{2} + \sqrt{-2+2\sqrt{2}}}$ *when* $\psi(z) = \sqrt{2} - (\sqrt{2}-1)\sqrt{\frac{1-z}{1+2(\sqrt{2}-1)z}}$;
(iii) $\alpha = \sqrt{2} - 1$ *when* $\psi(z) = \sqrt{1+z}$;
(iv) $\alpha = e - 1$ *when* $\psi(z) = e^z$;
(v) $\alpha = 2 - \sqrt{2}$ *when* $\psi(z) = z + \sqrt{1+z^2}$;
(vi) $\alpha = \frac{e-1}{e+1}$ *when* $\psi(z) = \frac{2}{1+e^{-z}}$;
(vii) $\alpha = \sin 1$ *when* $\psi(z) = 1 + \sin z$.

Then Theorem 2.8, Theorem 2.9, Theorem 2.10 and Theorem 2.11 hold true for the class $\mathcal{S}^*(\psi)$.

2.2. **Majorization result for $\mathcal{S}^*_\wp$.**

Theorem 2.13. *Let $f \in \mathcal{A}$ and $g \in \mathcal{S}^*_\wp$. If f is majorized by g in $\mathbb{D}$. Then*

$$|f'(z)| \leq |g'(z)| \quad in \quad |z| \leq r_0, \tag{2.23}$$

where $r_0 \approx 0.380056$ is the smallest positive root of the equation

$$(1-r^2)(1-re^r) - r = 0.$$

Proof. Let $f \in \mathcal{A}$ and $g \in \mathcal{S}^*_\wp$, then we have

$$\frac{zg'(z)}{g(z)} = 1 + \omega(z)e^{\omega(z)}$$

or

$$\frac{g(z)}{g'(z)} = \frac{z}{1+\omega(z)e^{\omega(z)}}, \tag{2.24}$$

where ω is a Schwarz function. Let $\omega(z) = Re^{i\theta}$, where $R \leq r = |z| < 1$, then we have

$$\left|\frac{g(z)}{g'(z)}\right| = \left|\frac{z}{1+\omega(z)e^{\omega(z)}}\right| \leq \frac{r}{1-re^r}. \tag{2.25}$$

Now by definition of majorization, we have $f(z) = \Phi(z)g(z)$, where $\Phi(z)$ is analytic function satisfying $|\Phi(z)| \leq 1$ in $\mathbb{D}$ such that

$$f'(z) = g'(z)\left(\Phi'(z)\frac{g(z)}{g'(z)} + \Phi(z)\right).$$

Thus we have

$$|f'(z)| \leq |g'(z)|\left(\frac{1-|\Phi(z)|^2}{1-r^2}\frac{r}{1-re^r} + |\Phi(z)|\right) = |g'(z)|h(r,\beta), \tag{2.26}$$

where

$$h(r,\beta) = \frac{(1-\beta^2)r}{(1-r^2)(1-re^r)} + \beta.$$

Now to arrive at equation (2.23), it suffices to find r_0 such that $h(r,\beta) \leq 1$ holds. Or equivalently,

$$0 \leq (1-r^2)(1-re^r) - (1+\beta)r =: k(r,\beta).$$

Since $\frac{\partial}{\partial\beta}k(r,\beta) = -r$, therefore $\min_\beta k(r,\beta) = k(r,1)$. Write $k(r,1)$ as $k(r)$ and note that $k(0) = 1$ and $k(1) = -1$. So by continuity of $k(r)$, there exists a point r_0 such that $k(r) \geq 0$ for $0 \leq r \leq r_0$, where r_0 is the smallest positive root of the equation

$$(1-r^2)(1-re^r) - r = 0.$$

This ends the proof. □

2.3. **Convolution radius.** First, we need to recall the following result due to Ruscheweyh and Sheil-Small:

Lemma 2.3 ([25], p. 126)**.** *Suppose that either $g \in \mathcal{C}$, $h \in \mathcal{S}^*$ or else $g, h \in \mathcal{S}^*_{1/2}$. Then for any analytic function G in $\mathbb{D}$, we have*

$$\frac{g * hG(z)}{g * h(z)} \in \overline{co}G(\mathbb{D}),$$

where $\overline{co}G(\mathbb{D})$ is the closed convex hull of $G(\mathbb{D})$.

Observe that if we consider the function $\phi \in \mathcal{P}$, which is starlike but not convex then we can not directly apply Theorem 1.1 as such function do exist, for instance $\phi(z) = 1 + ze^z$. But keenly observing the proof of Lemma 2.3, we see that the unit disk $\mathbb{D}$ can be replaced by the sub-disk $\mathbb{D}_r := \{z : |z| < r\}$, where $0 < r \leq 1$ and consequently, we obtain the following modified result. Since the proof is similar, so it is omitted here.

Lemma 2.4. *Suppose either $g \in \mathcal{C}$, $h \in \mathcal{S}^*$ or else $g, h \in \mathcal{S}^*_{1/2}$. Then for any analytic function G in $\mathbb{D}_r$, we have $(g * hG(z))/(g * h(z)) \in \overline{co}G(\mathbb{D}_r)$, where $r \in [0, 1]$.*

This immediately gives

Theorem 2.14. *Let r_0 be the radius of convexity of ψ. If $g \in \mathcal{C}$ and $f \in \mathcal{S}^*(\psi)$. Then $f * g \in \mathcal{S}^*(\psi)$ in $|z| < r$, where $r = \min\{r_0, 1\}$.*

Corollary 2.15. *Let $f \in \mathcal{S}^*_\wp$ and $g \in \mathcal{C}$. Then $f * g \in \mathcal{S}^*_\wp$ in $\mathbb{D}_{r_0}$, where $r_0 = (3 - \sqrt{5})/2$ is the radius of convexity of $\wp$.*

Now consider the operators $\mathcal{F}_i : \mathcal{A} \to \mathcal{A}$ defined by

$$\mathcal{F}_1(f)(z) = f * g_1(z) = zf'(z)$$

$$\mathcal{F}_2(f)(z) = f * g_2(z) = \frac{1}{2}(f(z) + zf'(z))$$

$$\mathcal{F}_3(f)(z) = f * g_3(z) = \frac{k+1}{z^k}\int_0^z t^{k-1}f(t)dt, \quad \Re k > 0,$$

where $g_3(z) = \sum_{n=1}^{\infty}(k+1)/(k+n)z^n$, $g_2(z) = (z - z^2/2)/(1-z^2)^2$ and $g_1(z) = z/(1-z)^2$. Note that the function g_1 is convex in $|z| < 2 - \sqrt{3}$, g_2 is convex in $|z| < 1/2$ while $g_3 \in \mathcal{C}$. The above defined operators were introduced by Alexander, Livingston and Bernardi, respectively. Now we obtain the following result, where $\mathcal{S}^*_{SG} := \mathcal{S}^*(\frac{2}{e^{-z}+1})$, $\mathcal{S}^*_C := \mathcal{S}^*(1 + 4z/3 + 2z^2/3)$ and $\mathcal{S}^*_s := \mathcal{S}^*(1 + \sin z)$:

Corollary 2.16. *Let $\mathcal{F}_i$, $i = 1$ to 3 be the operators as defined above.*

(*i*) *Let $f \in \mathcal{S}^*_\wp$. Then $\mathcal{F}_i(f) \in \mathcal{S}^*_\wp$ in $\mathbb{D}_{r_i}$, where $r_1 = 2 - \sqrt{3}$, $r_2 = (3 - \sqrt{5})/2$ and $r_3 = (3 - \sqrt{5})/2$.*

(*ii*) *Let $f \in \mathcal{S}^*_C$. Then $\mathcal{F}_i(f) \in \mathcal{S}^*_C$ in $\mathbb{D}_{r_i}$, where $r_1 = 2 - \sqrt{3}$, $r_2 = 1/2$ and $r_3 = 1/2$.*

(*iii*) *Let $f \in \mathcal{S}^*_s$. Then $\mathcal{F}_i(f) \in \mathcal{S}^*_s$ in $\mathbb{D}_{r_i}$, where $r_1 = 2 - \sqrt{3}$, $r_2 = 0.345$ and $r_3 = 0.345$.*

(*iv*) *Let $f \in \mathcal{S}^*_{SG}$. Then $\mathcal{F}_i(f) \in \mathcal{S}^*_{SG}$ in $\mathbb{D}_{r_i}$, where $r_1 = 2 - \sqrt{3}$, $r_2 = 1/2$ and $r_3 = 1$.*

In 2010, Ali et al. [2] dealt with the problem of finding $\mathcal{S}^*(\psi)$-radii of the convolution $f * g$, between two starlike functions. In particular, they showed that if $f, g \in \mathcal{S}^*$ and $h_\rho(z) = f * g(\rho z)/\rho$, then $h_\rho \in \mathcal{SL}^*$ for $0 \leq \rho \leq (\sqrt{5} - 2)/(\sqrt{2} - 1) \approx 0.09778$. They used the property of the function ψ being convex. Now using Theorem 2.14, we can obtain the result even for the case when $\psi(\mathbb{D})$ is starlike. Here we have shown the usability of radius of convexity of ψ.

Theorem 2.17. *Let $f, g \in \mathcal{S}^*$ and $h_\rho(z) := f * g(\rho z)/\rho$. Then*

(*i*) *$h_\rho \in \mathcal{S}^*_\wp$ for $0 \leq \rho \leq (2e - \sqrt{4e^2 - 2e + 1})/(2e - 1) \approx 0.0957$,*

(*ii*) *$h_\rho \in \mathcal{S}^*_C$ for $0 \leq \rho \leq (3 - \sqrt{7})/2 \approx 0.177124$,*

(*iii*) $h_\rho \in \mathcal{S}_s^*$ *for* $0 \le \rho \le (\sqrt{\sin 1^2 + 2\sin 1 + 4} - 2)/(2 + \sin 1) \approx 0.185835$,
(*iv*) $h_\rho \in \mathcal{S}_b^*$ *for* $0 \le \rho \le (2e - \sqrt{3e^2 + e^{2/e}})/(e + e^{1/e}) \approx 0.122919$,
(*v*) $h_\rho \in \mathcal{S}_{SG}^*$ *for* $0 \le \rho \le (\sqrt{7e^2 + 6e + 3} - 2(1 + e))/(3e + 1) \approx 0.108309$.

The constants are best possible.

Proof.

(*i*) Let $H(z) = z + \sum_{n=2}^{\infty} n^2 z^n = (z(1+z))/(1-z)^3$. It is easy to see that

$$\left| \frac{zH'(z)}{H(z)} - \frac{1+r^2}{1-r^2} \right| \le \frac{4r}{1-r^2}, \quad |z| = r < 1. \tag{2.27}$$

Now by Lemma 2.1, the disk (2.27) lies inside the cardioid $\wp(\mathbb{D})$, provided

$$\frac{4r}{1-r^2} \le \frac{1+r^2}{1-r^2} - 1 + \frac{1}{e}$$

which in turn gives $r \le r_0 := (2e - \sqrt{4e^2 - 2e + 1})/(2e - 1)$. Define the function $h : \mathbb{D} \to \mathbb{C}$ by $h(z) := f(z) * g(z)$. Then $h(z) = F(z) * G(z) * H(z)$, where F and G are, respectively defined as $zF'(z) = f(z)$ and $zG'(z) = g(z)$. Since $f, g \in \mathcal{S}^*$, it follows that $F * G \in \mathcal{C}$. Also, $H(r_0 z)/r_0 \in \mathcal{S}_\wp^*$. Hence, using Theorem 2.14, we have

$$F(z) * G(z) * H(\rho_0)/\rho_0 \in \mathcal{S}_\wp^*,$$

where $\rho_0 = \min\{r_0, r_c\} = r_0$ and $r_c = (3 - \sqrt{5})/2$ is the radius of convexity of $\wp$. For $z = -\rho_0$, $zH'(z)/H(z) = (1 + 4z + z^2)/(1 - z^2) = 1 - 1/e$, which implies that ρ_0 is sharp.

Rest part's proof also follow in a similar fashion. □

Theorem 2.18. *Let* $f, g \in \mathcal{S}^*$ *and* $h_\rho(z) := f * g(\rho z)/\rho$. *Then* $h_\rho(z) \in \mathcal{S}^*\left(\frac{1+Az}{1+Bz}\right)$ *for*

$$0 \le \rho \le \frac{2(B^2 - 1) + \sqrt{4(1 - B^2)^2 + (A - B)^2}}{A - B} =: \rho_0,$$

where $-1 < B < A \le 1$.

Proof. Since for the function $p(z) \prec (1 + Az)/(1 + Bz)$, we have

$$\left| p(z) - \frac{1 - AB}{1 - B} \right| \le \frac{A - B}{1 - B^2}. \tag{2.28}$$

Therefore, for the disk (2.27) to lie inside the disk (2.28), we must have

$$\frac{1 - AB}{1 - B^2} - \frac{A - B}{1 - B^2} \le \frac{1 + r^2}{1 - r^2} \le \frac{1 - AB}{1 - B^2} + \frac{A - B}{1 - B^2}$$

and

$$\frac{4r}{1 - r^2} \le \frac{A - B}{1 - B^2}$$

which upon simplification hold for $r \le r_0 = \sqrt{(A - B)/(2 + A + B)}$ and $r \le \rho_0$ respectively, where ρ_0 is the smallest positive root of the following equation

$$(A - B)r^2 + 4(1 - B^2)r - (A - B) = 0.$$

Since $\min\{r_0, \rho_0\} = \rho_0$ and the class $\mathcal{S}^*\left(\frac{1+Az}{1+Bz}\right)$ is closed under convolution with convex functions, now the result follows in a similar way as in the part (*i*) of Theorem 2.17. □

3. Extremal problem for the class $\mathcal{S}^*(\psi)$

In 1961, Goluzin [13] obtained the set of extremal functions $f(z) = z/(1-xz)^2$, $|x| = 1$ for the problem of maximization of the quantity $\Re\Phi\left(\log(f(z)/z)\right)$ or $|\Phi\left(\log(f(z)/z)\right)|$ over the class $\mathcal{S}^*$, where ϕ is a non-constant entire function. In 1973, MacGregor [21] proved the result for the class $\mathcal{S}^*(\alpha) := \{f \in \mathcal{A} : \Re(zf'(z)/f(z)) > \alpha, \alpha \in [0,1)\}$. We observed that the proof given by MacGregor can be generalized to the Ma-Minda class $\mathcal{S}^*(\psi)$. Thus we have the following result:

Theorem 3.1. *Suppose Φ is a non-constant entire function and $0 < |z_0| < 1$ and assume that the class $\mathcal{S}^*(\psi)$ is closed. Then maximum of either*

$$\Re\Phi\left(\log\frac{f(z_0)}{z_0}\right) \quad or \quad \left|\Phi\left(\log\frac{f(z_0)}{z_0}\right)\right| \tag{3.1}$$

for functions in the class $\mathcal{S}^*(\psi)$ is attained only when the function is of the form

$$f(z) = z\exp\int_0^{\zeta z}\frac{\psi(t)-1}{t}dt, \tag{3.2}$$

where $|\zeta| = 1$.

Proof. Since the class $\mathcal{S}^*(\psi)$ is compact, therefore the problem under consideration has a solution. Moreover, in view of a result of Goluzin [13], in (3.1) it suffices to consider the continuous functional

$$\Re\Phi\left(\log\frac{f(z_0)}{z_0}\right).$$

Let $f \in \mathcal{S}^*(\psi)$. Then using a result from [19], $f(z)/z \prec f_0(z)/z =: F(z)$, where $f_0(z) = z\exp\int_0^z \frac{\psi(t)-1}{t}dt$ or equivalenlty $\log(f(z)/z) \prec \log F(z)$. Thus,

$$g(z) = \Phi\left(\log\frac{f(z)}{z}\right) \prec \Phi(\log F(z)) = G(z).$$

Note that G is also non-constant as is Φ. Thus for each $r \in (0,1)$ by subordination principle, we obtain $g(\overline{\mathbb{D}}_r) \subset G(\overline{\mathbb{D}}_r) = \Omega$. Since $G(xz) \prec G(z)$ for $|x| \leq 1$ is obvious, therefore for $|z_0| = r$, we have $\{g(z_0) : g \prec G \text{ in } \mathbb{D}\} = \Omega$. Now by considering a support line to the compact set Ω, we conclude that

$$\max_{f\in\mathcal{S}^*(\psi)} \Re\Phi\left(\log\frac{f(z_0)}{z_0}\right) = \Re w_1, \quad w_1 \in \partial\Omega.$$

Since G is also an open map, therefore there exists a point z_1 where $|z_1| = r$ and $G(z_1) = w_1$ such that among finitely many w_1, for one suitable w_1, we have

$$\Phi\left(\log\frac{f(z_0)}{z_0}\right) = w_1,$$

where f is the solution for the extremal problem. Now by the well known Lindelöf Principle, we have

$$\Phi\left(\log\frac{f(z)}{z}\right) = \Phi(\log F(xz)), \tag{3.3}$$

that is, if f is the desired solution, then (3.3) holds for some x, $|x| = 1$. Since Φ is non-constant analytic function, so we may write

$$\Phi(w) = c_0 + c_n w^n + c_{n+1}w^{n+1} + \cdots;\ c_n \neq 0.$$

If we set $\log(f(z)/z) = \alpha_1 z + \alpha_2 z^2 + \cdots$ and $\log(F(z)) = \beta_1 z + \beta_2 z^2 + \cdots$, then from (3.3), comparing the coefficients, we get $c_n \alpha_1^n = c_n \beta_1^n$. Or equivalently, $\alpha_1^n = \beta_1^n$, which in particular implies that $|\alpha_1| = |\beta_1|$. Since $\log(f(z)/z) \prec \log F(xz)$, $|\alpha_1| = |\beta_1|$ is possible only if $\log(f(z)/z) = \log F(xyz)$ for some $|y| = 1$. Therefore, we conclude that

$$f(z) = z \exp \int_0^{uz} \frac{\psi(t) - 1}{t} dt,$$

where $|u| = 1$ if f is a solution to the extremal problem. □

Now as an application of the Theorem 3.1, we obtain a result due to MacGregor [21]:

Corollary 3.2. [21] *Suppose Φ is a non-constant entire function and $0 < |z_0| < 1$. Then the maximum of the expression* (3.1) *for functions in the class $\mathcal{S}^*(\alpha)$ is attained only when the function is of the form*

$$f(z) = \frac{z}{(1 - \zeta z)^{2-2\alpha}}, \ |\zeta| = 1.$$

Proof. If $f \in \mathcal{S}^*(\alpha)$, then $f(z)/z \prec 1/(1-z)^{2-2\alpha}$ and hence the result.

Corollary 3.3. *Suppose Φ is a non-constant entire function and $0 < |z_0| < 1$. Then the maximum of the expression* (3.1) *for functions in the class $\mathcal{S}^*_\wp$ is attained only when the function is of the form*

$$f(z) = z \exp(e^{\zeta z} - 1), \ |\zeta| = 1.$$

Proof. If $f \in \mathcal{S}^*_\wp$, then $f(z)/z \prec \exp(e^z - 1)$ and hence the result. □

4. Some sufficient conditions for $\mathcal{S}^*_\wp$

In this section, we determine the sufficient conditions for the functions $z/(1 + \sum_{k=1}^{\infty} a_k z^k)$, $z/(1+z^n)^k$ and certain other types of functions to be in $\mathcal{S}^*_\wp$.

Theorem 4.1. *Let $f(z) = z/(1 + \sum_{k=1}^{\infty} a_k z^k)$. If the coefficients of f satisfy*

$$|1 - a| + \sum_{k=1}^{\infty} (R_a + |1 - a - k|)|a_k| \leq R_a,$$

where a and R_a is as defined in Lemma 2.1. Then $f \in \mathcal{S}^*_\wp$.

Proof. For $f(z) = z/(1 + \sum_{k=1}^{\infty} a_k z^k)$, we have

$$\left| \frac{zf'(z)}{f(z)} - a \right| = \left| 1 - a - \frac{\sum_{k=1}^{\infty} k a_k z^k}{1 + \sum_{k=1}^{\infty} a_k z^k} \right|.$$

Thus by Lemma 2.1, $f \in \mathcal{S}^*_\wp$, if

$$\left| 1 - a - \frac{\sum_{k=1}^{\infty} k a_k z^k}{1 + \sum_{k=1}^{\infty} a_k z^k} \right| \leq R_a.$$

The above inequality holds whenever

$$|1 - a| + \sum_{k=1}^{\infty} |1 - a - k||a_k| r^k \leq R_a (1 - \sum_{k=1}^{\infty} |a_k| r^k)$$

or equivalently,

$$|1-a|+\sum_{k=1}^{\infty}(|1-a-k|+R_a)|a_k|r^k \leq R_a.$$

Letting r tends to 1^-, completes the proof. □

Theorem 4.2. *Let $f(z)=z/(1+z^k)^n$, where $n,k\in\mathbb{Z}^+$ are fixed. Then $f\in\mathcal{S}^*_\wp$ for*

$$|z|<\left(\frac{R_a-|1-a|}{R_a+|1-a-kn|}\right)^{1/k},$$

where a and R_a is as defined in Lemma 2.1.

Proof. For $f(z)=z/(1+z^k)^n$, we have

$$\left|\frac{zf'(z)}{f(z)}-a\right|=\left|1-a\frac{knz^k}{1+z^k}\right|.$$

Thus by Lemma 2.1, $f\in\mathcal{S}^*_\wp$, if

$$\left|1-a-\frac{knz^k}{1+z^k}\right|<R_a.$$

The above inequality holds whenever

$$|1-a|+|1-a-kn||z|^k<R_a(1-|z|^k)$$

which simplifies to

$$|z|^k<\frac{R_a-|1-a|}{R_a+|1-a-kn|}.$$

Hence the result. □

Theorem 4.3. *Let $p(z)$ be a polynomial such that $p(0)=1$ and $\deg p(z)=m$. Let $R=\min\{|z|:p(z)=0,z\neq 0\}$. Then the function $f(z)=z(p(z))^{\beta/m}\in\mathcal{S}^*_\wp$ for*

$$|z|<\frac{R(R_a-|1-a|)}{|\beta|+R_a-|1-a|},$$

where a and R_a is as defined in Lemma 2.1.

Proof. Assume that z_k, $(k=1,2,...,m)$ are zeros of the polynomial $p(z)$. For the function $f(z)=z(p(z))^{\beta/m}$, we have

$$\frac{zf'(z)}{f(z)}=1+\frac{\beta}{m}\sum_{k=1}^{\infty}\frac{z}{z-z_k}$$

or equivalently,

$$\frac{zf'(z)}{f(z)}-a=1-a+\frac{\beta}{m}\sum_{k=1}^{\infty}\left(\frac{z}{z-z_k}+\frac{r^2}{R^2-r^2}-\frac{r^2}{R^2-r^2}\right).$$

Thus by Lemma 2.1, $f\in\mathcal{S}^*_\wp$, if

$$(|1-a|-R_a)(R^2-r^2)+|\beta|(Rr+r^2)<0.$$

The above inequality is satisfied if $|z|=r<R(R_a-|1-a|)/(|\beta|+R_a-|1-a|)$. This completes the proof. □

5. Subordination results for $\mathcal{S}^*_\wp$

To prove our differential subordination results, we will need the following lemma due to Miller and Mocanu:

Lemma 5.1. [23] *Let q be the univalent in $\mathbb{D}$. Let θ and ϕ be analytic in a domain D containing $q(\mathbb{D})$, with $\phi(w) \neq 0$, when $w \in q(\mathbb{D})$. Set $Q(z) = zq'(z)\phi(q(z))$, $h(z) = \theta(q(z)) + Q(z)$ and suppose that Q is starlike. In addition, assume that*

$$\Re\frac{zh'(z)}{Q(z)} = \Re\left(\frac{\theta'(q(z))}{\phi(q(z))} + \frac{zQ'(z)}{Q(z)}\right) > 0.$$

If p is analytic in $\mathbb{D}$, with $p(0) = q(0)$, $p(\mathbb{D}) \subset D$ and

$$\theta(p(z)) + zp'(z)\phi(p(z)) \prec h(z), \tag{5.1}$$

then $p \prec q$ and q is the best dominant.

Theorem 5.1. *Let $0 < \alpha < 1$, $0 < B < A < 1$ and $k = 1 + \sqrt{2}$. If $p \in \mathcal{A}_0$ and satisfies the differential subordination*

$$1 + \beta zp'(z) \prec \wp(z),$$

then

(*i*) $p(z) \prec \phi_0(z) := 1 + \frac{z}{k}\left(\frac{k+z}{k-z}\right)$ *for* $\beta \geq \frac{k(e-1)(k+1)}{e(k-1)} \approx 3.68427$.
(*ii*) $p(z) \prec \sqrt{1+z}$ *for* $\beta \geq \frac{e-1}{\sqrt{2}-1}$.
(*iii*) $p(z) \prec \frac{1+Az}{1+Bz}$ *for* $\beta \geq \max\left\{\frac{(1-1/e)(1-B)}{A-B}, \frac{(e-1)(1+B)}{A-B}\right\}$.
(*iv*) $p(z) \prec 1 + \sin z$ *for* $\beta \geq \frac{e-1}{\sin 1}$.
(*v*) $p(z) \prec z + \sqrt{1+z^2}$ *for* $\beta \geq \frac{e-1}{\sqrt{2}}$.
(*vi*) $p(z) \prec e^z$ *for* $\beta \geq 1$.

The bounds are sharp.

Proof. We set $\theta(w) = 1$ and $\phi(w) = \beta \neq 0$. Let $\Psi_\beta(z, p(z)) := 1 + \beta zp'(z)$.

(*i*) The differential equation $\Psi_\beta(z, p(z)) = \wp(z)$ has an analytic solution given by

$$q_\beta(z) = 1 + \frac{1}{\beta}(e^z - 1).$$

Since the function $Q(z) = zq'_\beta(z)\phi(q_\beta(z)) = \wp(z) - 1 = ze^z \in \mathcal{S}^*$ and the function $h(z) = \theta(q_\beta(z)) + Q(z)$ satisfies

$$\Re\left(\frac{zh'(z)}{Q(z)}\right) = \Re\left(\frac{zQ'(z)}{Q(z)}\right) > 0.$$

Therefore, by Lemma 5.1, the following differential subordination implication holds:

$$\Psi_\beta(z, p(z)) \prec 1 + \beta zq'_\beta(z) \Rightarrow p \prec q_\beta.$$

Now to show that $p \prec \phi_0$, we need to show that $q_\beta \prec \phi_0$ and for $q_\beta \prec \phi_0$ it is necessary that $\phi_0(-1) \leq q_\beta(-1)$ and $q_\beta(1) \leq \phi_0(1)$ which gives $\beta \geq \beta_1$ and $\beta \geq \beta_2$ respectively, where

$$\beta_1 = \frac{k(e-1)(k+1)}{e(k-1)} \quad \text{and} \quad \beta_2 = \frac{k(e-1)(k-1)}{k+1}.$$

Let $\beta \geq \max\{\beta_1, \beta_2\} = \beta_1$. Since the difference of the square of the distances from the point $(1,0)$ to the points on the boundary curves $\phi_0(e^{i\theta})$ and $q_\beta(e^{i\theta})$, respectively is given by

$$T(\theta) := \frac{4(1+\sin^2\theta)}{(k^2-2k\cos\theta+1)^2} - \frac{e^{2\cos\theta}+1}{\beta^2} + \frac{2e^{\cos\theta}\cos(\sin\theta)}{\beta}, \quad 0 \leq \theta \leq \pi.$$

A calculation shows that $T'(\theta) \leq 0$. Therefore, $T(\theta)$ decreases on $[0,\pi]$, which further implies that the condition $\phi_0(-1) \leq q_\beta(-1) \leq q_\beta(1) \leq \phi_0(1)$, that is $\beta \geq \beta_1$ is also sufficient for $q_\beta \prec \phi_0$. Hence the result. Rest of the results follow in a similar fashion. □

Theorem 5.2. *Let $0 < \alpha < 1$, $0 < B < A < 1$ and $k = 1+\sqrt{2}$. If $p \in \mathcal{A}_0$ and satisfies the differential subordination*

$$1 + \beta\frac{zp'(z)}{p(z)} \prec \wp(z),$$

then

(*i*) $p(z) \prec e^z$ *for* $\beta \geq e-1$.
(*ii*) $p(z) \prec \phi_0(z)$ *for* $\beta \geq (1/e-1)/\log\left(1-\frac{k-1}{k(k+1)}\right) \approx 3.3583$.
(*iii*) $p(z) \prec \sqrt{1+z}$ *for* $\beta \geq \frac{e-1}{\log\sqrt{2}}$.
(*iv*) $p(z) \prec \frac{1+Az}{1+Bz}$ *for* $\beta \geq \max\left\{\frac{1-1/e}{\log((1-B)/(1-A))}, \frac{e-1}{\log((1+A)/(1+B))}\right\}$.
(*v*) $p(z) \prec 1+\sin z$ *for* $\beta \geq \frac{e-1}{\log(1+\sin 1)}$.
(*vi*) $p(z) \prec z+\sqrt{1+z^2}$ *for* $\beta \geq \frac{e-1}{\log(k)}$.

The bounds are sharp.

Proof. Let $\theta(w) = 1$ and $\phi(w) = \beta/\omega \neq 0$. Let $\Phi_\beta(z,p(z)) := 1+\beta\frac{zp'(z)}{p(z)}$.

(*i*) The differential equation $\Phi_\beta(z,p(z)) = \wp(z)$ has an analytic solution given by

$$q_\beta(z) = \exp\left(\frac{e^z-1}{\beta}\right).$$

Since the function $Q(z) = zq'_\beta(z)\phi(q_\beta(z)) = \wp(z) - 1 = ze^z \in \mathcal{S}^*$ and $h(z) = \theta(q_\beta(z)) + Q(z)$ satisfies $\Re(zh'(z)/Q(z)) = \Re(zQ'(z)/Q(z)) > 0$. Therefore, by Lemma 5.1, the following differential subordination implication holds:

$$\Phi_\beta(z,p(z)) \prec 1+\beta\frac{zq'_\beta(z)}{q_\beta(z)} \Rightarrow p \prec q_\beta.$$

Now to show that $p \prec e^z$, we need to show that $q_\beta \prec e^z$ and for $q_\beta \prec e^z$, it is necessary that $e^{-1} \leq q_\beta(-1)$ and $q_\beta(1) \leq e$ which gives $\beta \geq \beta_1$ and $\beta \geq \beta_2$ respectively, where

$$\beta_1 = (1-1/e) \quad \text{and} \quad \beta_2 = e-1.$$

Let $\beta \geq \beta_2$. To prove that $q_\beta(z) \prec e^z$, now it only suffices to show that $(e^z-1)/\beta \prec z$, which is equivalent to prove that for $z \in \overline{\mathbb{D}}$

$$|(e^z-1)/\beta| \leq 1, \tag{5.2}$$

and thus, for $z = e^{i\theta}$, we have

$$|(e^{e^{i\theta}}-1)/\beta| = \sqrt{(e^{2\cos\theta}+1-2e^{\cos\theta}\cos(\sin\theta))/\beta^2} =: \sqrt{T(\theta)}.$$

Note that $T(\theta) = T(-\theta)$. So we may consider the interval $0 \leq \theta \leq \pi$. Now since

$$T'(\theta) = 2e^{\cos\theta}(\sin(\theta - \sin\theta) - \sin\theta)/\beta^2 = 0$$

if and only if

$$\cos\left(\theta - \frac{\sin\theta}{2}\right)\sin\left(\frac{\sin\theta}{2}\right) = 0$$

if and only if $\theta \in \{0, \theta_0, \pi\}$, where $\theta_0 \approx 2.02098$ is the only root of the equation

$$2\theta - \sin\theta - \pi = 0.$$

Thus, we see that $T'(\theta) < 0$ over $(0, \theta_0)$ and $T'(\theta) > 0$ over (θ_0, π), which implies that $T(\theta)$ decreases over $[0, \theta_0]$ and increases over $[\theta_0, \pi]$ and therefore,

$$\max_{0\leq\theta\leq\pi} T(\theta) = \max\{T(0), T(\pi)\} = T(0) \leq 1$$

and (5.2) holds. Hence, $\beta \geq \beta_2$ is also sufficient for $q_\beta(z) \prec e^z$. Rest of the results follow in a similar fashion. □

Theorem 5.3. *Let $0 < \alpha < 1$, $0 < B < A < 1$ and $k = 1 + \sqrt{2}$. If $p \in \mathcal{A}_0$ and satisfies the differential subordination*

$$1 + \beta\frac{zp'(z)}{p^2(z)} \prec \wp(z),$$

then

(i) $p(z) \prec e^z$ *for* $\beta \geq \frac{e-1}{1-1/e}$.

(ii) $p(z) \prec \phi_0(z)$ *for* $\beta \geq (e-1)/\left(1 - \frac{1}{1+\frac{1}{k}\left(\frac{k+1}{k-1}\right)}\right)$.

(iii) $p(z) \prec \sqrt{1+z}$ *for* $\beta \geq \frac{(e-1)(k-1)}{k-2}$.

(iv) $p(z) \prec \frac{1+Az}{1+Bz}$ *for* $\beta \geq \max\left\{\frac{(1-1/e)(1-A)}{A-B}, \frac{(e-1)(1+A)}{A-B}\right\}$.

(v) $p(z) \prec 1 + \sin z$ *for* $\beta \geq \frac{(e-1)(1+\sin 1)}{\sin 1}$.

(vi) $p(z) \prec z + \sqrt{1+z^2}$ *for* $\beta \geq \frac{(e-1)k}{k-1}$.

The bounds are sharp.

Proof. We set $\theta(w) = 1$ and $\phi(w) = \beta/w^2 \neq 0$. Let $\mho_\beta(z, p(z)) := 1 + \beta\frac{zp'(z)}{p^2(z)}$.

(i) The differential equation $\mho_\beta(z, p(z)) = \wp(z)$ has an analytic solution given by

$$q_\beta(z) = \frac{1}{1 - \frac{1}{\beta}(e^z - 1)}.$$

Since the function $Q(z) = zq'_\beta(z)\phi(q_\beta(z)) = \wp(z) - 1 = ze^z \in \mathcal{S}^*$ and the function $h(z) = \theta(q_\beta(z)) + Q(z)$ satisfies

$$\Re\left(\frac{zh'(z)}{Q(z)}\right) = \Re\left(\frac{zQ'(z)}{Q(z)}\right) > 0.$$

Therefore, by Lemma 5.1, the following differential subordination implication holds:

$$\Psi_\beta(z, p(z)) \prec 1 + \beta z q'_\beta(z) \Rightarrow p \prec q_\beta.$$

Now to show that $p \prec e^z$, we need to show that $q_\beta \prec e^z$ and for $q_\beta \prec e^z$, it is necessary that $e^{-1} \leq q_\beta(-1)$ and $q_\beta(1) \leq e$ which gives $\beta \geq \beta_1$ and $\beta \geq \beta_2$ respectively, where

$$\beta_1 = \frac{1-1/e}{e-1} \quad \text{and} \quad \beta_2 = \frac{e-1}{1-1/e}.$$

Let $\beta \geq \max\{\beta_1, \beta_2\} = \beta_2$. Note that e^z and e^{-z} both maps $\mathbb{D}$ onto $\Omega := \{w \in \mathbb{C} : w = e^z \text{ and } z \in \mathbb{D}\}$. Therefore, to prove that $q_\beta \prec e^z$, or equivalently $q_\beta(z) \in \Omega$, it suffices to show that $q_\beta \prec e^{-z}$, which is equivalent to show that

$$1 - \frac{1}{\beta}(e^z - 1) \prec e^z. \tag{5.3}$$

Now since the difference of the square of the distances from the point $(1, 0)$ to the points on the boundary curves $e^{e^{i\theta}}$ and $1/q_\beta(e^{i\theta})$, respectively is given by

$$T(\theta) := \left(1 - \frac{1}{\beta^2}\right) S(\theta),$$

where $S(\theta) = 1 + e^{2\cos\theta} - 2e^{\cos\theta}\cos(\sin\theta)$. Here, we may consider the interval $0 \leq \theta \leq \pi$, since $T(\theta) = T(-\theta)$. A calculation shows that $S'(\theta) \leq 0$ for $\theta \in [0, \pi]$. Therefore, $T(\theta)$ decreases on $[0, \pi]$ and also

$$\min_{0\leq\theta\leq\pi} T(\theta) = \left(1 - \frac{1}{\beta^2}\right) S(\pi) = \left(1 - \frac{1}{\beta^2}\right)(1 + e^{-2} - 2e^{-1}) > 0,$$

which means (5.3) is true. Hence $\beta \geq \beta_2$ is also sufficient for $q_\beta \prec e^z$. Proof of the rest of the results are much akin and so are omitted here. □

Remark 5.1. If $p \in \mathcal{A}_0$, $\beta > 0$ and satisfies the differential subordination

$$1 + \beta\frac{zp'(z)}{p^n(z)} \prec \wp(z), \quad n \geq 2$$

then

$$p(z) \prec q_\beta(z) := \left(\frac{1}{1 - \frac{n-1}{\beta}(e^z - 1)}\right)^{\frac{1}{n-1}},$$

where q_β is the best dominant.

Theorem 5.4. *Let p be an analytic function in $\mathbb{D}$ with $p(0) = 1$. Let $0 < \alpha < 1$, $0 < B < A < 1$, $k = 1 + \sqrt{2}$ and $\wp(z) = 1 + ze^z$. Set*

$$\Psi_\beta(z, p(z)) = 1 + \beta z p'(z).$$

If $p \in \mathcal{A}_0$ and satisfies any of the following differential subordinations:

- *(i)* $\Psi_\beta(z, p(z)) \prec \phi_0(z)$ *for* $\beta \geq e(-1 + 2k\log(1 + 1/k))/k \approx 0.75822$.
- *(ii)* $\Psi_\beta(z, p(z)) \prec \sqrt{1+z}$ *for* $\beta \geq 2e(1 - \log 2) \approx 1.66822$.
- *(iii)* $\Psi_\beta(z, p(z)) \prec \varphi_\alpha(z) := 1 + \frac{z}{1-\alpha z^2}$ *for* $\beta \geq e\log(\frac{1+\sqrt{\alpha}}{1-\sqrt{\alpha}})/(2\sqrt{\alpha})$.
- *(iv)* $\Psi_\beta(z, p(z)) \prec \frac{1+Az}{1+Bz}$ *for* $\beta \geq \max\{\frac{e(A-B)}{B}\log(1 - B), \frac{A-B}{eB}\log(1 + B)\}$.
- *(v)* $\Psi_\beta(z, p(z)) \prec 1 + \sin z$ *for* $\beta \geq e\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!(2n+1)} \approx 2.57172$.
- *(vi)* $\Psi_\beta(z, p(z)) \prec z + \sqrt{1+z^2}$ *for* $\beta \geq e(3 - k + \log(k/2)) \approx 2.10399$.
- *(vii)* $\Psi_\beta(z, p(z)) \prec e^z$ *for* $\beta \geq e\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{(n!)(n)} \approx 2.16538$.

Then $p \prec \wp$ and the bounds are sharp.

Proof. We set $\theta(w) = 1$ and $\phi(w) = \beta \neq 0$. To prove our results, we make use of the inequality:

$$\wp(-1) \leq q_\beta(-1) \leq q_\beta(1) \leq \wp(1), \tag{5.4}$$

if $q_\beta \prec \wp$ which gives necessary as well as sufficient condition for the required subordination.

(i) The differential equation $\Psi_\beta(z, p(z)) = \phi_0(z)$ has an analytic solution given by

$$q_\beta(z) = 1 - \frac{1}{\beta k}\left(z + 2k \log\left(1 - \frac{z}{k}\right)\right).$$

Since the function

$$Q(z) = zq'_\beta(z)\phi(q_\beta(z)) = \phi_0(z) - 1 = \frac{z}{k}\left(\frac{k+z}{k-z}\right) \in \mathcal{S}^*$$

and the function $h(z) = \theta(q_\beta(z)) + Q(z)$ satisfies

$$\Re\left(\frac{zh'(z)}{Q(z)}\right) = \Re\left(\frac{zQ'(z)}{Q(z)}\right) > 0.$$

Therefore, by Lemma 5.1, the following differential subordination implication holds:

$$\Psi_\beta(z, p(z)) \prec 1 + \beta z q'_\beta(z) \Rightarrow p \prec q_\beta.$$

Now to show that $p \prec \wp$, we need to show that $q_\beta \prec \wp$. Now if $q_\beta \prec \wp$, then (5.4) must hold, that is, the inequalities $\wp(-1) \leq q_\beta(-1)$ and $q_\beta(1) \leq \wp(1)$ respectively gives $\beta \geq \beta_1$ and $\beta \geq \beta_2$, where

$$\beta_1 = \frac{e}{k}\left(2k \log\left(\frac{k+1}{k}\right) - 1\right) \quad \text{and} \quad \beta_2 = \frac{1}{ek}\left(2k \log\left(\frac{k}{k-1}\right) + 1\right).$$

Let $\beta \geq \max\{\beta_1, \beta_2\} = \beta_1$. Note that $q_\beta(\mathbb{D})$ is convex, symmetric about real axis and $q_\beta(\mathbb{D}) \subset q_{\beta_1}(\mathbb{D})$ for each $\beta \geq \beta_1$. Also, $|\arg(1+z)| \leq \arctan(r/(1-r))$ for $|z| = r < 1$. Now to prove $q_\beta \prec \wp$, in view of Lemma 2.1, it only suffices to show that $\max_{0\leq\theta\leq\pi} |\arg q_{\beta_1}(e^{i\theta})| = \max_{0\leq\theta\leq\pi} \left|\arg \wp(e^{i\theta})\right|$. Since

$$\begin{aligned}
&\max_{0\leq\theta\leq\pi} |\arg q_{\beta_1}(e^{i\theta})| \\
&= \max_{0\leq\theta\leq\pi} \left|\arctan(\Im q_{\beta_1}(e^{i\theta})/\Re q_{\beta_1}(e^{i\theta}))\right| \\
&= \max_{0\leq\theta\leq\pi} \left|\arctan\left(\frac{(-2k \arg\left(1 - \frac{e^{i\theta}}{k}\right) - \sin\theta)}{M - \cos\theta - k\log((1 + \frac{1}{k^2} - \frac{2\cos\theta}{k})}\right)\right| \\
&\leq \arctan\left(\frac{2k \arctan(\frac{1}{k-1}) + 1}{M(1 - 1/e)}\right) \approx 0.7719 \\
&< \max_{0\leq\theta\leq\pi} |\arg(1 + ze^z)| \approx 1.41022,
\end{aligned}$$

where $M = e(-1 + 2k\log(1 + 1/k))$. Therefore, $\beta \geq \beta_1$ is also sufficient for $q_\beta \prec \wp$ to hold. Hence the result. Other results follow similarly. □

Let $f \in \mathcal{A}$ and setting $p(z) = zf'(z)/f(z)$ and $p(z) = f(z)/z$. Then using Theorem 5.4, we obtain sufficient conditions for f to be in $\mathcal{S}^*_\wp$.

Corollary 5.5. *Let $f \in \mathcal{A}$ and set*

$$\Psi_\beta\left(z, \frac{zf'(z)}{f(z)}\right) = 1 + \beta\frac{zf'(z)}{f(z)}\left(1 - \frac{zf'(z)}{f(z)} + \frac{zf''(z)}{f'(z)}\right).$$

If $\Psi_\beta\left(z, \frac{zf'(z)}{f(z)}\right)$ satisfies any of the conditions (i)-(vii) of the Theorem 5.4, then $f \in \mathcal{S}^*_\wp$.

Corollary 5.6. *Let $f \in \mathcal{A}$ and set*

$$\Psi_\beta\left(z, \frac{f(z)}{z}\right) = 1 + \beta\left(f'(z) - \frac{f(z)}{z}\right).$$

If $\Psi_\beta\left(z, \frac{f(z)}{z}\right)$ satisfies any of the conditions (i)-(vii) of the Theorem 5.4, then $f \in \mathcal{S}^*_\wp$.

Theorem 5.7. *Let $0 < \alpha < 1$, $0 < B < A < 1$, $\gamma = \log(e/(e-1))$, $k = 1 + \sqrt{2}$ and $\wp(z) = 1 + ze^z$ and $p \in \mathcal{A}_0$. Set*

$$\Phi_\beta(z, p(z)) = 1 + \beta\frac{zp'(z)}{p(z)}.$$

If p satisfies any of the following differential subordinations:

- *(i)* $\Phi_\beta(z, p(z)) \prec \phi_0(z)$ *for* $\beta \geq (-1 + 2k\log(1 + 1/k))/\gamma k \approx 0.60812$.
- *(ii)* $\Phi_\beta(z, p(z)) \prec \sqrt{1+z}$ *for* $\beta \geq 2(1 - \log 2)/\gamma \approx 1.33799$.
- *(iii)* $\Phi_\beta(z, p(z)) \prec \varphi_\alpha(z)$ *for* $\beta \geq \log(\frac{1+\sqrt{\alpha}}{1-\sqrt{\alpha}})/(2\gamma\sqrt{\alpha})$.
- *(iv)* $\Phi_\beta(z, p(z)) \prec (1+Az)/(1+Bz)$ *for* $\beta \geq \max\left\{\frac{(A-B)\log(1/(1-B))}{\gamma B}, \frac{(A-B)\log(1+B)}{B\log(1+e)}\right\}$.
- *(v)* $\Phi_\beta(z, p(z)) \prec 1 + \sin z$ *for* $\beta \geq (1/\gamma)\sum_{n=0}^{\infty} \frac{(-1)^(3n+2)}{(2n+1)!(2n+1)} \approx 2.06264$.
- *(vi)* $\Phi_\beta(z, p(z)) \prec z + \sqrt{1+z^2}$ *for* $\beta \geq (3 - k + \log(k/2))/\gamma \approx 1.6875$.
- *(vii)* $\Phi_\beta(z, p(z)) \prec e^z$ *for* $\beta \geq (1/\gamma)\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{(n!)(n)} \approx 1.736740$.

Then $p \prec \wp$ and all the above bounds are sharp.

Proof. Let $\theta(w) = 1$ and $\phi(w) = \beta/\omega \neq 0$. To prove our results, we make use of the following inequality:

$$\wp(-1) \leq q_\beta(-1) \leq q_\beta(1) \leq \wp(1) \tag{5.5}$$

so that $q_\beta \prec \wp$. Observe that (5.5) gives necessary as well as sufficient condition for the required subordination.

(*i*) The differential equation $\Phi_\beta(z, p(z)) = \phi_0(z)$ has an analytic solution given by

$$q_\beta(z) = \exp\left(-(z + 2k\log(1 - z/k))/\beta k\right). \tag{5.6}$$

Since the function

$$Q(z) = zq'_\beta(z)\phi(q_\beta(z)) = \phi_0(z) - 1 = \frac{z}{k}\left(\frac{k+z}{k-z}\right) \in \mathcal{S}^*$$

and $h(z) = \theta(q_\beta(z)) + Q(z)$ satisfies $\Re(zh'(z)/Q(z)) = \Re(zQ'(z)/Q(z)) > 0$. Therefore, by Lemma 5.1, the following differential subordination implication holds:

$$\Phi_\beta(z, p(z)) \prec 1 + \beta\frac{zq'_\beta(z)}{q_\beta(z)} \Rightarrow p \prec q_\beta.$$

Now to show that $p \prec \wp$, we need to show that $q_\beta \prec \wp$. Thus, (5.5) gives necessary condtion for the subordination $p \prec \wp$ to hold. That is, $\beta \geq \beta_1$ and $\beta \geq \beta_2$, respectively, where

$$\beta_1 = \frac{-1 + 2k\log(1 + 1/k)}{k\log(e/(e-1))} \quad \text{and} \quad \beta_2 = \frac{-1 + 2k\log(k/(k-1))}{k\log(1+e)}.$$

Let $\beta \geq \beta_1$. Also for simplicity, we write the function given in (5.6) as follows for $z = e^{i\theta}$:

$$q_\beta(z) := \exp(u + iv),$$

where u and v are function of $\theta \in [-\pi, \pi]$ given as

$$u(\theta) := \left(\frac{-1}{\beta k}\right)\left(\cos\theta + 2k\log\sqrt{1 + \frac{1}{k^2} - \frac{2\cos\theta}{k}}\right)$$

and

$$v(\theta) := \left(\frac{-1}{\beta k}\right)\left(\sin\theta + 2k\arctan\left(\frac{\sin\theta}{\cos\theta - k}\right)\right).$$

Further, without loss of generality, we may assume that $\beta = \beta_1$. Now the difference of the square of the distances from the point $(1, 0)$ to the points on the boundary curves $\wp(e^{i\theta})$ and $q_\beta(e^{i\theta})$, respectively is given by

$$d(\theta) := e^{2\cos\theta} - 1 - e^{2u(\theta)} + 2e^{u(\theta)}\cos(v(\theta)).$$

Since $d(\theta) = d(-\theta)$. So we may consider the interval $0 \leq \theta \leq \pi$. Now we need to prove that $d(\theta) \geq 0$ and so we consider

$$d'(\theta) = -e^{2\cos\theta}\sin\theta - 2e^{u(\theta)}(e^{u(\theta)}u'(\theta) + u'(\theta)\cos(v(\theta)) - \sin(v(\theta))v'(\theta)).$$

It is easy to see that $d'(\theta) = 0$ for $\theta = 0$ and π. Further a numerical computation also shows that $d'(\theta) = 0$ for a unique $\theta_0 \in (0, \pi)$ such that $d(\theta)$ increases on $[0, \theta_0]$ and decreases on $[\theta_0, \pi]$, where $\theta_0 \approx 0.351807$. Note that $d(0) > 0$ and $d(\pi) = 0$. Therefore,

$$\min_{0 \leq \theta \leq \pi} d(\theta) = 0.$$

Hence $\beta \geq \beta_1$ is also sufficient for $q_\beta \prec \wp$ and this completes the proof. Similarly we can prove the other results. □

Let $f \in \mathcal{A}$ and set $p(z) = zf'(z)/f(z)$ and $p(z) = f(z)/z$. Then using Theorem 5.7, we obtain sufficient conditions for f to be in $\mathcal{S}^*_\wp$.

Corollary 5.8. *Let $f \in \mathcal{A}$ and set*

$$\Phi_\beta\left(z, \frac{zf'(z)}{f(z)}\right) = 1 + \beta\left(1 - \frac{zf'(z)}{f(z)} + \frac{zf''(z)}{f(z)}\right).$$

If $\Phi_\beta\left(z, \frac{zf'(z)}{f(z)}\right)$ satisfies any of the conditions (i)-(vii) of Theorem 5.7, then $f \in \mathcal{S}^*_\wp$.

Corollary 5.9. *Let $f \in \mathcal{A}$ and set*

$$\Phi_\beta\left(z, \frac{f(z)}{z}\right) = \beta\frac{zf'(z)}{f(z)} + (1 - \beta).$$

If $\Phi_\beta\left(z, \frac{f(z)}{z}\right)$ satisfies any of the conditions (i)-(vii) of Theorem 5.7, then $f \in \mathcal{S}^*_\wp$.

The proof of the following result is omitted here as it is much akin to the proof of Theorem 5.4 and 5.7.

Theorem 5.10. *Assume that* $0 < \alpha < 1$, $0 < B < A < 1$, $k = 1 + \sqrt{2}$ *and* $\wp(z) = 1 + ze^z$. *Let* $p \in \mathcal{A}_0$. *Set*

$$\mho_\beta(z, p(z)) = 1 + \beta \frac{zp'(z)}{p^2(z)}.$$

If p *satisfies any of the following differential subordinations:*

(*i*) $\mho_\beta(z, p(z)) \prec \phi_0(z)$ *for* $\beta \geq -((e+1)/ke)(1 + 2k\log(1 - 1/k)) \approx 0.896489$.
(*ii*) $\mho_\beta(z, p(z)) \prec \sqrt{1+z}$ *for* $\beta \geq 2(e-1)(1 - \log 2) \approx 1.05451$.
(*iii*) $\mho_\beta(z, p(z)) \prec \varphi_\alpha(z)$ *for* $\beta \geq ((e-1)/2\sqrt{\alpha}) \log(\frac{1+\sqrt{\alpha}}{1-\sqrt{\alpha}})$.
(*iv*) $\mho_\beta(z, p(z)) \prec 1 + \sin z$ *for* $\beta \geq (e-1)\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!(2n+1)} \approx 1.62563$.
(*v*) $\mho_\beta(z, p(z)) \prec e^z$ *for* $\beta \geq ((e+1)/e)\sum_{n=1}^{\infty} \frac{1}{n!n} \approx 1.80273$.

Then $p \prec \wp$ *and the bounds are sharp.*

Let $f \in \mathcal{A}$ and set $p(z) = zf'(z)/f(z)$ and $p(z) = f(z)/z$. Then using Theorem 5.10, we obtain the following results:

Corollary 5.11. *Let* $f \in \mathcal{A}$ *and set*

$$\mho_\beta\left(z, \frac{zf'(z)}{f(z)}\right) = 1 + \beta\left(\frac{zf'(z)}{f(z)}\right)^{-1}\left(1 - \frac{zf'(z)}{f(z)} + \frac{zf''(z)}{f(z)}\right).$$

If $\mho_\beta\left(z, \frac{zf'(z)}{f(z)}\right)$ *satisfies any of the conditions (i)-(v) of the Theorem 5.10, then* $f \in \mathcal{S}^*_\wp$.

Corollary 5.12. *Let* $f \in \mathcal{A}$ *and set*

$$\mho_\beta\left(z, \frac{f(z)}{z}\right) = 1 + \beta\left(\frac{z}{f(z)}\left(\frac{zf'(z)}{f(z)} - 1\right)\right).$$

If $\mho_\beta\left(z, \frac{f(z)}{z}\right)$ *satisfies any of the conditions (i)-(v) of the Theorem 5.10, then* $f \in \mathcal{S}^*_\wp$.

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] Aktaş İ, Baricz Á and Orhan H. Bounds for radii of starlikeness and convexity of some special functions. Turkish J Math, 2018, **42**(1): 211–226

[2] Ali R M, Ravichandran V and Jain N K. Convolution of certain analytic functions. J Anal, 2010, **18**: 1–8

[3] Bansal D and Prajapat J K. Certain geometric properties of the Mittag-Leffler functions. Complex Variables and Elliptic Equations, 2016, **61**(3): 338–350

[4] Banga S and Kumar S S. Applications of differential subordinations to certain classes of starlike functions. J Korean Math Soc, 2020, **57**(2): 331–357

[5] Baricz Á, Dimitrov D K, Orhan H, and Yağmur N. Radii of starlikeness of some special functions. Proc Amer Math Soc, 2016, **144**(8): 3355–3367

[6] Baricz Á, Toklu E and Kadioğlu E. Radii of starlikeness and convexity of Wright functions. Math Commun, 2018, **23**(1): 97–117

[7] Biernacki M and Krzyż J. On the monotonity of certain functionals in the theory of analytic functions. Ann Univ Mariae Curie-Skłodowska Sect A,1957, **9**(1955): 135–147

[8] Bulut S and Engel O. The radius of starlikeness, convexity and uniform convexity of the Legendre polynomials of odd degree. Results Math, 2019, **74**(1), Art 48, 9 pp.

[9] Cho N E, Kumar S, Kumar v and Ravichandran V. Convolution and radius problems of analytic functions associated with the tilted Carathéodory functions. Math Commun, 2019, **24**(2): 165–179

[10] Cho N E, Kumar V, Kumar S S and Ravichandran V. Radius problems for starlike functions associated with the sine function. Bull Iranian Math Soc, 2019, **45**(1): 213–232

[11] Gangadharan A, Ravichandran V and Shanmugam T N. Radii of convexity and strong starlikeness for some classes of analytic functions. J Math Anal Appl, 1997, **211**(1): 301–313

[12] Goel P and Kumar S S. Certain Class of Starlike Functions Associated with Modified Sigmoid Function. Bull Malays Math Sci Soc, 2020, **43**(1): 957–991

[13] Goluzin G M. On a variational method in the theory of analytic functions. Amer Math Soc Transl, 1961, **18**(2): 1–14

[14] Kargar R, Ebadian A and Sokół J. Radius problems for some subclasses of analytic functions. Complex Anal Oper Theory, 2017, **11**(7): 1639–1649

[15] Kumar V, Cho N E, Ravichandran V and Srivastava H M. Sharp coefficient bounds for starlike functions associated with the Bell numbers. Math Slovaca, 2019, **69**(5): 1053–1064

[16] Kumar S S and Kamaljeet G. A Cardioid Domain and Starlike Functions. Analysis and Mathematical Physics, 2020 (Communicated)

[17] Kumar S and Ravichandran V. A subclass of starlike functions associated with a rational function. Southeast Asian Bull Math, 2016, **40**(2): 199–212

[18] Kuroki K, Owa S. Notes on new class for certain analytic functions. RIMS Kokyuroku, 2011, **1772**: 21–25

[19] Ma W C and Minda D. A unified treatment of some special classes of univalent functions. *Proceedings of the Conference on Complex Analysis, Tianjin*, Conf Proc Lecture Notes Anal, I Int Press, Cambridge, MA, 1992: 157–169

[20] MacGregor T H. Majorization by univalent functions. Duke Math J, 1967, **34**: 95–102

[21] MacGregor T H. Hull subordination and extremal problems for starlike and spirallike mappings. Trans Amer Math Soc, 1973, **183**: 499–510

[22] Mendiratta R, Nagpal S and Ravichandran V. On a subclass of strongly starlike functions associated with exponential function. Bull Malays Math Sci Soc, 2015, **38**(1): 365–386

[23] Miller S S and Mocanu P T. *Differential subordinations*, Monographs and Textbooks in Pure and Applied Mathematics. New York: Marcel Dekker, Inc, 2000

[24] Raina R K and Sokół J. Some properties related to a certain class of starlike functions. C R Math Acad Sci Paris, 2015, **353**(11): 973–978

[25] Ruscheweyh S and Sheil-Small T. Hadamard products of Schlicht functions and the Pólya-Schoenberg conjecture. Comment Math Helv, 1973, **48**: 119–135

[26] Sharma K, Jain N K and Ravichandran V. Starlike functions associated with a cardioid. Afr Mat, 2016, **27**(5-6): 923–939

[27] Sokół J and Stankiewicz J. Radius of convexity of some subclasses of strongly starlike functions. Zeszyty Nauk Politech Rzeszowskiej Mat, 1996, (19): 101–105

[28] Tang H and Deng G. Majorization problems for some subclasses of starlike functions. J Math Res Appl, 2019, **39**(2): 153–159.

[29] Wang L M. The tilted Carathéodory class and its applications. J Korean Math Soc, 2012, **49**: 671–686

Department of Applied Mathematics, Delhi Technological University, Delhi–110042, India

E-mail address: spkumar@dce.ac.in

Department of Applied Mathematics, Delhi Technological University, Delhi–110042, India

E-mail address: gangania.m1991@gmail.com

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3007079, IEEE Transactions on Industrial Electronics

Suppression of Lower Order Harmonics for the full modulation range for a 2-level inverter fed IM drive with a switched capacitive filter technique forming a 42-sided Voltage Space Vector Structure

Rahul Dewani, *Student Member, IEEE*, Rakesh R, *Student Member, IEEE*,
K. Gopakumar, *Fellow, IEEE*, Umanand Loganathan
Dariusz Zielinski, *Member, IEEE*, Leopoldo G. Franquelo, *Fellow, IEEE*

***Abstract*—The work proposes an inverter topology for the first time, for suppression of all lower order harmonics (5^{th}, 7^{th}, 11^{th}, 13^{th}, ...) up to 39^{th} harmonic. The lower order harmonic suppression is achieved by switching the 42-sided polygonal vectors, which eliminates harmonics in phase voltage without using passive LC filtering technique. The power circuit topology uses a single DC link supply for primary inverter. The secondary inverter is fed with a capacitive supply. A novel topology is proposed for the dense 6-level secondary inverter. The lower order harmonics in phase voltage are suppressed using secondary inverter, which acts like a switched-capacitive filter. The work also proposes a novel capacitor voltage balancing scheme for the first time, which achieves voltage balancing for secondary inverter capacitors. The voltages of secondary inverter capacitors are balanced using the power balance relationship between primary inverter and motor load. An experimental study of the topology and capacitor balancing scheme is presented for the validation of proposed scheme.**

***Index Terms*—Harmonic suppression, Polygonal space vector structure, Induction Motor drives, Transformer-less Filtering, Lower Order Harmonic Elimination, Pulse width modulation (PWM).**

I. Introduction

INDUCTION MOTORS have emerged as the preferred solution for medium to high power industrial applications. The induction motor can be driven with a 3-phase, 2-level inverter, fed with active DC link supply for variable speed applications. The 2-level inverter generates a hexagonal Space Vector (SV) structure which has lower linear modulation range of operation. When the inverter is driven in over-modulation range, lower order harmonics in phase voltage, cause undesirable lower order harmonic torque pulsations. The lower order torque ripple is not suitable for industrial applications, and can even lead to mechanical breakdown of motor shaft [1], [2].

The lower order torque harmonics can be eliminated by suppressing lower order harmonics in phase voltage. The lower order harmonics can be attenuated by using a passive LC filter [3], [4]. Since the phase voltage has 5^{th} and 7^{th} order harmonics in over-modulation, the cut-off frequency of LC filter, must be lower than $5.\omega_o$ (ω_o is the fundamental frequency). An LC filter of such a low cut-off frequency will be very bulky and expensive. The LC filter also limits the bandwidth of the speed control loop, degrading transient performance. Also, in linear modulation range the 2-level inverter needs to be switched at high frequencies to suppress lower order harmonics resulting in higher switching losses. Hence, this solution for harmonic suppression has many drawbacks in implementation.

Another technique used for suppression of lower order harmonics in phase voltage is Selective Harmonic Elimination(SHE) [5]–[8]. The technique is based on mathematical computations to find the optimum PWM switching sequence for the inverter operation. The resulting phase voltage will be devoid of the harmonics for which the PWM scheme has been designed. The technique lacks intuitive insight, as it involves solving an additional non-linear equation for each harmonic being eliminated. Hence, the non-linear equations are precomputed and optimal PWM sequence is stored in look-up tables. The method is limited in the number of lower order harmonics that can be suppressed, as the number of non-linear equations to be solved keeps increasing with the number of harmonics that need to be suppressed.

The multi-level topologies proposed in literature [9]–[12], achieve harmonic suppression in linear range of operation. Since the closest levels are time-averaged to generate the reference pole voltage, the switching ripple is reduced in the phase voltage. While this method is useful for harmonic suppression in linear range of operation, in general, multi-level topologies generate lower order harmonics in over-modulation range of operation. This imposes a limitation on the performance that can be achieved with multi-level inverters.

The topology proposed in this work overcomes the draw-

Manuscript received March 14, 2020; revised May 24, 2020; accepted June 21, 2020.

Rahul Dewani, Rakesh R, K. Gopakumar and Umanand Loganathan are with the Department of Electronic Systems Engineering,Indian Institute of Science,Bangalore-560012,India.(email: rahuldewani@iisc.ac.in; rrakesh225@gmail.com; kgopa@iisc.ac.in; lums@iisc.ac.in).

Dariusz Zielinski is with the Department of Electrical Drives and Machines, Lublin University of Technology, Lublin, Poland. (email: d.zielinski@pollub.pl).

Leopoldo G. Franquelo is with the Electronic Engineering Department, University of Seville, 41092, Spain. (email: lgfranquelo@ieee.org).

0278-0046 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 06:36:29 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3007079, IEEE Transactions on Industrial Electronics

backs of other works in literature, by rejecting lower order harmonics in phase voltage without compromising on the bandwidth of speed control loop. The harmonic rejection scheme is extended over the entire range of linear modulation up to the base speed, without the need of complex computations. The power circuit topology proposed in literature [13], for polygonal structures have reported the necessity of multiple DC link sources. In the work presented in this paper, a harmonic elimination scheme with a single DC link supply (feeding primary inverter) is proposed. A secondary inverter fed with capacitive supply is used to suppress lower order harmonics in phase voltage.

The primary and secondary inverters are feeding an Open End IM (OEIM) from opposite ends in this scheme. The primary inverter delivers the active power for motor operation. The secondary inverter delivers only reactive power for harmonic elimination, hence a capacitive supply can be used for secondary inverter [14], [15]. A new power circuit topology for secondary inverter is proposed in this work for the first time, which generates a 6-level Space Vector Structure (SVS). The primary and secondary space vectors are superimposed to generate a 42-sided polygonal SVS.

The secondary inverter capacitor voltages are balanced using a novel capacitor voltage balancing scheme for the first time, which does not compromise on suppression of lower order harmonics. The works proposed in literature for balancing of secondary inverter [15], [16], achieve capacitor balancing by modifying the timings of the vectors being switched, by a factor referred as K_{ratio} correction factor in literature. This brings about an error in the reference vector being generated for motor operation, which can cause lower order harmonics in phase voltage. This work overcomes the drawbacks for the first time, by introducing a novel capacitor voltage balancing scheme. The capacitor balancing is achieved by controlling the power balance of primary inverter and motor load. The capacitor voltages are controlled by modifying the radius of the 42-sided polygon, by small percentage, so as to achieve power balance.

II. Inverter Topology

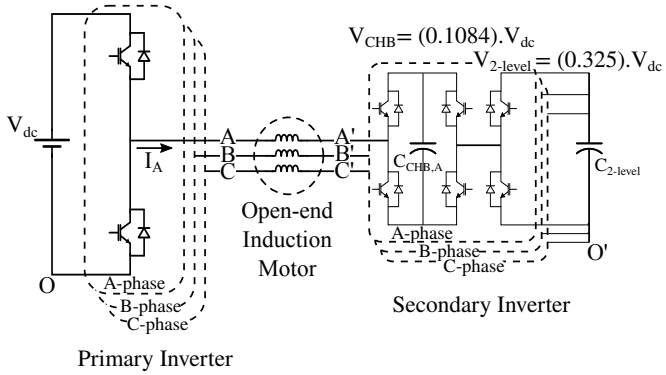


Fig. 1. Proposed Power Circuit Topology

The power circuit topology proposed in this work, is shown in Fig.1. The primary 2-level inverter fed by active DC link supply, provides the active power for motor operation. The secondary inverter, feeding the other end of OEIM is used for switched-capacitive filtering of lower order harmonics in motor phase voltage. The secondary inverter consists of a 2-level inverter fed with capacitive supply, followed by Cascaded H-Bridge (CHB) stages in the three phases.

A dense 6-level SV structure with reduced component count is generated by using asymmetric voltages of $3xV_{dc}$ and xV_{dc} for 2-level and CHB stages, where x is a constant ratio. The secondary 2-level stage generates a 2-level SV structure of radius $3xV_{dc}$, and the CHB stage generates a 3-level SV structure of radius $2xV_{dc}$. The combination of the 2-level and 3-level SV structures results in a dense 6-level SV structure of radius $5xV_{dc}$ for secondary inverter. The secondary structure is superimposed on the primary SV structure to generate the resultant 42-sided vectors, as shown in Fig.2. The figure also shows the 42-sided polygonal vectors (OP_1, OP_2, ...). The 42-sided vectors are generated by superposition of primary and secondary vectors. For example, the 42-sided vector OP_4 is generated by switching $V_{pri,1}$ vector from primary inverter and $O'P_4$ from secondary inverter. The vector $O'P_4(= 0.4374V_{dc}\angle 108.5^o)$ can be generated from the secondary SV structure, if the vector $O'P_4$ is inside the secondary 6-level SV structure of radius $5xV_{dc}$. The ratio x is designed to be 0.1084 so that $O'P_4$, can be generated by the secondary SV structure. The 2-level and CHB capacitors are maintained at voltages of $0.325V_{dc}$ and $0.1084V_{dc}$ for 42-sided polygonal switching.

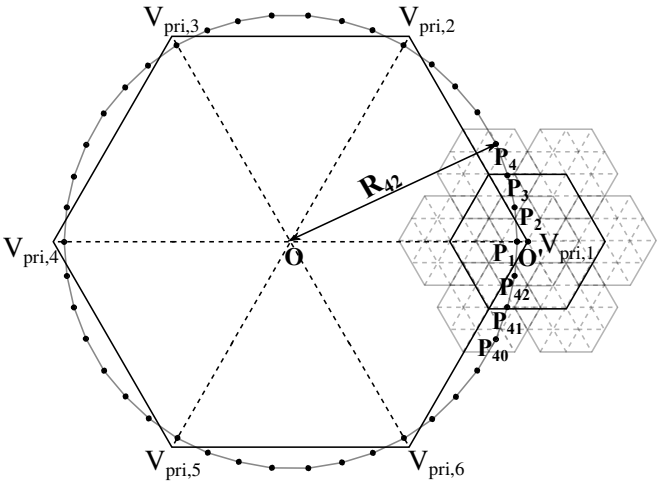


Fig. 2. SV structure of secondary inverter, superimposed on primary vector $\mathrm{V_{pri,1}}$, along with 42-sided vectors.

For power balance of secondary inverter capacitors, the power drawn form primary inverter should be same as the power delivered to the motor load. The power balance is maintained, if the fundamental phase voltage generated by the primary inverter, is equal to the fundamental phase voltage generated by switching the vectors of the 42-sided polygon. Using this condition, at the extreme case of over-modulation (where primary 2-level inverter is in six step mode of operation), the radius of 42-sided polygon can be determined. The fundamental component of phase voltage in six step mode of operation, $V_{pri,fund}$ for a hexagonal SV structure of radius V_{dc} is given by,

$$V_{pri,fund} = 0.6366V_{dc} \tag{1}$$

0278-0046 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 06:36:29 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3007079, IEEE Transactions on Industrial Electronics

Similarly, for a 42-sided polygon of radius R_{42} operating in 42-stepped mode of operation, $V^*_{pri,fund}$, the fundamental component of phase voltage, can be determined to be,

$$V^*_{pri,fund} = 0.6660R_{42} \tag{2}$$

From the power balance condition, the fundamental phase voltage should be same in the two cases,

$$\begin{aligned} &V^*_{pri,fund} = V_{pri,fund} \\ \implies &R_{42} = 0.956V_{dc} \end{aligned} \tag{3}$$

Where, R_{42} is the nominal radius of the 42-sided polygonal vectors that is necessary to maintain the active power balance.

III. Capacitor Balancing strategy

Based on the explanation in the previous section, the radius of the 42-sided polygonal reference vectors is maintained at $0.956V_{dc}$ (Equation-(3)). The PWM switching pattern for primary and secondary inverters can be obtained by assuming the secondary inverter capacitors are balanced to their nominal voltages, shown in Fig.1. For the 42-sided PWM scheme, the real power will flow from primary inverter to motor load. Secondary inverter only provides reactive power, therefore, it can be fed with capacitive supply maintained at a nominal voltage using the capacitor balancing scheme proposed next.

Assume that the secondary inverter CHB capacitors are maintained at the nominal voltage of $0.1084V_{dc}$, as shown in Fig.1. The secondary inverter 2-level capacitor can get discharged due to circuit non-idealities (or overcharged, overcharging is discussed later). To balance the 2-level capacitor voltage, it needs to be charged back to it's nominal voltage by sourcing active power to secondary inverter. To source power to secondary inverter, a fraction of active power drawn from primary inverter is diverted to secondary inverter as explained next. To source active power to secondary inverter, the radius of 42-sided polygonal vectors (OP_1, OP_2, OP_3,.., OP_{42}) is reduced slightly from the nominal value of $0.956V_{dc}$, by changing the PWM of the secondary inverter alone. By reducing the radius of 42-sided polygonal reference vectors, the phase voltage of motor load is reduced momentarily. As a result, the power going to motor load is reduced momentarily. During this process, the PWM pattern of primary inverter is not disturbed. Since the primary inverter PWM is not disturbed, the power drawn from primary inverter does not change. Thus, the power drawn from primary inverter is maintained constant, while the power delivered to motor load is reduced. Since, the power drawn from primary inverter is now greater than the power drawn by motor load, the excess power is diverted to secondary inverter 2-level capacitor momentarily. The power diverted to secondary inverter increases the energy stored in 2-level capacitor, by charging the 2-level capacitor voltage back to it's nominal value.

The secondary 2-level capacitor may get overcharged due to circuit non-idealities. To balance the 2-level capacitor voltage, it needs to be discharged back to it's nominal voltage by drawing active power from secondary inverter. To draw power from secondary inverter, the active power delivered to motor load is increased momentarily by increasing motor phase voltage as explained next. To draw active power from secondary inverter, the radius of 42-sided polygonal vectors is increased from the nominal value of $0.956V_{dc}$, by changing the PWM of the secondary inverter alone. By increasing the radius of 42-sided polygonal reference vectors, the phase voltage of motor load is increased momentarily. As a result, the power going to motor load is increased momentarily. During this process, the PWM pattern of primary inverter is not disturbed. Since the primary inverter PWM is not disturbed, the power drawn from primary inverter does not change. Thus, the power drawn from primary inverter is maintained constant, while the power delivered to motor load is increased. Since, the power drawn from primary inverter is now lesser than the power drawn by motor load, power is drawn from the secondary inverter 2-level capacitor momentarily. The power drawn from secondary inverter decreases the energy stored in 2-level capacitor, by discharging the 2-level capacitor voltage back to it's nominal value. Thus, the above mentioned control scheme is used to maintain the 2-level capacitor voltage within $\pm 5\%$ of $0.325V_{dc}$.

In the balancing scheme for 2-level secondary capacitor discussed so far, the CHB capacitors of secondary inverter are assumed to be balanced at their nominal voltages. In practice, they may deviate from their nominal voltages. A fast switching cycle capacitor voltage balancing scheme is proposed for CHB stage capacitors alone in sub-section.III-B.

A. Voltage Balancing of 2-level stage capacitor

Assuming that the CHB stage capacitor voltages are balanced at their nominal value, there may be a momentary disturbance in the secondary 2-level capacitor voltage due to circuit non-idealities. The 2-level capacitor needs to be balanced by diverting power appropriately to secondary inverter. The PWM of secondary inverter has to be modified in such a way that, either a fraction of real power from primary inverter is given to secondary inverter, or real power is drawn from secondary inverter and given to motor load. During this process the PWM pattern of primary inverter is not disturbed.

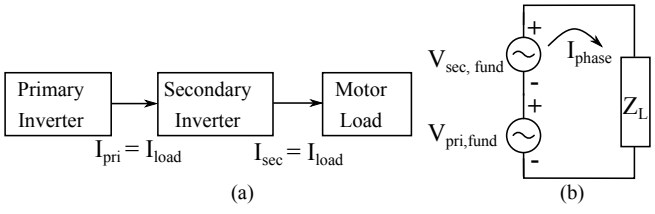


Fig. 3. **(a)** Block diagram of power circuit. **(b)** Single phase equivalent of power circuit.

The primary and secondary phase voltages are generated by PWM switching pattern of the inverters. The currents for primary and secondary inverters is same as motor phase current (Fig.3(a)). Thus, a single phase equivalent model can be visualized as shown in Fig.3(b). The primary and secondary fundamental phase voltages can be visualized as voltage sources in series. If the phase voltage of load impedance (motor load) is maintained to be at phase voltage of primary inverter, then secondary inverter does not contribute any fundamental phase voltage. If the fundamental phase voltage of secondary inverter is maintained to be zero, it does not contribute any real power.

0278-0046 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 06:36:29 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3007079, IEEE Transactions on Industrial Electronics

To balance the secondary 2-level capacitor voltage, the radius of 42-sided vectors is changed, by changing the switching pattern of secondary inverter alone. By changing the radius of 42-sided vectors from the nominal value, fundamental phase voltage of motor load is changed. Since, the PWM of primary inverter is not disturbed, the phase voltage of primary inverter remains same. If the motor phase voltage is increased, the secondary inverter generates positive phase voltage and delivers real power to motor load. Whereas, if the motor phase voltage is reduced, the secondary inverter generates negative phase voltage and absorbs a fraction of the real power given by primary inverter (refer Fig.3).

The real power drawn from primary inverter is given by,

$$P_{pri} = 3/2 * V_{pri,fund} * I_{phase} * cos(\theta) \tag{4}$$

Where, $V_{pri,fund}$ and I_{phase}, represent the amplitudes of fundamental components of phase voltage and current of the primary inverter, respectively, and $cos(\theta)$ represents the power factor. Similarly, the real power delivered to the motor can be expressed as,

$$P_{motor} = 3/2 * V_{motor,fund} * I_{phase} * cos(\theta) \tag{5}$$

Where, $V_{motor,fund}$ represents the amplitude of fundamental component of motor phase voltage. In steady state, the power drawn from the primary inverter, equals the power delivered to the motor load. During transients, if the power drawn from the primary inverter is more than the power delivered to the motor load, the excess power will be stored in the secondary inverter capacitors. This relation can be expressed as,

$$\begin{aligned} P_{pri} &= P_{motor} + P_{sec} \\ \implies P_{sec} &= P_{pri} - P_{motor} \end{aligned} \tag{6}$$

Where, P_{sec} is the real power going into secondary inverter. Substituting P_{pri} and P_{motor} from (4) and (5), in (6),

$$P_{sec} = 3/2 * I_{phase} * cos(\theta) * [V_{pri,fund} - V_{motor,fund}]$$

Note that, $cos(\theta)$ and I_{phase} are the Power Factor and amplitude of phase current, hence the two are always positive. Thus, the power realtion can be expressed as,

$$P_{sec} = K * [V_{pri,fund} - V_{motor,fund}]$$

Where, $K = 3/2 * I_{phase} * cos(\theta)$, is always positive for motoring operation. For $K > 0$, the above relation can be expressed as,

$$\begin{aligned} &if, \quad V_{pri,fund} > V_{motor,fund} \\ &then, V_{pri,fund} - V_{motor,fund} > 0 \\ &\implies P_{sec} > 0 \end{aligned} \tag{7}$$

$$\begin{aligned} &elseif, V_{pri,fund} < V_{motor,fund} \\ &then, \ V_{pri,fund} - V_{motor,fund} < 0 \\ &\implies \ P_{sec} < 0 \end{aligned} \tag{8}$$

Using the above relations, the real power of secondary inverter, P_{sec}, can be controlled by changing the fundamental component of motor phase voltage alone, without disturbing $V_{pri,fund}$. The fundamental motor phase voltage is controlled by changing the radius of 42-sided vectors.

Since the secondary inverter has a capacitive supply, when real power is supplied to secondary inverter, the capacitor voltage changes to store the real power in the capacitor. So, if the capacitor voltage deviates from it's nominal value, the fundamental motor phase voltage is momentarily changed to bring the capacitor voltage back to it's nominal value. The control scheme for voltage balancing of 2-level capacitor is summarized as follows:

1) Sense the voltage of 2-level capacitor of secondary inverter and compare the voltage with the desired nominal value. The sensed voltage may be greater (or lesser) than the nominal voltage due to momentary disturbances caused by circuit non-idealities. Initially, the primary inverter phase voltage and motor phase voltage are equal.
$$\begin{aligned} V_{motor,fund} &= V_{pri,fund} \\ \implies P_{sec} &= 0 \end{aligned}$$
2) If the sensed voltage is greater (or lesser), then active power should be drawn (or supplied) from the secondary inverter. The motor phase voltage is increased (or decreased) to $V^*_{motor,fund}$ by changing the PWM of secondary inverter alone, so that the power drawn by motor load exceeds the power supplied by primary inverter. As a result, the secondary inverter delivers active power to motor load.
$$\begin{aligned} V^*_{motor,fund} &= V_{pri,fund}.(1 + \alpha\%) \\ \implies V_{pri,fund} &< V^*_{motor,fund} \\ \implies P^*_{sec} &< 0 \end{aligned}$$
Where, α is less than 4%. Real power, P^*_{sec}, is drawn from the secondary inverter due to control action. As a result, the overcharged 2-level capacitor starts discharging.
3) When the secondary capacitor voltage is close to its nominal value, the motor phase voltage is reduced back to $V_{pri,fund}$, such that,
$$\begin{aligned} V_{motor,fund} &= V_{pri,fund} \\ \implies P_{sec} &= 0 \end{aligned}$$

B. Voltage Balancing of CHB stage capacitors

The CHB stage capacitor voltages are balanced within every switching cycle. The fast charge balancing for CHB stage capacitor reduces the capacitor size and voltage ripple. To achieve fast capacitor voltage balancing, SV redundancies for every vector in secondary inverter SV structure are identified.

TABLE I
SWITCHING STATE CODES CORRESPONDING TO SECONDARY INVERTER VOLTAGES.

Voltage	Switching State code
$+0.325 * V_{dc}$	+3
$+0.1084 * V_{dc}$	+1
0	0
$-0.1084 * V_{dc}$	-1

The CHB stages can be switched to generate voltages of $+0.1084V_{dc}$, 0 or $-0.1084V_{dc}$. The capacitor fed 2-level stage of secondary inverter can be switched to generate voltages of 0 or $+0.325V_{dc}$. The CHB stage voltages and 2-level voltages

0278-0046 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 06:36:29 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3007079, IEEE Transactions on Industrial Electronics

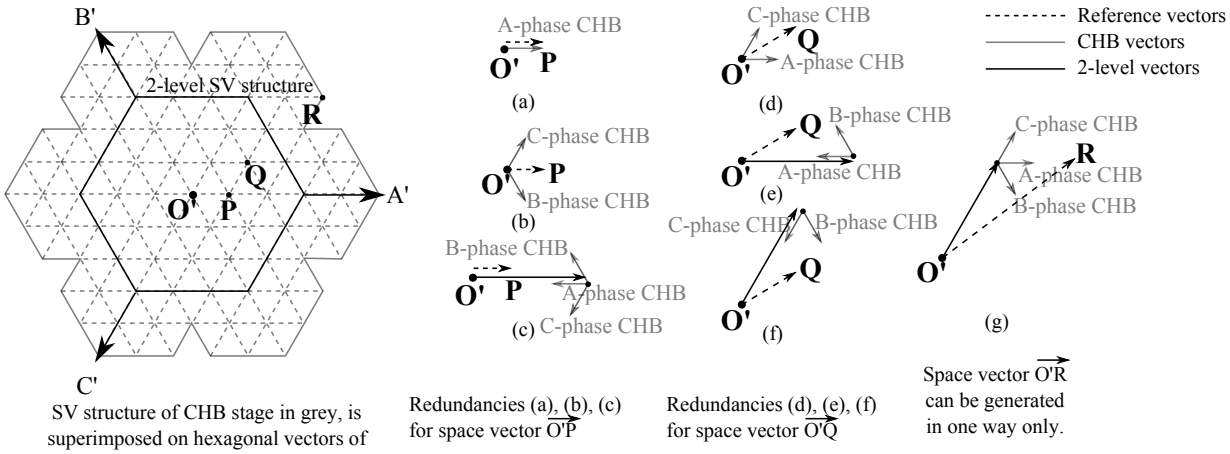


Fig. 4. Space Vector redundancies for reference vectors OP, OQ and OR.

in the three phases can be represented by the switching state codes represented in Table.I.

Consider the vector $O'P$ (Fig.4), which is to be applied for t_p time, to generate the 42-sided vectors, in a switched averaged sense. $O'P$ has three SV redundancies (a), (b), (c), shown in Fig.4. The dashed lines show the reference vector $O'P$ that should be generated by superposition of CHB stage vectors and 2-level stage vectors. The CHB stage vectors are marked by gray lines, and the 2-level stage vectors are marked by bold lines. To generate redundancy (a), state $+1$ (voltage of $0.1084V_{dc}$) is applied in A-phase of CHB stage alone, and state 0, in B and C phases. The 2-level stage applies state 0 in all 3 phases. Similarly, to generate redundancy (c), states [-1,+1,+1] are applied in the three phases CHB stage and states [+3,0,0] are applied in the three phases of 2-level stage. Since, the polarity of A-phase CHB stage is opposite in the two redundancies (+1 for redundancy (a) and -1 for redundancy (c)), one of the redundancies charges the A-phase capacitor, whereas the other redundancy discharges the capacitor. The charge balance can be maintained on A-phase capacitor, if the two redundancies (a) and (c), are applied for equal time. Similarly, for redundancy (b), the states for CHB stage and 2-level stage are [0,-1,-1] and [0,0,0]. The CHB states for redundancies (b) and (c), are opposite in B and C phases, if the two redundancies are applied for equal times, then charge balance is maintained in B and C phase capacitor voltages.

The secondary vector $O'P$ can be generated by applying any of the three redundancies. Thus, the three redundancies are applied for times t_{p1}, t_{p2} and t_{p3}, such that,

$$t_p = t_{p1} + t_{p2} + t_{p3}$$

Applying the the redundancies for times t_{p1}, t_{p2}, t_{p3}, is equivalent to applying the vector $O'P$ for time t_p. The three redundancy times are selected to be $t_p/3$, for voltage balancing of capacitors in the three phases, as explained earlier. The switching state codes in the three phases for CHB stage and 2-level stage for the three redundancies are shown in Table.II.

If the capacitor voltages are disturbed due to circuit non-idealities, the timings for each of the redundancies is slightly adjusted, so as to reduce the error in capacitor voltages.

The 2-level capacitive supply is assumed to be stiff voltage, in the discussion for CHB stage capacitor balancing, as it is maintained at the nominal voltage by power balance relations discussed earlier.

TABLE II
SPACE VECTOR REDUNDANCIES FOR VECTOR O'P

Reference Vector($O'P$) redundancy	CHB state	2-level state
(a)	[+1,0,0]	[0,0,0]
(b)	[0,-1,-1]	[0,0,0]
(c)	[-1,+1,+1]	[+3,0,0]

TABLE III
SPACE VECTOR REDUNDANCIES FOR VECTOR O'Q

Reference Vector($O'Q$) redundancy	CHB state	2-level state
(d)	[+1,0,-1]	[0,0,0]
(e)	[-1,+1,0]	[+3,0,0]
(f)	[0,-1,+1]	[+3,+3,0]

The vector $O'Q$ has three SV redundancies (d), (e), (f), shown in Fig.4. The states for CHB stage and 2-level stage are shown in Table.III. A-phase CHB capacitor is balanced by applying redundancies (d) and (e) for equal time, since they have opposite polarity for A-phase CHB state. Similarly, B-phase CHB capacitor is balanced by applying redundancies-(e) and (f) for equal time, since they have opposite polarity for B-phase CHB state. If the 3 redundancies are applied for equal times, then charge balance is maintained on all three CHB capacitors. Thus, if the vector $O'Q$ needs to be applied for t_q time, then each of the redundancies are applied for $t_q/3$ duration to balance the CHB capacitors in the 3 phases. If the capacitor voltages are disturbed due to circuit non-idealities, the timings for each of the redundancies is slightly adjusted, so as to reduce the error in capacitor voltages.

Vector $O'R$ does not have sufficient redundancies to achieve switching cycle balancing. If the vector $O'R$ is to be applied for t_r time, the capacitors will get charged or discharged for t_r time. The capacitors are balanced back to their nominal voltage during the application of next vector. It can be noted that the outer vectors like $O'R$ only need to be switched while generation of 42-sided vectors OP_{40} and OP_4, shown in Fig.2. The vectors that need to be applied to generate, other 42-sided vectors (OP_{41}, OP_{42}, OP_1, OP_2 and OP_3), shown in Fig.2, have three redundancies, so that CHB stage capacitor voltages can be balanced during there vector periods. Thus, the CHB stage capacitors can be balanced in switching cycle periods.

0278-0046 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 06:36:29 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3007079, IEEE Transactions on Industrial Electronics

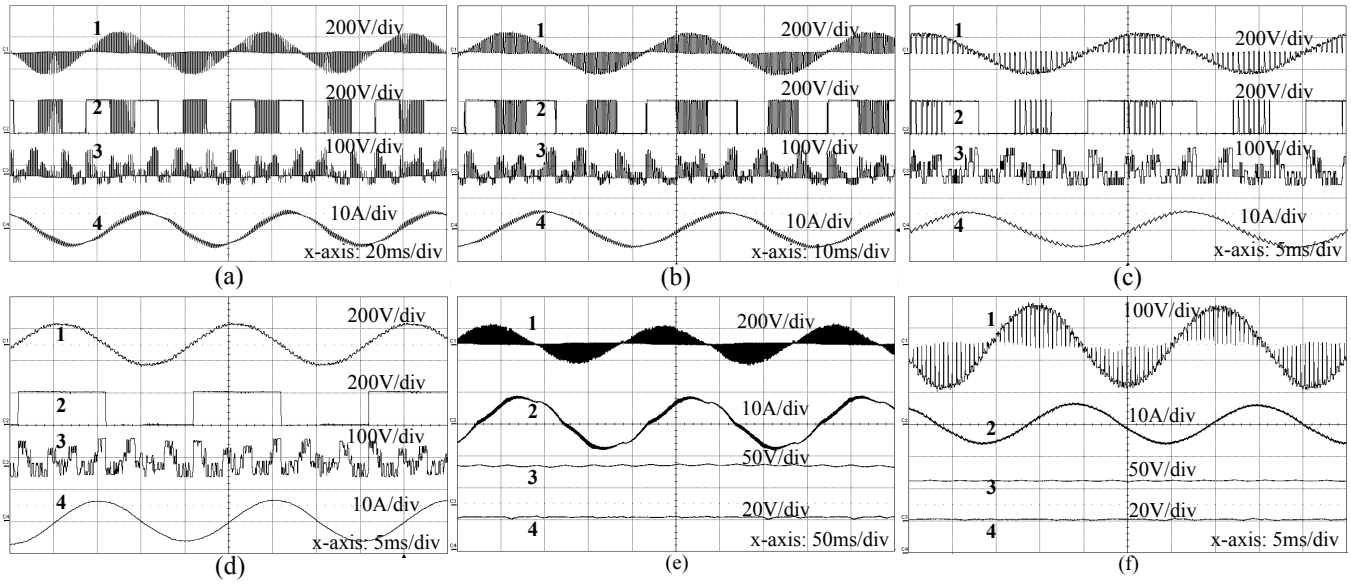


Fig. 5. Steady state Hardware results at- **(a)** 15Hz **(b)** 24Hz **(c)** 40Hz **(d)** 50Hz
Traces: **(1)** Motor phase voltage ($V_{AA'}$), **(2)** Primary pole voltage (V_{AO}), **(3)** Secondary pole voltage ($V_{A'O'}$), **(4)** Phase current (I_A).
Hardware results: Capacitor balancing results at- **(e)** 5Hz **(f)** 48Hz
Traces: **(1)** Motor phase voltage ($V_{AA'}$), **(2)** Phase current (I_A), **(3)** Secondary 2-level capacitor voltage ($V_{2-level}$), **(4)** A-phase CHB capacitor voltage ($V_{CHB,A}$).

IV. Experimental Results

TI-DSP TMS320F28335 and FPGA Spartan-XC3S200 platform are used for the hardware implementation of the proposed scheme. SKM75GB12T4 half-bridge modules are used to realize inverter switches. The 3-phase induction motor used for experiments is a 4-pole, 5kW, 200V, 50Hz Induction Motor. The block diagram of experimental setup is shown in Fig.6.

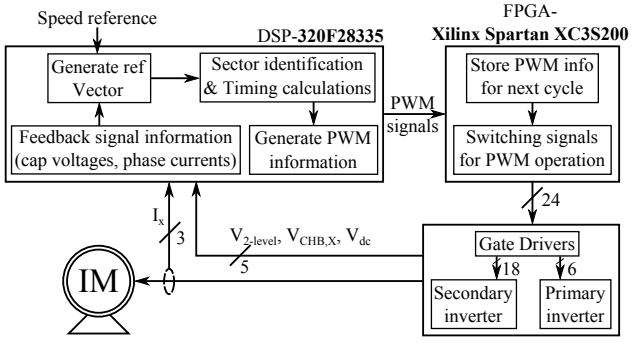


Fig. 6. Block Diagram of Experimental Setup.

The experimental results for steady state inverter operation are shown in Fig.5. The figure shows motor phase voltage, primary pole voltage, secondary pole voltage and phase current for 15Hz, 24Hz and 40Hz operation in steady state. The secondary inverter CHB capacitors and 2-level capacitor are balanced at 21.7V ($0.1084V_{dc}$) and 65V ($0.325V_{dc}$) for 200V DC link supply voltage. The secondary 6-level inverter pole voltage can vary from -21.7V to 86.7V. The envelope of phase voltage is identical for different speeds of operation. For lower speeds of operation the fundamental phase voltage is reduced, by switching the zero vector to reduce the average voltage being applied. The motor inductance further attenuates the harmonics in phase currents. Hence, the phase current is a sinusoidal signal. Fig.5(d) shows the experimental results for 50Hz operation, the extreme case of over modulation. The motor phase voltage and current are nearly sinusoidal signals. The primary inverter pole voltage is a square wave at the extreme case of over modulation. The low frequency switching on primary inverter results in lower switching losses. The secondary inverter fed with capacitive supply, is switched at higher frequencies for suppression of lower order harmonics. Since, the secondary inverter operates on voltages much lower than DC link supply, the switching losses are reduced.

Fig.5(e)-(f) show motor phase voltage, phase current, secondary 2-level capacitor voltage and A-phase CHB capacitor voltage for 5Hz and 48Hz operation in steady state. It can be observed that the capacitor voltages are balanced at their nominal voltages of 65V and 21.7V for 2-level and CHB capacitors of secondary inverter respectively.

The capacitor voltages are balanced in one switching cycle period. The capacitors are designed to have within $\pm 5\%$ deviation from nominal voltages, in the worst case. The capacitance value is determined using the relation,

$$C = \frac{I_{pk}.T_{sw}}{\Delta V_c} \tag{9}$$

Where, I_{pk}, T_{sw} and ΔV_c, are the peak of phase current, switching cycle period and deviation in capacitor voltage. For the base speed of 50Hz, the switching cycle period for 42-sided vectors is $476\mu sec$ $(= 1/(50 * 42))$. For peak current of 10A and $\pm 5\%$ voltage deviation in the worst case, the capacitance should be greater than $4.38mF$. Capacitance values of $4.4mF$ are used for experimental set up.

Fig.7 shows the FFT results for 50Hz operation, the extreme case of over modulation. The primary pole voltage generates lower order harmonics that are canceled by the secondary inverter pole voltage. The lower order harmonics are suppressed

0278-0046 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 06:36:29 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3007079, IEEE Transactions on Industrial Electronics

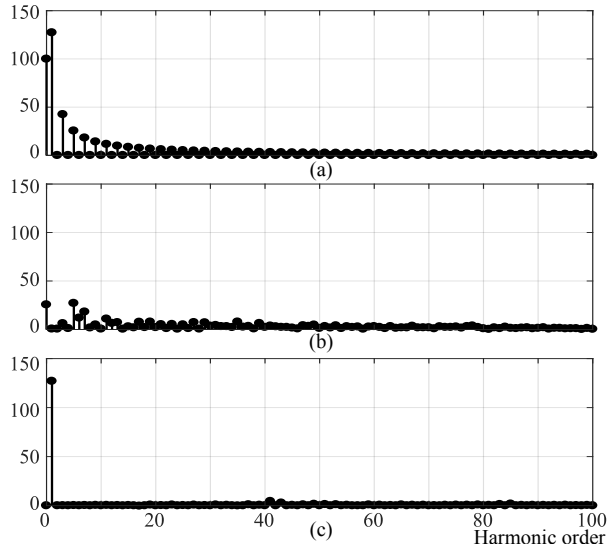


Fig. 7. FFT results at 50Hz operation. **(a)** Primary pole voltage (V_{AO}) **(b)** Secondary pole voltage ($V_{A'O'}$) **(c)** Motor phase voltage ($V_{AA'}$)

in the motor phase voltage, resulting in smooth torque profile for motor operation.

The experimental results for motor acceleration from 5Hz to 50Hz acceleration are shown in Fig.8. The waveforms for motor phase voltage, phase current, 2-level capacitor voltage and CHB A-phase capacitor voltage are shown for acceleration results. The motor phase current increases during acceleration. The capacitor voltages are balanced close to their nominal value throughout transients. Thus, the inverter topology is experimentally tested for steady state and transient operation.

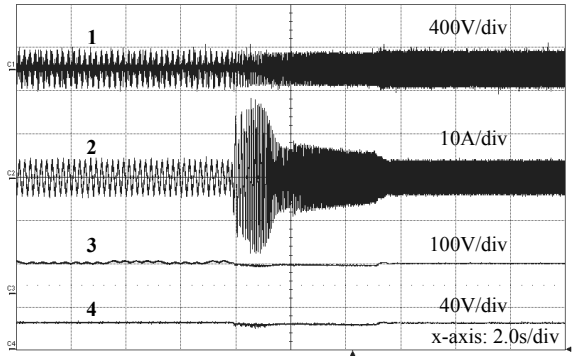


Fig. 8. Transient results: Accelaration from 5Hz to 50Hz.
Traces: **(1)** Phase voltage ($V_{AA'}$), **(2)** Phase current (I_A), **(3)** Secondary 2-level capacitor voltage ($V_{2-level}$), **(4)** A-phase CHB capacitor voltage ($V_{CHB,A}$)

Fig.9 shows the comparison of WTHD performance of a 2-level inverter and the proposed scheme. The proposed scheme achieves better harmonic performance due to suppression of lower order harmonics by secondary inverter.

V. Conclusion

A scheme for suppression of all lower order harmonics up to 39^{th} harmonic in motor phase voltage is proposed. The harmonic suppression is achieved by generating a 42-sided voltage space vector structure. The harmonics generated by the primary inverter are suppressed by using a secondary inverter, which acts like a switched-capacitive filter. The new power circuit topology proposed for secondary inverter

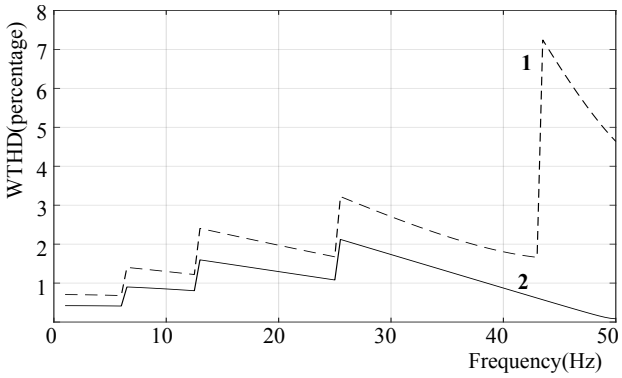


Fig. 9. Weighted Total Harmonic Distortion (WTHD) plots.
Traces: **(1)** WTHD of phase voltage for 2-level inverter **(2)** WTHD of phase voltage for proposed 42-sided polygonal SVS.

generates a dense 6-level SV structure. A switching cycle balancing scheme for CHB stage capacitors using space vector redundancies is proposed. A new capacitor balancing scheme based on power balance is proposed in this work for the first time, which balances secondary inverter 2-level capacitor voltage while suppressing all lower order harmonics up to 39^{th} harmonic. The harmonic elimination can be observed in the FFT waveforms of phase voltage. The power circuit topology proposed in the work has been studied and verified extensively in hardware. The work offers a new solution for elimination of lower order harmonics in phase voltage using minimum component count, without using conventional bulky LC filters. Here, high frequency switching is on low voltage side alone, the component size and system cost are reduced, and with the use of switched capacitive filtering, the lower order harmonics are eliminated for the full modulation range up to the base speed.

References

[1] A. Tripathi and G. Narayanan, "Evaluation and minimization of low-order harmonic torque in low-switching-frequency inverter-fed induction motor drives," *IEEE Transactions on Industry Applications*, vol. 52, no. 2, pp. 1477–1488, March 2016.

[2] J. Song-Manguelle, J. Nyobe-Yome, and G. Ekemb, "Pulsating torques in pwm multi-megawatt drives for torsional analysis of large shafts," *IEEE Transactions on Industry Applications*, vol. 46, no. 1, pp. 130–138, Jan 2010.

[3] V. Dzhankhotov and J. Pyrhönen, "Passive lc filter design considerations for motor applications," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 10, pp. 4253–4259, Oct 2013.

[4] V. Dzhankhotov, J. Pyrhonen, P. Silventoinen, M. Kuisma, and T. A. Minav, "A new passive hybrid air-core foil filter for modern power drives," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 5, pp. 1757–1762, May 2011.

[5] D. Ahmadi, K. Zou, C. Li, Y. Huang, and J. Wang, "A universal selective harmonic elimination method for high-power inverters," *IEEE Transactions on Power Electronics*, vol. 26, no. 10, pp. 2743–2752, Oct 2011.

[6] J. R. Wells, B. M. Nee, P. L. Chapman, and P. T. Krein, "Selective harmonic control: a general problem formulation and selected solutions," *IEEE Transactions on Power Electronics*, vol. 20, no. 6, pp. 1337–1345, Nov 2005.

[7] Li Li, D. Czarkowski, Yaguang Liu, and P. Pillay, "Multilevel selective harmonic elimination pwm technique in series-connected voltage inverters," *IEEE Transactions on Industry Applications*, vol. 36, no. 1, pp. 160–170, Jan 2000.

[8] H. Zhao, T. Jin, S. Wang, and L. Sun, "A real-time selective harmonic elimination based on a transient-free inner closed-loop control for cascaded multilevel inverters," *IEEE Transactions on Power Electronics*, vol. 31, no. 2, pp. 1000–1014, Feb 2016.

0278-0046 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 06:36:29 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2020.3007079, IEEE Transactions on Industrial Electronics

[9] N. Flourentzou, V. G. Agelidis, and G. D. Demetriades, "Vsc-based hvdc power transmission systems: An overview," *IEEE Transactions on Power Electronics*, vol. 24, no. 3, pp. 592–602, March 2009.
[10] H. Abu-Rub, J. Holtz, J. Rodriguez, and G. Baoming, "Medium-voltage multilevel converters—state of the art, challenges, and requirements in industrial applications," *IEEE Transactions on Industrial Electronics*, vol. 57, no. 8, pp. 2581–2596, Aug 2010.
[11] L. G. Franquelo, J. I. Leon, and E. Dominguez, "Recent advances in high-power industrial applications," in *2010 IEEE International Symposium on Industrial Electronics*, July 2010, pp. 5–10.
[12] S. Rohner, S. Bernet, M. Hiller, and R. Sommer, "Modelling, simulation and analysis of a modular multilevel converter for medium voltage applications," in *2010 IEEE International Conference on Industrial Technology*, March 2010, pp. 775–782.
[13] A. Das, K. Sivakumar, R. Ramchand, C. Patel, and K. Gopakumar, "A combination of hexagonal and 12-sided polygonal voltage space vector pwm control for im drives using cascaded two-level inverters," *IEEE Transactions on Industrial Electronics*, vol. 56, no. 5, pp. 1657–1664, May 2009.
[14] N. A. Azeez, S. Pramanick, K. Gopakumar, and C. Cecati, "A 5th and 7th order harmonic suppression scheme using capacitive filtering for 2-level vsi fed im drive," in *2014 International Symposium on Power Electronics, Electrical Drives, Automation and Motion*, June 2014, pp. 780–783.
[15] S. Pramanick, N. A. Azeez, R. Sudharshan Kaarthik, K. Gopakumar, and C. Cecati, "Low-order harmonic suppression for open-end winding im with dodecagonal space vector using a single dc-link supply," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 9, pp. 5340–5347, Sep. 2015.
[16] K. Raj R, K. Gopakumar, M. Boby, A. K. Yadav, L. G. Franquelo, and S. S. Williamson, "Multilevel 24-sided polygonal voltage-space-vector structure generation for an im drive using a single dc source," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 2, pp. 1023–1031, Feb 2019.



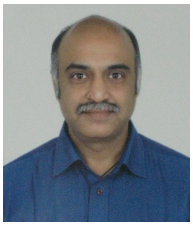
Rahul Dewani (S'19) received the B.Tech. and M.Tech. degrees from the Delhi Technological University, Delhi, India and Indian Institute of Technology, Delhi, India, in 2014 and 2016, respectively. He was with Texas Instruments, Bangalore, India in 2016, and Aura Semiconductor, Bangalore, India from 2016 to 2018. He is currently working towards the Ph.D. degree at the Department of Electronic Systems Engineering, Indian Institute of Science. His research interests include renewable energy and IM drive applications.



Rakesh R (S'18) received the B.Tech. degree in electrical and electronics engineering from the College of Engineering, Trivandrum, India, in 2014 and the M.Tech. degree in electronic systems engineering from the Indian Institute of Science, Bangalore, India, in 2017. He is currently working towards the Ph.D. degree at the Department of Electronic Systems Engineering, Indian Institute of Science. His research interests include areas of power converters and drives.



K. Gopakumar (M'94-SM'96-F'11) received the B.E., M.Sc. (Engg.), and Ph.D. degrees from the Indian Institute of Science, Bangalore, India, in 1980, 1984, and 1994, respectively. He was with the Indian Space Research Organization, Bangalore, India from 1984 to 1987. He currently holds the position of professor at the Department of Electronic Systems Engineering, Indian Institute of Science. Dr. Gopakumar is a Fellow of Institution of Electrical and Telecommunication Engineers, India and Indian National Academy of Engineers. He is currently a Co-Editor in Chief of IEEE Transaction on Industrial Electronics and also a Distinguished Lecturer of IEEE Industrial Electronics Society(IES). His research interests include PWM converters and high power drives.



Umanand Loganathan is an Associate Professor at Department of Electronics Systems Engineering (DESE) in Indian Institute of Science (IISc). He received the Bachelor's degree in electronics and communication from Bangalore University, Bangalore, India, in 1987, the M.Tech. degree in electronics design, and the Doctoral degree in the area of control of high-performance induction motor drives from the Indian Institute of Science, Bangalore, India, in 1989 and 1996 respectively. His major research interests are in photovoltaic system design, bond graph modeling of power electronic systems, high performance control of induction motor, designing for reliability and hybrid electric vehicles.



Dariusz Zielinski received his Ph.D. degree in Electrical Engineering from the Lublin University of Technology in 2017. Currently he supervises the research Laboratory for the conversion and storage of energy from renewable sources. He plays a leading role in the projects PLUGinEV (electric vehicle fast-charging system) and Active Prosumer, financed by the National Centre for Research and Development. His research interests cover power electronics, grid-tied converters, microgrids and energy storage systems. His most significant works focus on control algorithms for multi-channel power converters, bidirectional power converters, hybrid energy storage systems, grid synchronisation, island mode detection methods and DC-DC converters. He is a member of IEEE.



Leopoldo G. Franquelo (M'84–SM'96–F'05) was born in Malaga, Spain. He received the M.Sc. and Ph.D. degrees in electrical engineering from the University of Sevilla, Seville, Spain, in 1977 and 1980, respectively. His research interests include modulation techniques for multilevel inverters and its application to power electronic systems for renewable energy systems. Dr. Franquelo was a Member-at-Large of IES AdCom from 2002 to 2003, the Vice President for Conferences from 2004 to 2007, and the President Elect of the IES from 2008 to 2009. He was the recipient of a number of best paper awards from journals of the IEEE, and the Eugene Mittelmann Award and the Anthony J. Hornfeck Service Award from Industrial Electronics Society (IES) in 2012 and 2015, respectively. He has been an IES Distinguished Lecturer since 2006, an Associate Editor for the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS since 2007, the Co-Editor in Chief since 2014, and the Editor in Chief from 2016 to 2018. He was the President of the IEEE Industrial Electronics Society.

0278-0046 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 28,2020 at 06:36:29 UTC from IEEE Xplore. Restrictions apply.

Nanoscale Advances

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Check for updates

Cite this: DOI: 10.1039/d0na00309c

The prospective application of a graphene/MoS_2 heterostructure in Si-HIT solar cells for higher efficiency†

Chandra Kamal Borah,[a] Pawan K. Tyagi[bc] and Sanjeev Kumar *[a]

The efficiency of a Si-HIT (heterojunction with intrinsic thin layer) solar cell based on a graphene/MoS_2 heterostructure has been optimized by varying the various parameters of graphene (Gr) as a transparent conducting electrode (TCE) and n-type molybdenum disulfide (n-MoS_2) as an emitter layer. The photovoltaic performance of a graphene/n-MoS_2/a-Si:H/p-cSi/Au single facial HIT solar cell has been studied using AFORS-HET v2.5 simulation software. A maximum output efficiency of 25.61% has been achieved. The obtained results were compared with the results from a commercially available a-Si:H layer and p-cSi wafer after simulation. Moreover, the dependence of the cell performance on changes in the TCE and the back contact materials has also been studied. Finally, it has been demonstrated that the graphene layer and n-MoS_2 layer could act as a TCE and an efficient emitter layer, respectively, in a n-MoS_2/p-cSi based HIT solar cell.

Received 18th April 2020
Accepted 8th June 2020

DOI: 10.1039/d0na00309c

rsc.li/nanoscale-advances

Open Access Article. Published on 23 June 2020. Downloaded on 6/24/2020 4:25:19 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

1. Introduction

The output efficiency (η) of a simple silicon (Si) heterojunction solar cell degrades owing to the presence of a large number of defects at the surface of the Si wafer. This deficiency of heterojunction solar cells can be overcome by the introduction of a thin intrinsic layer between the two doped layers of the heterojunction. Such a structure is known as a heterojunction with intrinsic thin layer (HIT) structure.[1] In a HIT structured solar cell, the intrinsic layer-like a-Si:H reduces the interfacial recombination state density to the minimum level between two doped layers[2] and reduces the leakage current.[3] Recently, SANYO Ltd reported attaining a 20% output efficiency in HIT solar cells based on n-type Si.[4] Furthermore, Panasonic has achieved the world's highest efficiency of 25.6% in a HIT cell based on an n-type c-Si.[5]

In a solar cell, a transparent conductive electrode (TCE) is required to ensure the easy transportation of carriers from the emitter layer to the external circuit. The low sheet resistance (10 Ω sq^{-1}) and high optical transparency (85%) mean that indium tin oxide (ITO) is a commonly used TCE in photovoltaic (PV) applications. However, the high cost, rarity and mechanical rigidity of ITO have limited its use for future flexible devices and these drawbacks initiated the search for new TCE materials.[6] Recently, pristine graphene has been reported to have an optical transparency as high as 97.7% (single layer graphene) with a work function of 4.31–4.5 eV (ref. 7 and 8) and a sheet resistance of 120 Ω sq^{-1} in few-layer graphene.[6] In addition, the intercalated graphene with ferric chloride is reported to have a high optical transparency of approximately 96% with a low sheet resistance of 8.8 Ω sq^{-1} and a high carrier density of 8.9×10^{14} cm^{-2} at room temperature and a high stability in air.[6] At room temperature, the carrier mobility in the graphene layer exceeds 10^5 cm^2 V^{-1} s^{-1}.[9] These extraordinary properties have led to the application of multilayer graphene as a TCE. Recently, Lazar *et al.*[7] reported that doping with boron (B) and $-NH_2$ can increase the work function of graphene (Φ_{Gr}) to 1.20 eV and 0.46 eV from 4.31 eV (pristine graphene). Kim *et al.*[8] demonstrated that, after being irradiated by an α beam the work function of graphene increases to 4.9 eV. Conversely, the low photoresponsivity, poor external quantum efficiency, and low absorption of light reduces the application of graphene as an emitter layer instead of the TCE layer in HIT solar cell.[10] Unlike graphene, another 2D material, molybdenum disulfide (MoS_2) has emerged as a complement 2D material for graphene as an emitter layer in HIT solar cells. MoS_2 has a mobility of 200 cm^2 V^{-1} s^{-1} in the monolayer and approximately 517 cm^2 V^{-1} s^{-1} in few layers, as well as a high on/off ratio in the order of 10^8 (ref. 11 and 12) and a fast photo-response time of 40 μs.[13] MoS_2 is reported to have an indirect bandgap of 1.2 eV in the bulk form and a direct bandgap of 1.9 eV in the single layer form.[14] Moreover, double layer MoS_2 has more than 85% optical transparency in the visible region.[15] Recently, it has been reported that the work function of MoS_2 changes with a change in the number of layers.[16] It has also been demonstrated that MoS_2

[a]Centre of Advanced Research, Department of Physics, Rajiv Gandhi University, Arunachal Pradesh-791112, India. E-mail: sanjeev.kumar@rgu.ac.in
[b]Department of Physics, Central University of Haryana, Haryana-123029, India
[c]Department of Applied Physics, Delhi Technological University, Delhi-110042, India

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0na00309c

Open Access Article. Published on 23 June 2020. Downloaded on 6/24/2020 4:25:19 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

can be used to form significant n-type doping with a Benzyl Viologen (BV) dopant[17] and polyethyleneimine (PEI)[18]. PEI reduces the sheet resistance of MoS_2 2.6 times.[18] These properties mean that MoS_2 can be utilized as an emitter layer in MoS_2/silicon HIT solar cells.

Experimentally and theoretically, it has been demonstrated that MoS_2 can act as an emitter layer in heterojunction silicon solar cells. Tsai *et al.*[19] fabricated a heterojunction solar cell of monolayer MoS_2 and p-cSi and achieved a power conversion efficiency of 5.23%. Hao *et al.*[20] also fabricated a bulk p–n junction after depositing a 40 nm thick layer of n-MoS_2 onto p-cSi substrates and a 1.3% photoconversion efficiency was achieved.

The high transparency, as well as the high mobility of the graphene layer with a tunable bandgap and work function, along with practical and theoretical demonstrations of graphene/MoS_2 and MoS_2/Si heterojunctions, motivated us to propose a single facial HIT solar cell structure of graphene/n-MoS_2/a-Si:H/p-cSi/Au. Here, we have used graphene as a TCE material. No reports have been published on the use of graphene as a TCE in HIT solar cells so far. In addition to graphene and MoS_2, there are a few interesting candidates such as carbon nanotubes, black phosphorous and MXene, that show potential for use in silicon solar cells. So far, a power conversion as high as 17.2% has been observed in carbon nanotube based silicon solar cells.[21–23] After proposing the structure of the cell, the influence of various parameters of the layers used on the cell performance were studied using AFORS-HET v2.5 simulation software. The main objective of the present simulation is to study the independent effects owing to the variation of the parameters of the layers on the cell performance and to demonstrate the prospective application of graphene as an effective TCE material in a n-MoS_2/p-cSi HIT solar cell.

2. Physical model, structural model and formation of the junctions

2.1 Physical model

AFORS-HET is solar cell simulation software, used to study and analyze the various properties of a heterojunction solar cell. For the proposed structure, this software solves Poisson's equation and the transport equation for electrons and holes in one dimension. These equations are as follows:

$$\frac{\varepsilon_0\varepsilon_r}{q}\frac{\partial^2\varphi(x,t)}{\partial x^2} = p(x,t) - n(x,t) + N_D - N_A + \sum_{\text{defects}}\rho_t(x,t)$$

$$-\frac{1}{q}\frac{\partial j_n(x,t)}{\partial x} = G_n(x,t) - R_n(x,t) - \frac{\partial}{\partial t}n(x,t)$$

$$\frac{1}{q}\frac{\partial j_p(x,t)}{\partial x} = G_p(x,t) - R_p(x,t) - \frac{\partial}{\partial t}p(x,t)$$

In which, n represents the electron density, p is the hole density, φ is the electric potential, q is the electron charge, $\varepsilon_0\varepsilon_r$ is the absolute/relative dielectric constant, and $N_{D/A}$ represents the donors/acceptors concentrations, which are assumed to be completely ionized. Here, n, p and φ are the independent variables for which the system of differential equations is solved.

The charge stored in the defects is described by a distribution function, f_t, specifying the probability that defects with a defect density N_t are at position E if the bandgap is occupied with electrons:

$$\text{Acceptor-type defect}: \quad \rho_t(x,t) = -\int dE f_t(E,x,t)N_t(E)$$

$$\text{Donor-type defect}: \quad \rho_t(x,t) = -\int dE[1 - f_t(E,x,t)]N_t(E)$$

The electron/hole currents ($j_{n/p}$) flow owing to the gradient of the corresponding quasi-Fermi energy ($E_{F_{n/p}}$). Within the semiconducting layer, this is equivalent to the sum of the diffusion and drift current with the corresponding mobility $\mu_{n/p}$.

$$j_n(x,t) = q\mu_n n(x,t)\frac{\partial E_{F_n}(x,t)}{\partial x} = -\frac{\mu_n kT}{q}\frac{\partial n(x,t)}{\partial x} + \mu_n n(x,t)\frac{\partial\varphi(x,t)}{\partial x}$$

$$j_p(x,t) = q\mu_p p(x,t)\frac{\partial E_{F_p}(x,t)}{\partial x} = -\frac{\mu_p kT}{q}\frac{\partial p(x,t)}{\partial x} - \mu_p p(x,t)\frac{\partial\varphi(x,t)}{\partial x}$$

If the heterostructure is illuminated (specifying the spectral distribution $\phi(\lambda)$ of the incoming photon flux), the super-bandgap optical generation rate $G_n = G_p\left(\text{for } \frac{hc}{\lambda} \geq E_g\right)$ from the valence band into the conduction band of the semiconductor layers can be evaluated by considering the Lambert–Beer absorption model by specifying the spectral absorption coefficient α, as well as specifying the dielectric properties (n,k) of each layer. An optical sub-bandgap generation $\left(\text{for } \frac{hc}{\lambda} < E_g\right)$ from a defect to the conduction/valence band can be defined by specifying the optical emission coefficients $e_n^o(E)$,$e_p^o(E) \neq 0$ for the defect state.

Recombination from the conduction band into the valence band may occur directly (band to band recombination, Auger recombination) and *via* trap states (Shockley–Read–Hall recombination, (SHR)).

$$R_{n,p}(x,t) = R_{n,p}^{\text{Band–Band}}(x,t) + R_{n,p}^{\text{Auger}}(x,t) + R_{n,p}^{\text{SRH}}(x,t)$$

The transportation across the heterojunction interface is modeled by drift-diffusion. In order to do so, an interface layer is considered. The specified interface defects are distributed homogeneously within this layer. The cell currents have to be

 This journal is © The Royal Society of Chemistry 2020

Open Access Article. Published on 23 June 2020. Downloaded on 6/24/2020 4:25:19 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

evaluated directly from the gradient of the corresponding quasi-Fermi energies. The electron affinity (χ), the bandgap (E_g) and the effective conduction/valence band density of states (N_C, N_V) will depend on the position x within the interface layer.

$$E_{F_n}(x) = -q\chi(x) + q\varphi(x) - kT \ln \frac{n(x)}{N_C(x)}$$

$$E_{F_p}(x) = -q\chi(x) + q\varphi(x) + kT \ln \frac{n(x)}{N_V(x)}$$

The electric potential is fixed to zero at one contact. At the second contact, a boundary condition has to be specified, which relates the external cell voltage/current to the internal quantities. Furthermore, the electron and hole currents moving into the metal contacts need to be modeled. The Schottky contact can either be voltage controlled, or current controlled.

2.2. Structural model

Layered structure MoS_2 can be directly deposited on Si substrates. However, owing to their structural mismatch (*i.e.* layered structure of MoS_2 and diamond-like structure of Si), large quantities of lattice defects are produced and this increases the recombination at the interface. Thus, the introduction of an intrinsic (i) buffer layer a-Si:H could be beneficial to balance the carrier injection and reduce the leakage current

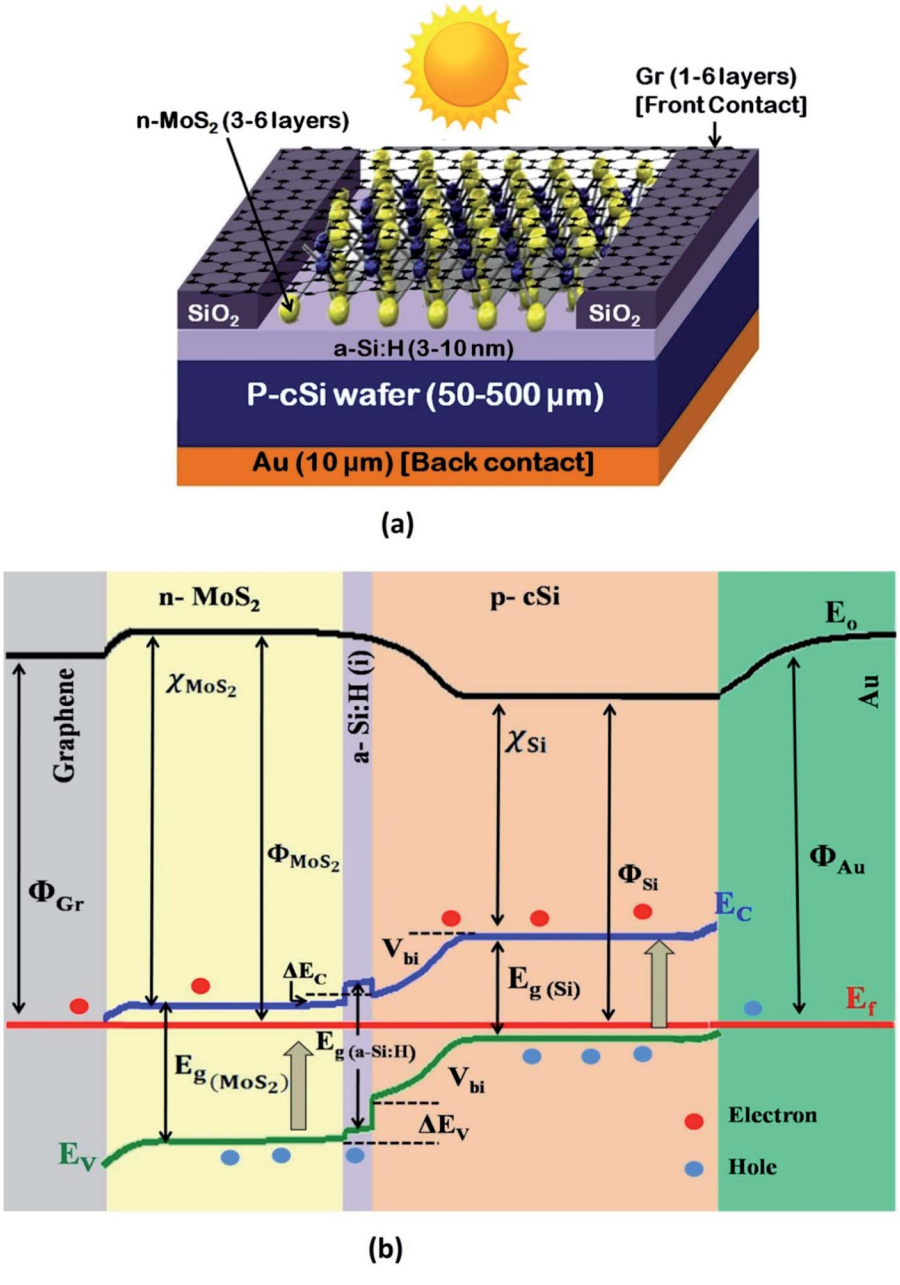


Fig. 1 (a) The proposed HIT solar cell structure, and (b) band diagram as generated using AFORS-HET software for the proposed structure.

Open Access Article. Published on 23 June 2020. Downloaded on 6/24/2020 4:25:19 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

in the cell.[3] Therefore, the as-synthesized MoS_2 is an n-type semiconductor[24] and p-type silicon is more popular in the PV industries.[4] Thus, in this simulation, we have used p-cSi as a base material. Moreover, the minority carrier diffusion length in p-cSi is also reported to be higher than n-cSi.[25] Furthermore, Fan *et al.*[26] have reported that ultrathin gold (Au) contacts provide a better electrical conductivity and optical reflectance-scattering to maintain the performance of the cell compared with ITO. Also, it has been demonstrated that an Au/Si interface formed by using novel transfer printing technology makes excellent electrical contact and exhibits a low contact resistance in comparison to that formed by vacuum deposition of Au on Si.[27] Therefore, the proposed structure of the single facial HIT solar cell is configured as graphene/n-MoS_2/a-Si:H(i)/p-cSi/Au, as shown in Fig. 1. No lattice mismatch between the layers has been considered for this simulation. Ideally, a semiconductor is called 2D if it satisfies the condition $\langle t \rangle \leq \lambda_F$, in which $\langle t \rangle$ is the average system of the 2D electron system, λ_F is the Fermi wavelength which is equal to $\sqrt{\frac{2\pi}{n_{2D}}}$ nm, and n_{2D} is the carrier density in cm^{-2}.[28] Following this condition as reported by Sarma *et al.*[29] and Suh *et al.*[30] even $n \sim 10^5$ cm^{-3} in graphene and $n \sim 1.8 \times 10^{14}$ cm^{-3} in MoS_2 cannot satisfy the above described conditions respectively. It has also been reported that doping in 2D materials increases the interlayer spacing.[31] Thus, we consider graphene and MoS_2 quasi 3D in nature instead of 2D. The quantum confinement in the structure was achieved by controlling the layer number of n-MoS_2. In the proposed HIT structure, as shown in Fig. 1a the carrier transport occurred along the *c*-axis. Details of the front contact and the back contact parameters of the proposed structure are given in Table 1[6,7,26,32–37] and the parameters, as well as the ranges of the n-MoS_2, a-Si:H and p-cSi are given in Table 2. The thickness of the single graphene was taken as 0.334 nm and the monolayer MoS_2 as 0.65 nm.[19,38,39] However, we started our simulation from five-layer (5L) graphene (1.67 nm) and six-layer (6L) MoS_2 (3.9 nm) by considering the high overlapping probability of the graphene and p-cSi. It was also reported that the Raman spectrum of 6L MoS_2 is very close to the bulk[37] and the optical transparency of 5L graphene (88.5%) is more than that of ITO (85%).[6] The parameters considered for optimization of the MoS_2 layer were the donor concentration (N_D), N_C/N_V, E_g, layer thickness, χ, dielectric constant (ε_r) and electron mobility (μ_n). The values of these parameters were kept within range and these and the results of the above mentioned previous literature are tabulated in Table 2.[3,40–47] To simulate and optimize the other parameters the bulk and interfacial defect distribution and recombination rates were kept constant in the software. Furthermore, the MS-Schottky contact model was selected for the graphene/n-MoS_2 and p-cSi/Au interfaces and the drift-diffusion model was selected for the n-MoS_2/a-Si:H and a-Si:H/p-cSi interfaces. The values and parameters of these interfaces[48] are listed in Table 3. The values of these parameters were varied accordingly under air mass 1.5 (AM1.5) illuminations under a power density of 100 mW cm^{-2} and a temperature of 300 K. In order to study the effect of the layer parameters on the cell performance, the open-circuit voltage (V_{OC}), short circuit current density (J_{SC}), fill factor (FF) and efficiency (η) were analyzed. After optimizing the parameters of n-MoS_2 further simulation was extended to the a-Si:H layer and the thickness and bandgap were optimized. The values and ranges of these parameters were obtained from previously published studies.[4,49–51] Further simulations were performed to optimize the parameters, such as the N_A, N_C/N_V, χ, and thickness for p-cSi. The values and the ranges of the parameters for the p-cSi layer were taken from the previously published research.[4,47*and*48] The defect properties were kept as the default. Finally, we optimized the parameters of the contact materials, that is graphene, Au, ITO, Ag, and so forth. The work function (Φ) and the number of layers were optimized and other parameters were fixed, as found in the literature.[7,8,34–37,52]

Table 1 Front and back contact parameters

	Front contact			Back contact	
Parameter	Initially	After optimization		Initially	After optimization
Material	Graphene	ITO	Al	Au	Ag
Thickness	1L–5L (1L = 0.334 nm)	80 nm	80 nm	10 μm	10 μm
Optical properties	$n = 2.7$ $k = 1.45$	ITO *nk* (default)	$n = 1.19$ $k = 7.05$	Au *nk* (default)	Ag *nk* (default)
Work function (eV)	4.31 4.50 4.54 4.77 4.90	4.4 4.45 4.5	4.06 4.20 4.26	5.4	4.64 4.74
Absorption loss	0.023 (constant)	ITO abs (default)	0.056	0	0
External reflection	1L = 0 2L = 0.023 3L = 0.046 4L = 0.070 5L = 0.092	ITO Ref (default)	0.868	0	0
Surface condition	Plane	Plane	Plane	Plane	Plane
Internal reflection	0	0	0	0	0

 This journal is © The Royal Society of Chemistry 2020

Open Access Article. Published on 23 June 2020. Downloaded on 6/24/2020 4:25:19 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Table 2 Parameters of the layers

Parameter	n-MoS_2	a-Si:H(i)	p-cSi
Thickness	3L–6L (1L = 0.65 nm)	3–10 nm	50–500 μm
Dielectric constant (ε_r)	4–14	11.9	11.9
Electron affinity χ (eV)	3.74–4.45	3.9	4.05
Band gap, E_g (eV)	1.41–1.48	1.6–2.0	1.12
Effective conduction band density, N_C (cm^{-3})	3×10^{18} to 9×10^{20}	2.5×10^{20}	3×10^{19} to 1×10^{21}
Effective valence band density, N_V (cm^{-3})	3×10^{18} to 9×10^{20}	2.5×10^{20}	3×10^{19} to 1×10^{21}
Effective electron mobility, μ_n ($cm^2\ V^{-1}\ s^{-1}$)	517	20	1104
Effective hole mobility, μ_p ($cm^2\ V^{-1}\ s^{-1}$)	8.5	5	420
Acceptor concentration N_A (cm^{-3})	0	0	10^{15} to 10^{17}
Donor concentration N_D (cm^{-3})	10^{12} to 10^{18}	0	0
Electron thermal velocity, V_e (cm s^{-1})	10^7	10^7	10^7
Hole thermal velocity, V_h (cm s^{-1})	10^7	10^7	10^7
Layer density, (g cm^{-1})	5.06	2.328	2.328
Auger electron recombination coefficient, ($cm^6\ s^{-1}$)	$\sim 10^{-24}$	0	2.2×10^{-31}
Auger hole recombination coefficient, ($cm^6\ s^{-1}$)	$\sim 10^{-24}$	0	9.9×10^{-32}
Band-to-band recombination coefficient, ($cm^3\ s^{-1}$)	$\sim 10^{-7}$	0	9.5×10^{-15}
Defect type	Single	Conduction tail	Single
Defect charge	Acceptor	Acceptor	Acceptor
Total defect density (cm^{-3})	10^{18}	6.4×10^{19}	10^{10}
Specific defect density ($cm^{-3}\ eV^{-1}$)	10^{18}	1.83×10^{21}	10^{10}
Defect level energy (eV)	0.6	0.035 (Urbach)	0.56
Electron and hole capture cross section (cm^{-2})	10^{-14} (default)	7×10^{-16}	10^{-14}
Optical properties	$n = 4.47$; $k = 1.01$	a-Si *nk* (default)	c-Si *nk* (default)

Table 3 Interfacial Parameters

Interface	Numerical model	Remarks
Graphene/n-MoS_2	MS-Schottky contact	Surface recombination velocity for electrons and holes = 10^7 cm s^{-1} (default)
p-cSi/Au		
n-MoS_2/a-Si:H	Drift diffusion model	Interface charge = 10^{11} cm^{-2}
a-Si:H/p-cSi		

2.3 Formation of the junctions

If n-MoS_2 and p-cSi are brought into contact, a heterojunction forms and bending of the energy bands occur close to the Fermi level. This built-in potential (V_{bi}) is developed at the MoS_2/p-cSi interface. This voltage facilitates the separation of photo-generated carriers in the solar cell. Generally, in the p–n junction solar cell, the interfacial defect density reduces the overall efficiency of the cell. Therefore, better surface passivation between the two layers is required to reduce the defect density at the interface. Here, an intrinsic layer a-Si:H is inserted between n-MoS_2 and p-cSi. This layer can be deposited on Si using the plasma-enhanced chemical vapor deposition method at a low temperature (<200 °C) without degrading the bulk properties of the Si substrate.[4] Also, it introduces band offsets between the active and emitter band which are also beneficial to obstructing the reverse flow of the charge carriers. The schematic energy band diagram of the proposed HIT solar cell is shown in Fig. 1b. The build-in potential (eV_{bi}) in the HIT solar cell is defined as:

$$\begin{aligned} eV_{bi} &= E_{g(MoS_2)} + \Delta E_V + kT \ln\left(\frac{N_A N_D}{N_{V(Si)} N_{C(MoS_2)}}\right) \\ &= E_{g(Si)} - \Delta E_C + kT \ln\left(\frac{N_A N_D}{N_{V(Si)} N_{C(MoS_2)}}\right) \end{aligned} \quad (1)$$

In which e, E_g, ΔE_C and ΔE_V, k and T are the electric charge, bandgap, the conduction band offset and valence band offset, Boltzmann constant, and the temperature in Kelvin, respectively. N_A and N_D represent the acceptor and donor concentrations and N_C and N_V are the conduction band and valence band density.[53] In this HIT cell, a barrier for hole transportation is built at the interface between n-MoS_2 and the pristine graphene corresponding to the potential energy of eV_{bi} which can be expressed using the following equation:[54]

$$eV_{bi} = e(\Phi_{MoS_2} - \Phi_{Gr}) \quad (2)$$

In which, Φ_{MoS_2} and Φ_{Gr} are the work functions of n-MoS_2 and pristine graphene respectively. As reported in previously published literature,[16and55] the Φ of a six layer MoS_2 (~5.4 eV) is greater than that of pristine graphene (~4.1 eV).

3. Results and discussion

3.1 Consequences of the n-MoS_2 layer

In this section, we have optimized the donor N_D of the n-MoS_2 layer in the range from 10^{12} to 10^{18} cm^{-3}, and other parameters such as the thickness, N_C/N_V, E_g, χ and ε_r are kept at feasible values as shown in Fig. 2a and b. In Fig. 2b, it can be seen that as

Open Access Article. Published on 23 June 2020. Downloaded on 6/24/2020 4:25:19 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

N_D was increased from 10^{12} cm^{-3} to 10^{16} cm^{-3}, the V_{OC} remained constant at the value of 636.7 mV. At $N_D > 10^{16}$ cm^{-3}, the V_{OC} increased to 646.1 mV as N_D was increased to 10^{18} cm^{-3}. At a low value of N_D, the consistency of V_{OC} indicates that it does not depend on N_D unless the condition $\Delta n_o \gg N_D$ is followed, as reported previously,[56] in which Δn_o represents the excess electron–hole pairs. At $N_D > 10^{16}$ cm^{-3}, the V_{OC} increases owing to enhancement in the eV_{bi} according to eqn (2). As the mobility of the photo-generated carriers remains almost constant even with an increase of N_D, this might keep J_{SC} constant at 32.33 mA cm^{-2}. Similarly, the FF was found to be constant at 76.71% until the N_D reached 10^{16} cm^{-3} and then the FF slightly reduced to 75.73%. This is due to an increase in the recombination rate at a higher N_D (10^{18} cm^{-3}). As we have observed, the N_D controls the V_{OC}, thus the η of the cell remained constant at 15.79% as the N_D increases, and a maximum of $\eta = 15.82\%$ was obtained at $N_D = 10^{18}$ cm^{-3}.

The N_C/N_V is related to the effective mass of the electrons (m_e^*) and holes (m_h^*) as follows: $N_{C/V} \propto 2m_{e/h}^*$. Furthermore, $m_{e/h}^*$ is related to the mobility ($\mu_{e/h}$) as follows: $\mu_{e/h} = \frac{e\tau}{m_{e/h}^*}$. From these two relationships, it can be concluded that with the change in mobility ($\mu_{e/h}$) the conduction and valence band density (N_C/N_V) also changes. It has been reported that the mobility of the charge carriers in MoS_2 changes with the change in the number of layer. MoS_2 has a mobility of 200 cm^2 V^{-1} s^{-1} in the monolayer and approximately 517 cm^2 V^{-1} s^{-1} in few layers.[11] Therefore, after optimization of N_D to 10^{18} cm^{-3}, the N_C/N_V was optimized in the range 3×10^{18} to 9×10^{20} cm^{-3} as shown in Fig. 2c and d. It can be seen in Fig. 2d that the V_{OC} increased from 646.1 to 668 mV with an increase of the N_C/N_V

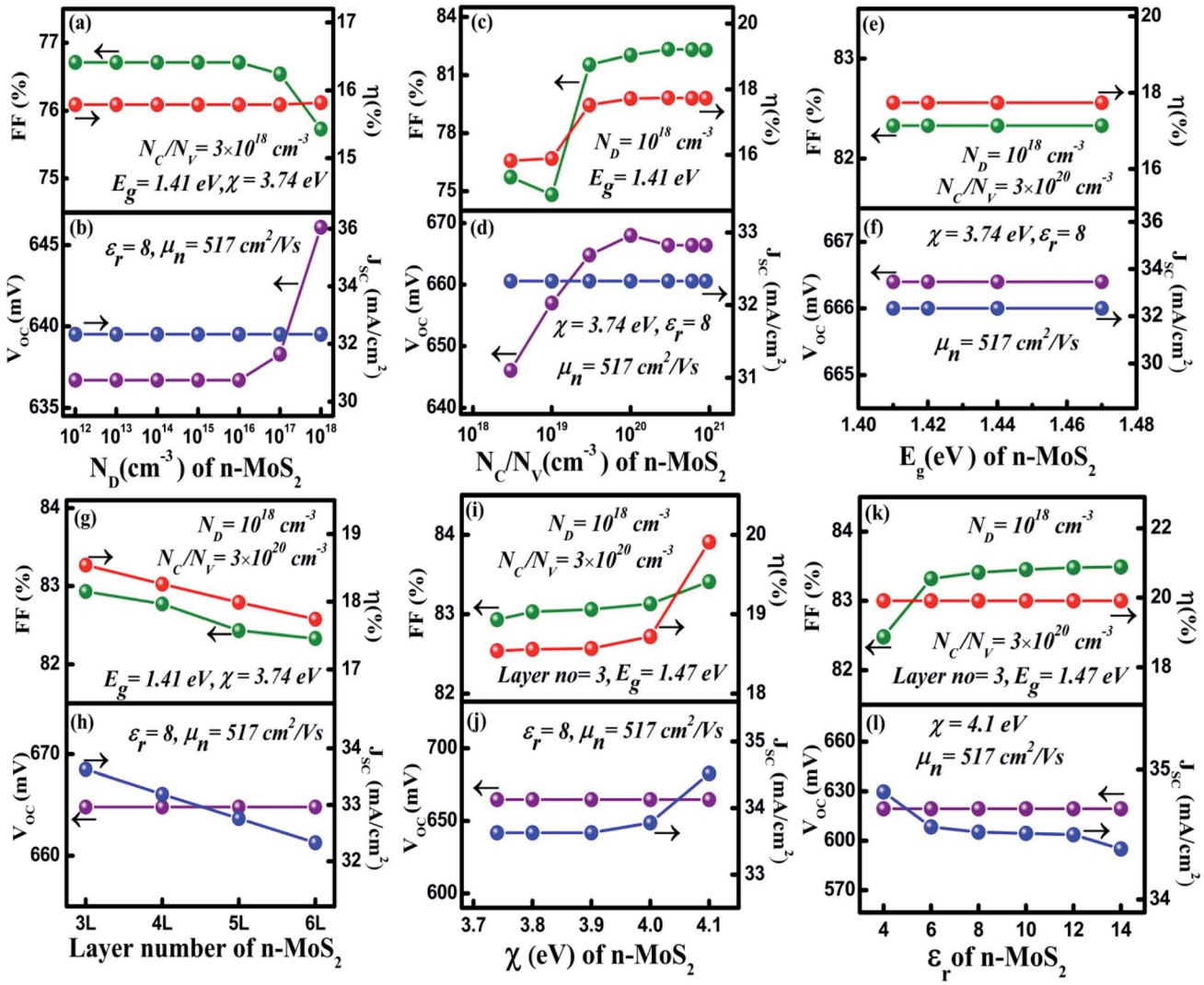


Fig. 2 Optimization of the n-MoS_2 layer: (a) and (b) donor concentration N_D (cm^{-3}), (c) and (d) effective conduction band and valance band density N_C/N_V (cm^{-3}), (e) and (f) the bandgap energy E_g (eV), (g) and (h) the number of layers of MoS_2, (i) and (j) the electron affinity χ (eV), and (k) and (l) the dielectric constant ε_r. The p-cSi parameters were maintained as: thickness = 100 μm, $N_A = 1 \times 10^{16}$ cm^{-3}, $N_C/N_V = 3 \times 10^{19}$ cm^{-3} and $\chi = 4.05$ eV, and the graphene parameters were maintained as: number of layers = 5, reflectance = 0.092 and work function = 4.31 eV. The other parameters are given in Tables 1 and 2.

Open Access Article. Published on 23 June 2020. Downloaded on 6/24/2020 4:25:19 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

from 3×10^{18} cm^{-3} to 10^{20} cm^{-3}. The increment in V_{OC} can be understood by considering the equation: $V_{OC} = \frac{kT}{q} \ln\left(\frac{np}{{n_i}^2}\right)$ in which n_i is the carrier concentration, n is the intrinsic electron concentration and p represents the intrinsic hole concentration. Here, $n = n_o + \Delta n$ and $p = p_o + \Delta p$, in which n_o and p_o are the concentrations of the electrons and holes in the equilibrium condition and Δn, Δp are the excess photo-generated electron and hole concentrations, respectively. As, ${n_o}{p_o} = {n_i}^2$, $n = N_D$ and $\Delta n = \Delta p$, V_{OC} can therefore be rewritten as $V_{OC} = \frac{kT}{q} \ln\left(1 + \frac{\Delta p}{N_D} + \frac{\Delta p^2}{{n_i}^2} + \frac{\Delta p N_D}{{n_i}^2}\right)$.[57] We can assume that V_{OC} increased with N_C/N_V owing to an increase of Δp and Δn. As shown in Fig. 2d, the V_{OC} decreased a little bit to 666.4 mV as the N_C/N_V increased to 9×10^{20} cm^{-3}, which is due to the decrement of eV_{bi} according to eqn (1). The J_{SC} was found to remain constant at 32.33 mA cm^{-2} with an increase of N_C/N_V from 3×10^{18} to 9×10^{20} cm^{-3}. Although the generation rate (G) of the photo-generated carriers is reported to increase with higher values of N_C/N_V.[58] The constant J_{SC} indicates that the photo-generated carriers either recombine or are trapped close to the metal/semiconductor interface. As with the V_{OC}, the FF is also found to increase and a maximum value of 82.33% was obtained at $N_C/N_V = 3 \times 10^{20}$ cm^{-3}. Therefore, η was found to be enhanced from 15.82% to a maximum value of 17.73%.

After optimization of the N_D and N_C/N_V to values of 10^{18} cm^{-3} and 3×10^{20} cm^{-3}, respectively we varied the bandgap (E_g) in the range 1.41–1.47 eV (ref. 39) and results are shown in Fig. 2e and f. As we can see, all of the cell parameters are constant and a value of $\eta = 17.74\%$ was obtained in this range of E_g.

For the simulation with different numbers of layers of n-MoS_2 as shown in Fig. 2g and h, the V_{OC} was found to remain constant at a value of 664.8 mV, whereas the J_{SC} reduced from 33.63 to 32.33 mA cm^{-2}, when the number of layers was increased from 3 to 6. This happens because the increase in the number of layers reduces the transmittance of photons to reach the junction, which results in a reduction of the electron–hole pair generation rate. The reduction of the transmittance with the increase in the n-MoS_2 layer number leads to a reduction of the FF. The η was found to decrease from 18.54% to 17.74% as the number of layers increased from 3 to 6. Hence, we used 3L as the optimized number of layers of MoS_2 for further simulation.

Fig. 2i and j show the cell parameters for the optimization of χ in the ranges 3.7 to 4.1 eV. Here we have considered E_g = 1.47 eV for 3L MoS_2.[42] The V_{OC} remained constant at a value of 664.8 mV when χ (eV) was increased from 3.7 to 4 eV. Upon a further increase in the χ to 4.1 eV, a small increase in the V_{OC} to approximately 691.4 eV was obtained owing to an increase in the eV_{bi}. A constant J_{SC} of 33.63 mA cm^{-2} was found in the range 3.74–3.90 eV of χ and this was enhanced to 34.16 mA cm^{-2} for χ = 4.1 eV. A small increase in the ΔE_C (see Fig. 1b) was observed with the increase in χ, which leads to a slightly higher V_{OC} and J_{SC} at $\chi = 4.1$ eV. The FF was observed to increase from 82.93% (@ $\chi = 3.74$ eV) to 83.41% (@ $\chi = 4.1$ eV) as a result of the reduction in the sheet resistance. Therefore, η was found to increase from 18.54% to 19.19%.

It has been found from electric transportation measurements that the dielectric response of the MoS_2 layers is sensitive to the substrate that is used and the dielectric effect relies on the thickness of the sample.[43] The effect of the dielectric constant (ε_r) on the cell parameters is depicted in Fig. 2k and l. It can be observed in Fig. 2l, that the V_{OC} does not depend on the variation of ε_r and it was found to be a constant at 691.4 eV. However, the J_{SC} shows a light decrement from 34.83 to 34.39 mA cm^{-2} with an increase of ε_r up to 14. In a p–n junction, $eV_{bi} \propto \frac{1}{\varepsilon_r}$, which affect the relative permittivity of the MoS_2 layer.[54,57] The generation of the e–h pair increases in the MoS_2 layer with an increase in the ε_r following Coulombs law as $F \propto \frac{1}{\varepsilon_r}$, in which F is the electrostatic force between the electron and holes, but at the same time the increase in ε_r reduces eV_{bi}. As a consequence of both of these effects, we obtained a constant V_{OC}. In contrast, under short circuit conditions, the photo-generated carriers experienced reduced eV_{bi}, which boosts the recombination of the carriers. Thus, the J_{SC} was found to decrease. The FF showed a minor increase from 82.43% to 83.41% when ε_r was increased from 4 to 6. After this the increase in FF slowed down and only reached 83.49% for $\varepsilon_r = 14$. In Fig. 2k, η was found to largely depend on the V_{OC} and remained constant at 19.91% with ε_r. The cell parameters of the proposed cell were found to be independent of the electron mobility (μ_n) of the n-MoS_2 layer, which has been varied in the range of 517–50 cm^2 V^{-1} s^{-1}.

3.2 Effect of the a-Si:H(i) layer thickness

In the HIT solar cell, an intrinsic layer of a-Si:H is inserted, mainly for passivation of the c-Si surface in the cell.[51] Moreover, the insertion of the thin intrinsic layer also contributes to the depletion region, as well as to the drift current and electric field strength.[59] Fig. 3a and b shows the dependence of the cell parameters on the thickness variation of the a-Si:H(i) layer is in the 3 to 10 nm range. It can be observed in Fig. 3b that along with the increase of a-Si:H(i) thickness from 3 to 4 nm, the V_{OC} was slightly enhanced from 691.4 to 693 mV owing to a minor

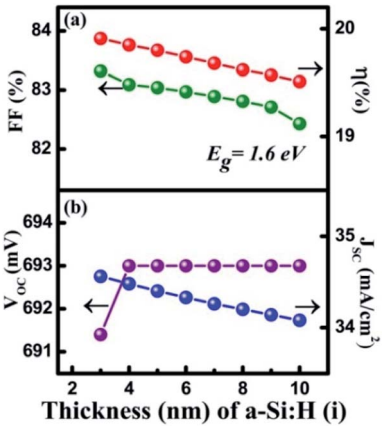


Fig. 3 Optimization of the a-Si:H(i) layer: (a) and (b) thickness (nm) in which E_g was maintained at 1.6 eV.

Open Access Article. Published on 23 June 2020. Downloaded on 6/24/2020 4:25:19 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

increase in the barrier height at the a-Si:H(i) layer and p-cSi interface.[43] The V_{OC} then became constant at 693 mV upon further increasing the thickness of a-Si:H(i). The J_{SC} was found to decrease linearly from 34.56 to 34.08 mA cm^{-2}. It is reported that a thicker intrinsic layer increases the series resistance in the cell.[50] Therefore, we observed a tiny change in the J_{SC}. Thus, the FF was found to reduce from 83.32% to 82.43%. Consequently, the η obtained reduced from 19.91% to 19.51%. To observe the cell performance for the optimized bandgap energy of a-Si:H(i), a further simulation was carried out by varying the E_g a-Si:H(i) in the range 1.6 to 2 eV. However, the maximum η = 19.19% was only obtained at E_g = 1.6 eV. This might be because the absorption of the photons by the MoS_2 layer only corresponds to this E_g of the a-Si:H(i) layer.

3.3 Optimization of the p-cSi layer

After optimizing the n-MoS_2 and a-Si:H(i) parameters to their optimal values, the simulation was further performed to optimize the parameters of the p-cSi layer, while keeping the parameters of n-MoS_2 and a-Si:H(i) at their optimal values, as discussed in the previous section. The results of the cell performance against variation in the N_A variation of p-cSi are shown in Fig. 4a and b. It is observed in Fig. 4b that when the N_A is increased from 10^{15} to 10^{17} cm^{-3}, the V_{OC} also increased from 691.2 to 699.2 mV owing to the higher eV_{bi} with N_A according to eqn (1). Similarly, J_{SC} was also found to increase from 34.23 to 39.15 mA cm^{-2} as the increased eV_b pushes the photo-generated carriers effectively from the depletion region into the contact region without experiencing a recombination effect or the charge carrier trapped effect in the neutral region. This resulted in an increment in the J_{SC}.[57] The increment in the FF from 81.48% to 82.76% can be attributed to the reduction of the sheet resistance as J_{SC} is also increased. These changes lead to an improvement in η from 19.46% to 21.88%. Hence, $N_A = 10^{17}$ cm^{-3} with a maximum η of 21.88% is considered to be the optimal value for the c-Si layer.

To understand the effect of N_C and N_V on the performance of the HIT solar cell, we varied these parameters within the range of 3 × 10^{19} to 1 × 10^{21} cm^{-3} for the p-cSi layer. It can be observed from Fig. 4c and d, with the increase in the N_C/N_V, the V_{OC} decreased linearly from 699.2 to 560.2 mV. According to eqn (1), with the increase in the N_V of p-cSi layer the eV_{bi} decreases and therefore V_{OC} also decreases. In contrast, the value of J_{SC} shows a small increase from 37.81 to 37.91 mA cm^{-2} with N_C/N_V. It has also been reported that N_C/N_V influences the bandgap as well as the absorption of photons.[58] It can be estimated that for N_C/N_V > 3 × 10^{19} cm^{-3}, the high absorption of photons in the active silicon layer increased the G of electron–

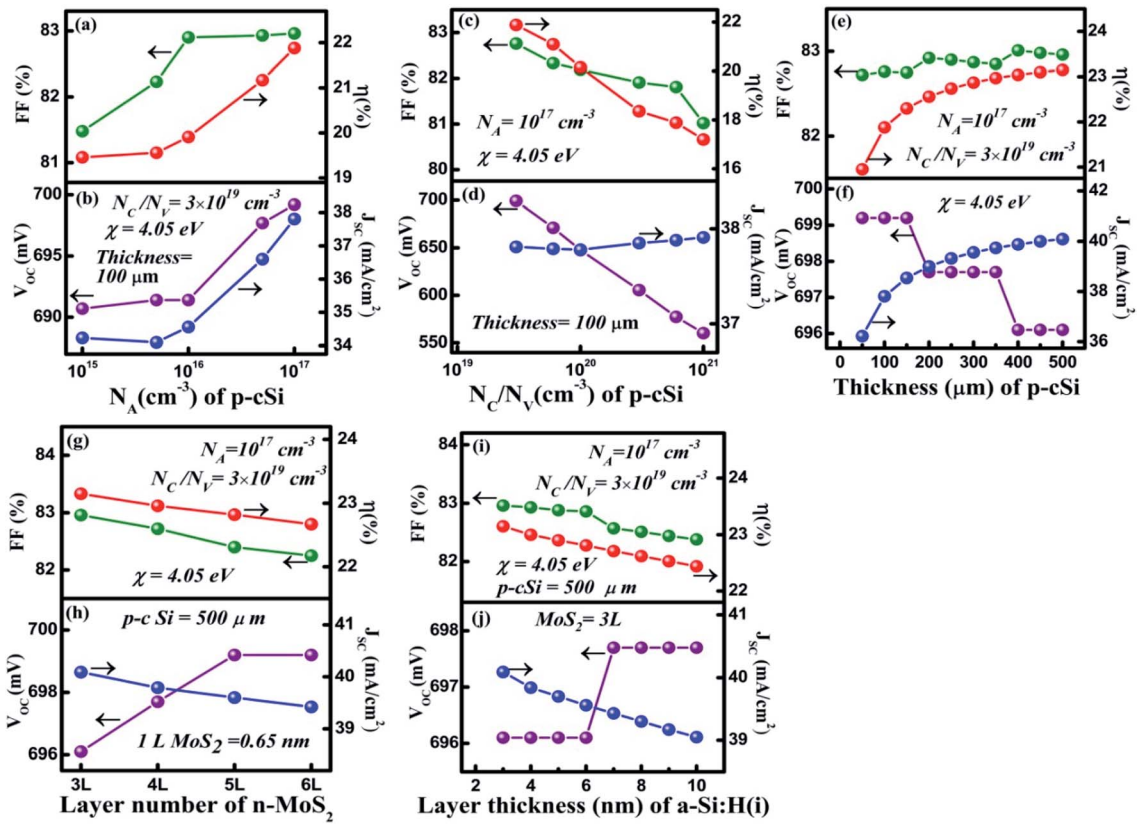


Fig. 4 Optimization of the p-cSi layer: (a) and (b) acceptor concentration N_A (cm^{-3}), (c) and (d) effective conduction band and valance band density N_C/N_V (cm^{-3}), (e) and (f) thickness (μm), (g) and (h) number of layers of MoS_2 and (i) and (j) layer thickness of a-Si:H(i). The optimized parameters of a-Si:H(i) were maintained as: thickness = 3 nm, E_g = 1.6 eV, and the parameters for n-MoS_2 were maintained as: $N_D = 10^{18}$ cm^{-3}, N_C/N_V = 3 × 10^{20} cm^{-3}, number of layers = 3, E_g = 1.47 eV, χ = 4.1 eV and ε_r = 6.

Open Access Article. Published on 23 June 2020. Downloaded on 6/24/2020 4:25:19 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

hole pairs. In Fig. 4d, the FF showed a deterioration from 82.76% to 81.02%, which resulted in the η dropping to 17.2% from 21.88%. The maximum value of $\eta = 21.88\%$ was obtained at $N_C/N_V = 3 \times 10^{19}$ cm^{-3}.

In solar cell technology, the thickness of the p-cSi wafer plays an important role in the fabrication of HIT solar cells. Fig. 4e and f represents the variation of cell parameters as a function of the p-cSi wafer thickness from 50 to 500 μm. In Fig. 4f, the V_{OC} remains constant at 699.2 mV. Upon a further increase in the thickness, the V_{OC} reduced sharply to 697.7 mV and again remained constant up to a thickness of 350 μm. Above this thickness, the V_{OC} was found to deteriorate to the value 696.1 mV and becomes independent of the change of thickness. This variation may be due to the variation in the recombination of the photo-generated carriers under open circuit conditions. The observed enhancement in the J_{SC} from 36.22 to 40.09 mA cm^{-2} can be ascribed to the boost in the generation of the charge carriers with the increase in the p-cSi thickness under short circuit conditions. The FF was observed to increase slowly from 82.72% to 82.96% owing to the reduction in the series resistance. These processes help to increase the η from 20.95% to 23.15% at a thickness of 500 μm.

To study the role of the number of layers of n-MoS_2 and the a-Si:H(i) layer thickness with the best optimized p-cSi parameters, the proposed cell was again simulated by varying the number of layers of MoS_2 and the thickness of a-Si:H and the results are depicted in Fig. 4g, h and i, j respectively. From Fig. 4g and h, it can be clearly seen that the V_{OC} increased from 696.1 to 699.2 mV when the number of layers of n-MoS_2 was increased from 3 to 5. Furthermore, the increase in the number of layers to 6V_{OC} remains constant at 699.2 mV. On the other hand, the J_{SC}, FF, and η were observed to decrease with the increase in the number of layers. The J_{SC} decreased from 40.09% to 39.42%, the FF from 82.96% to 82.25 and the η from 23.15% to 22.67%. In Fig. 4i and j, as the layer thickness of a-Si:H(i) was changed to 6 from 3 nm, the V_{OC} gave a constant value of 696.1 mV and increased slightly to 697.7 mV when the thickness was increased to 6 nm. The V_{OC} then became independent of the

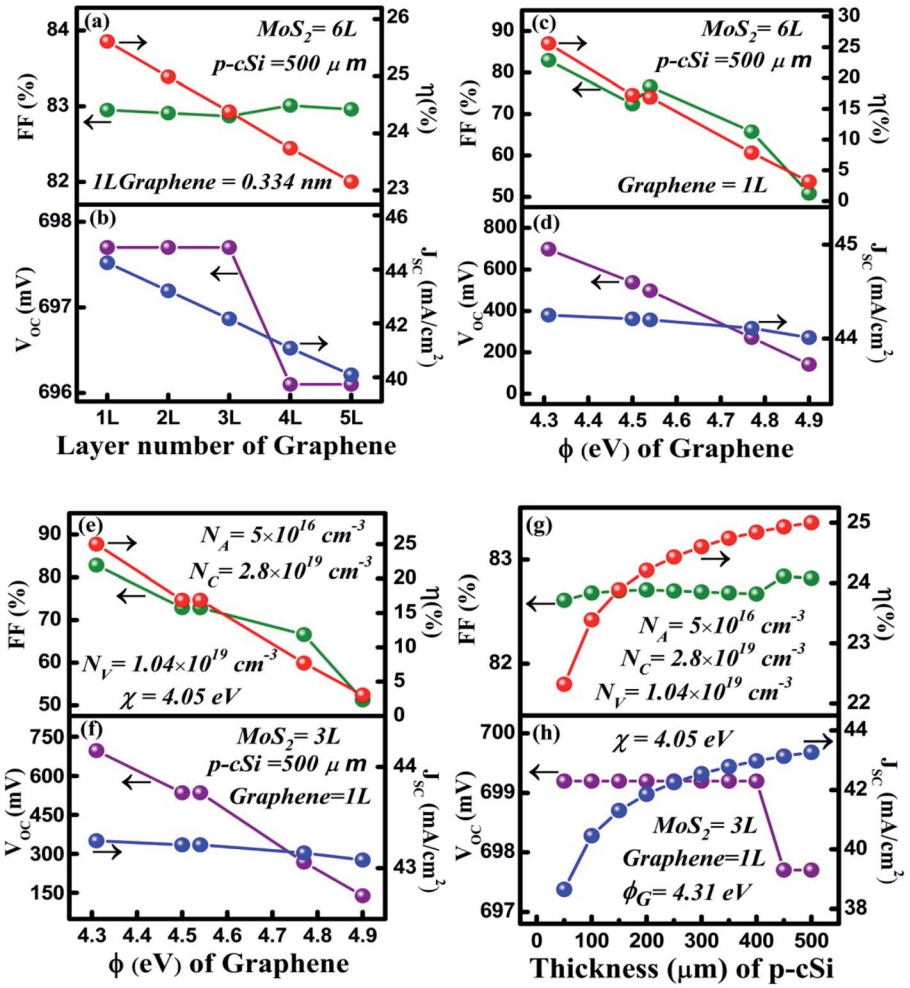


Fig. 5 Optimization of the graphene (front contact) layer: (a) and (b) optimization of the graphene layer number, (c) and (d) optimization of the Φ (eV) of the graphene layer in which the parameters of n-MoS_2 and p-cSi are maintained at their best-optimized values, (e) and (f) optimization of the Φ (eV) of the graphene layer and (g) and (h) optimization of the thickness of a commercially available silicon wafer with practically available p-cSi parameters.

Open Access Article. Published on 23 June 2020. Downloaded on 6/24/2020 4:25:19 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

thickness even when it was increased to 10 nm. The J_{SC} was found to decline from 40.09 to 39.05 mA cm^{-2} with the thickness of a-Si:H(i). The FF was found to decrease from 82.96% to 82.38% and a minimum η of 22.14% was achieved at 10 nm.

3.4 Simulation and optimization of the front contact (graphene) parameters

In a solar cell, contact plays a transparent conducting electrode role in the solar cell by allowing the light to transmit through and maintain a high electrical conductivity. Fig. 5a and b shows the dependence of the HIT solar cell parameters on the variation of the graphene layer (front contact) number. It is observed in Fig. 5b that with the increase in the number of graphene layers from 1 to 3, the V_{OC} remained constant at 697.7 mV. The reason behind this consistency might be the decrease in the recombination rate which limits the reduction of the charge carrier generation under open-circuit voltage conditions. However, upon further increasing of the graphene layer number to 4, the absorbed light might not be sufficient to produce an efficient electron–hole pair, which causes the V_{OC} to slightly decrease to 696.5 mV and then become constant with an increase in the layer number. On the other hand, the J_{SC} was observed to decline linearly from 44.25 to 40.09 mA cm^{-2}. When the number of layers of graphene was increased the optical transmittance reduced and thus the G of the charge carriers decreased and therefore the J_{SC} reduced too. Under short circuit conditions, the photo-generated carriers could be trapped close to the graphene/n-MoS_2 junction. Therefore, we have observed a linear decrease in the J_{SC}, which also, in turn, reduced the η from 25.61% to 23.15%. An insignificant change in the FF was found from 82.95% to 82.96%.

The presence of a strong covalent bond between the C and Mo atoms or the presence of a dangling bond of sulfur present in the interfaces affects the interfacial barrier height and a variation in the barrier height results in a variation in the contact resistance.[60,61] Usually, the contact resistance is a vital factor that affects the solar cell performance.[62] However, many experimental reports suggest that by controlling the Fermi level (*i.e.* the work function) of graphene, the contact resistance can be minimized in the graphene/MoS_2 interfaces.[63,64] Therefore, in this work, we have focused primarily on the work function of graphene. The depicted results shown in Fig. 5c and d, show the variation of the cell parameters with the variation of the graphene work function (Φ_{Gr}) from 4.31 to 4.9 eV. A large reduction in the V_{OC} was observed in Fig. 5d, in which it decreased from 697.7 to 141.4 mV with an increase of Φ_{Gr}. From eqn (2), it can be observed that with the increase in Φ_{Gr}, the potential barrier energy eV_{bi} of graphene/n-MoS_2 decreases. It has also been reported that when the Φ_{TCO} (contact) of Si HIT solar is low, the eV_{bi} of the TCO (transparent conducting oxide)/n-type emitter has the same direction as that of the eV_{bi} of the n-emitter/p-Si. On the other hand, when the Φ_{TCO} (contact) is high, the eV_{bi} of the TCO/n-type emitter contact will have an inverted direction to that of the eV_{bi} of the n-emitter/p-Si junction. Moreover, with the increase in the Φ of the TCO contact in the silicon HIT solar cell, the depletion region in the emitter region increases and widens and results in the overlap of the contact/n-type emitter region and the n-type emitter/p-type silicon junction region if the emitter is not sufficiently thin enough.[4] Thus, the higher Φ_{Gr} might increase the possibility of overlapping between the graphene/n-MoS_2 depletion region and the n-MoS_2/p-cSi depletion region, which has resulted in a reduction in the V_{OC}. In contrast, owing to the overlap of the graphene/n-MoS_2 depletion region and the n-MoS_2/p-cSi depletion region, the eV_{bi} of n-MoS_2/p-cSi reduces. This results in a reduction of the built-in electric field which is unable to push the photo-generated carriers effectively to their respective sides under short circuit conditions and thus the J_{SC} is observed to decline linearly from 44.25 to 40.09 mA cm^{-2} with an increase in the Φ_{Gr}. The overall reduction in the V_{OC} and J_{SC} reduced η from 25.61% to 23.15%. The maximum η = 25.61% was achieved at Φ_{Gr} = 4.31 eV of pristine graphene.

To validate our simulated results, a comparison has been performed by considering the best-optimized values of n-MoS_2 and a-Si:H(i) with the practically available p-cSi parameters by varying the Φ_{Gr} and p-cSi thickness within the range of 4.31–4.9 eV and the 50–500 μm range respectively, as depicted in Fig. 5e–h and Table 4, using the practically available data for p-cSi.[52,65] It can be seen from Fig. 5e and f, that the maximum η =

Table 4 Summary of the most optimized cell

Cell parameter	Most optimized cell (before optimization of the parameters of the graphene layers)	Graphene layer number optimized cell (after optimization of n-MoS_2 and p-cSi)	n-cSi wafer optimized cell for practically available silicon parameters
V_{OC} (mV)	696.1	697.7	697.7
J_{SC} (mA cm^{-2})	40.09	44.25	43.27
FF (%)	82.96	82.95	82.82
η (%)	23.15 [@ 3L n-MoS_2 at $N_D = 10^{18}$ cm^{-3}, $N_C/N_V = 3 \times 10^{20}$ cm^{-3}, E_g = 1.47 eV, χ = 4.47 eV, ε_r = 6, μ_n = 517 cm^2 V^{-1} s^{-1}, μ_p = 8.5 cm^2 V^{-1} s^{-1}, and @ 6 nm a-Si:H(i) at E_g = 1.6 eV and @ 500 μm p-cSi at $N_A = 1 \times 10^{17}$ cm^{-3}, $N_C/N_V = 3 \times 10^{19}$ cm^{-3}, χ = 4.05 eV and @ 5L graphene at Φ = 4.31 eV]	25.16 [@ 3L n-MoS_2 at $N_D = 10^{18}$ cm^{-3}, $N_C/N_V = 3 \times 10^{20}$ cm^{-3}, E_g = 1.47 eV, χ = 4.47 eV, ε_r = 6, μ_n = 517 cm^2 V^{-1} s^{-1}, μ_p = 8.5 cm^2 V^{-1} s^{-1}, and @ 6 nm a-Si:H(i) at E_g = 1.6 eV and @ 500 μm p-cSi at $N_A = 1 \times 10^{17}$ cm^{-3}, $N_C/N_V = 3 \times 10^{19}$ cm^{-3}, χ = 4.05 eV and @ 1L graphene at Φ = 4.31 eV]	25 [@ 3L n-MoS_2 at $N_D = 10^{18}$ cm^{-3}, $N_C/N_V = 3 \times 10^{20}$ cm^{-3}, E_g = 1.47 eV, χ = 4.47 eV, ε_r = 6, μ_n = 517 cm^2 V^{-1} s^{-1}, μ_p = 8.5 cm^2 V^{-1} s^{-1}, and @ 6 nm a-Si:H(i) at E_g = 1.6 eV and @ 500 μm p-cSi at $N_A = 5 \times 10^{16}$ cm^{-3}, $N_C/N_V = 2.8 \times 10^{19}/1.04 \times 10^{19}$ cm^{-3}, χ = 4.05 eV and @ 1L graphene at Φ = 4.31 eV]

Open Access Article. Published on 23 June 2020. Downloaded on 6/24/2020 4:25:19 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

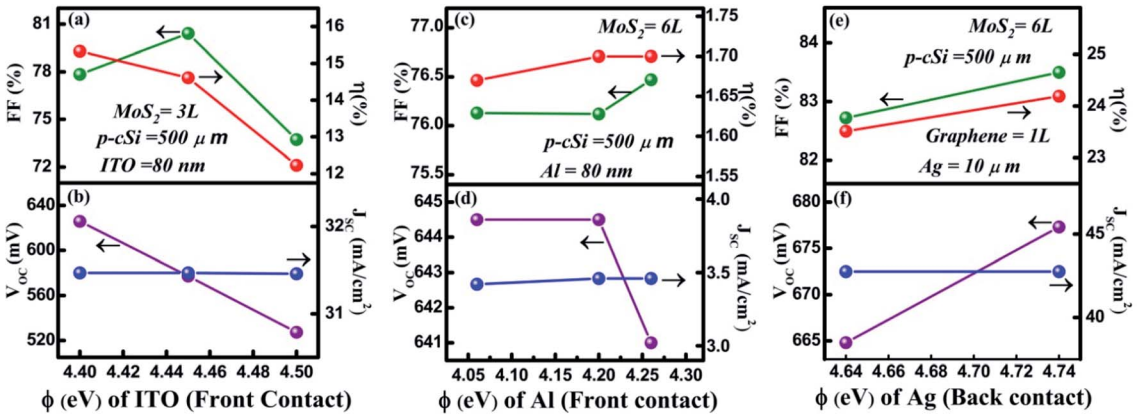


Fig. 6 Changing the front and back contact: (a) and (b) variation in the Φ (eV) of ITO, (c) and (d) variation in the Φ (eV) of Al, and (e) and (f) variation in the Φ (eV) of Ag.

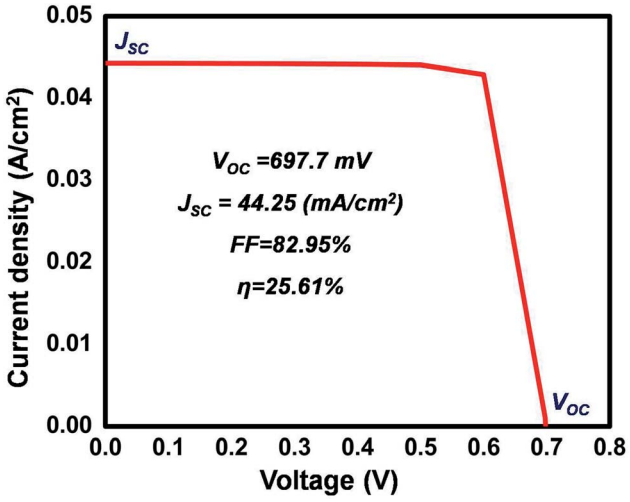


Fig. 7 *IV* characteristics of the most optimized cell.

25% was obtained at $\Phi_{Gr} = 4.31$ eV, which is close to our best-optimized results as shown in Fig. 5c and d. Similarly, in Fig. 5g and h, the same maximum $\eta = 25\%$ was achieved for a 500 μm thick p-cSi wafer.

3.5 Variation and simulation of the front contact and back contact material

In this section, a comparison of the cell parameters has been performed by changing the front contact material graphene to ITO and Al[19] and the back contact to Ag. As seen in Fig. 6a and b the thickness of ITO is considered to be 80 nm and the Φ_{ITO} has been varied from 4.40–4.50 eV.[34] The maximum values for V_{OC}, J_{SC}, FF and η were obtained and found to be 625.8 mV, 31.47 mA cm^{-2}, 80.41% and 15.33% respectively. Similarly, in Fig. 6c and d, when Φ_{Al} was varied, the maximum values for V_{OC}, J_{SC}, FF, and η were found to be 650.8 mV, 4.36 mA cm^{-2}, 82.79% and 2.353%, respectively. In both cases, η was significantly lower than that of η obtained using the graphene layer as a contact.

Fig. 6e and f shows the variation in the cell parameters when the back contact was changed from Au to Ag and the Φ of Ag was maintained from 4.64 to 4.74 eV. It was observed that in this range of Φ for Ag, the V_{OC} increased from 664.8 to 677.3 mV, whereas the J_{SC} remained constant at 42.76 mA cm^{-2}. The FF and η were observed to decrease from 82.72% to 83.5% and 23.51% to 24.19%, respectively. This η was close to our simulated η_{max} value, that is, 25.61% when we used Au as a contact. Here, the back contacts have been considered as the uniform thin film deposited on the backside of p-cSi.

3.6 maximum power and series resistance

The proposed solar cell as shown in Fig. 1a was simulated and characterized under a standard AM1.5G spectrum and forward voltage in the range of 0 to 0.7 V. The *IV* characteristics of the best optimized solar cell generated by the software are shown in Fig. 7. The maximum power delivered by the cell, as well as the

Table 5 Graphene and MoS_2 silicon-based solar cells

Type of study	Solar cell structure	V_{OC} (V)	J_{SC} (mA cm^{-2})	FF (%)	η (%)	Ref.
Experimental	ITO/graphene/MoS_2/n-cSi/(Ti/Ag)	0.59	36.8	73	15.8	66
Experimental	Au/graphene/MoS_2/n-Si/In	0.56	33.4	60	11.1	67
Experimental	Au/graphene/MoS_2/n-Si/Au	0.50	28.1	47	6.56	68
Experimental	Pd/n-MoS_2/i-SiO_2/p-Si/In	0.30	5.5	42	4.5	69
Experimental	Ni/graphene/MoS_2/p-csi/Al	0.51	—	—	2.58	70
Theoretical	TCO/MoS_2/SiGe:H/p-Si/Al	0.652	40.01	83.7	21.85	3
Theoretical	Graphene/n-MoS_2/a-Si:H/p-cSi/Au	0.697	44.25	82.95	25.61	This work

Open Access Article. Published on 23 June 2020. Downloaded on 6/24/2020 4:25:19 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

series resistance (R_s), was calculated from the *IV* characteristics and found to be 25.4 mW cm^{-2} and 2.5 Ω cm^2 was achieved respectively (see ESI†).

Finally, a comparison between the experimental and theoretical demonstration of graphene, MoS_2, and Si-based solar cells was performed and the results are detailed in Table 5. However, except for our work none of the studies used graphene as a contact material or TCE.

4. Conclusions

In order to develop an efficient HIT solar cell, a graphene/n-MoS_2/a-Si:H(i)/p-cSi/Au structure has been proposed and simulated using AFORS-HET v 2.5 software. The individual effects of the six parameters of the n-MoS_2 layer were analyzed individually. After optimizing the parameters of the n-MoS_2, a-Si:H(i), and p-cSi layers, an optimum efficiency of 23.15% was achieved. As the contact material in a solar cell has an adverse effect on the cell performance, a further simulation was performed to observe the effect of graphene on the proposed HIT solar cell by varying the number of layers and the work function. From this simulation, a maximum efficiency of 25.61% was achieved for single-layer graphene at a work function of 4.1 eV. On the other hand, by changing the front contact graphene to ITO and Al we achieved efficiencies of 15.33% and 2.35%. This confirms that graphene could be an efficient contact material for use in an n-MoS_2/a-Si:H(i)/p-cSi/Au HIT solar cell. Our simulated results establish that the graphene layer could be an effective contact TCE material and n-MoS_2 could be an efficient emitter layer in a single facial n-MoS_2 and p-cSi based HIT solar cell.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors are very thankful to Helmholtz-Zentrum Berlin for providing us with the AFORS-HET simulation software and the Department of Science and Technology (DST), Ministry of India for their financial support.

References

1 N. Dwivedi, S. Kumar, A. Bisht, K. Patel and S. Sudhakar, *Sol. Energy*, 2013, **88**, 31–41.
2 M. Belarbi, M. Beghdad and A. Mekemeche, *Sol. Energy*, 2016, **127**, 206–215.
3 S. Xu, X. Zeng, W. Wang, G. Zhou, Y. Hu and S. Wu, *J. Miner. Mater. Char. Eng.*, 2017, **5**, 323–338.
4 L. Zhao, C. L. Zhou, H. L. Li, H. W. Diao and W. J. Wang, *Sol. Energy Mater. Sol. Cells*, 2008, **92**, 673–681.
5 K. Masuko, M. Shigematsu, T. Hashiguchi, D. Fujishima, M. Kai and N. Yoshimura, *IEEE J. Photovolt.*, 2014, **4**, 1433–1435.
6 I. Khrapach, F. Withes, T. H. Bointon, D. K. Polyushkin, W. L. Barnes, S. Russo and M. F. Craciun, *Adv. Mater.*, 2012, **24**, 2488–2849.
7 P. Lazar, R. Zboril, M. Pumera and M. Otyepka, *Phys. Chem. Chem. Phys.*, 2014, **16**, 14231–14235.
8 J. H. Kim, J. H. Hwang, J. Suh, S. Tongay, S. Kwon, C. C. Hwang, J. Wu and J. Y. Park, *Appl. Phys. Lett.*, 2013, **103**, 171604.
9 A. S. Mayorov, R. V. Gorbachev, S. V. Morozov, L. Britnell, R. Jalil, L. A. Ponomarenko, P. Blake, K. S. Novoselov, K. Watanbe, T. Taniguchi and A. K. Geim, *Nano Lett.*, 2011, **11**, 2396–2399.
10 X. Zhong, W. Zhou, Y. Peng, Y. Zhou, F. Zhou, Y. Yin and D. Tang, *RSC Adv.*, 2015, **5**, 45239.
11 B. Radisavljevic, A. Radenovic, J. Brivio, V. Giacometti and A. Kis, *Nat. Nanotechnol.*, 2011, **6**, 147.
12 B. Radisavljevic and A. Kis, *Nat. Mater.*, 2013, **12**, 815.
13 D. S. Tsai, K. K. Liu, D. H. Lien, M. L. Tsai, C. F. Kang and C. A. Lin, *ACS Nano*, 2013, **7**, 3095–3911.
14 X. Yang, W. Fu, W. Liu, J. Hong, Y. Cai, C. Jin, M. Xu, H. Wang, D. Yang and H. Chen, *J. Mater. Chem. A*, 2014, **2**, 7727.
15 J. M. Yun, Y. J. Noh, J. S. Yeo, Y. J. Go, S. I. Na, H. G. Jeong, J. Kim, S. Lee, S. S. Kim, H. Y. Kim and D. Y. Kim, *J. Mater. Chem. C*, 2013, **1**, 3777.
16 S. Choi, Z. Shaolin and W. Yang, *J. Korean Phys. Soc.*, 2014, **64**, 1550–1555.
17 J. H. Li, D. Lee, D. Qu, X. Liu, J. Ryu and A. Seabaugh, *Nat. Commun.*, 2015, **6**, 6564.
18 Y. Du, H. Liu, A. T. Neal, M. Si and P. D. Ye, *IEEE Electron Device Lett.*, 2013, **34**, 1328–1330.
19 M. L. Tsai, S. H. Su, J. K. Chang, D. S. Tsai, C. H. Chen and C. I. Wu, *ACS Nano*, 2014, **8**, 8317–8322.
20 L. Hao, Y. Liu, W. Gao, Z. Han, Q. Xue and H. Zeng, *J. Appl. Phys.*, 2015, **117**, 114502.
21 D. D. Tune, N. Mallik, H. Fornasier and B. S. Flavel, *Adv. Energy Mater.*, 2020, **10**, 1903261.
22 M. Bat-Erdene, M. Batmunkh, S. A. Tawfik, M. Fronzi, M. J. Ford, C. J. Shearer, L. Yu, M. Dadkhah, J. R. Gascooke, C. T. Gibson and J. G. Shapter, *Adv. Funct. Mater.*, 2017, 1704488.
23 L. Yu, A. S. R. Bati, T. S. L. Grace, M. Batmunkh and J. G. Shapter, *Adv. Energy Mater.*, 2019, 1901063.
24 S. Lin, P. Wang, X. Li, Z. Wu, Z. Xu and S. Zhang, *Appl. Phys. Lett.*, 2015, **107**, 153904.
25 Z. Qiao, J. L. Ji and Y. L. Zhang, *Chin. Phys. B*, 2017, **26**, 068802.
26 L. Fan, F. Wang, J. Liang, X. Yao, J. Fang and D. Zhang, *J. Semicond.*, 2017, **38**, 014003.
27 H. Keum, H. J. Chung and S. Kim, *ACS Appl. Mater. Interfaces*, 2013, **5**, 6061–6065.
28 *Introduction to Quantum Transport in Low-Dimensional Systems*, http://www.hri.res.in/∼nanotr16/notes/KJThomas-1.pdf, accessed July 2018.
29 S. D. Sarma, S. Adam, E. H. Hwang and E. Rossi, *Rev. Mod. Phys.*, 2011, **83**, 407.

 This journal is © The Royal Society of Chemistry 2020

Open Access Article. Published on 23 June 2020. Downloaded on 6/24/2020 4:25:19 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

30 J. Suh, T. E. Park, D. Y. Lin, D. Fu, J. Park and H. J. Jung, *Nano Lett.*, 2014, **14**, 6976–6982.
31 Z. Lin, B. R. Carvalho, E. Kahn, R. Lv and R. Rao, *2D Matter.*, 2016, **3**, 022002.
32 A. I. Pantazi and V. Yannopapas, *Opt. Commun.*, 2016, **374**, 24–28.
33 K. Patel and P. K. Tyagi, *Carbon*, 2017, **116**, 744–752.
34 Y. Park, V. Choong, Y. Gao, B. R. Hsieh and C. W. Tang, *Appl. Phys. Lett.*, 1996, **68**, 2699–2701.
35 A. Kahn, Fermi level, work function and vacuum level, *Mater. Horiz.*, 2016, **3**, 7–10.
36 R. M. Eastment and C. H. B. Mee, *J. Phys. F: Met. Phys.*, 1973, **3**, 1738–1745.
37 A. W. Dweydari and C. H. B. Mee, *Phys. Status Solidi A*, 1975, **27**, 223–230.
38 Z. H. Ni, H. M. Wang, J. Kasim, H. M. Fan, T. Yu and Y. H. Wu, *Nano Lett.*, 2007, **7**, 2758–2763.
39 F. J. Nelson, V. K. Kamineni, T. Zhang, E. S. Comfort, J. U. Lee and A. C. Diebold, *Appl. Phys. Lett.*, 2010, **97**, 253110.
40 C. P. Lu, G. Li, J. Mao, L. M. Wang and E. Y. Andrei, *Nano Lett.*, 2014, **14**, 4628.
41 B. Radisavljevic and A. Kis, *Nat. Mater.*, 2013, **12**, 815–820.
42 K. F. Mak, C. Lee, J. Hone, J. Shan and T. Y. Heinz, *Phys. Rev. Lett.*, 2010, **105**, 136805.
43 E. J. G. Santos and E. Kaxiras, *ACS Nano*, 2013, **7**, 10741–10746.
44 S. L. Howell, D. Jariwala, C. C. Wu, K. S. Chen, V. K. Sangwan and J. Kang, *Nano Lett.*, 2015, **15**, 2278–2284.
45 O. Salehzadesh, N. H. Tran, X. Liu, I. Shihand and Z. Mi, *Nano Lett.*, 2014, **14**, 125–4130.
46 H. Zhang, Y. Ma, Y. Wan, X. Rong, Z. Xie and W. Wang, *Sci. Rep.*, 2015, **5**, 8440.
47 W. Zhou, X. Zou, S. Naijmaei, Z. Liu, Y. Shi and J. Kong, *Nano Lett.*, 2013, **13**, 2615–2622.
48 X. Garros, G. Reimbold, J. Cluzel, D. Muñoz and P.-J. Ribeyron, *Microelectron. Eng.*, 2011, **88**, 1247–1250.
49 L. Qin, Y. Xiao-Jun, L. Cheng and C. Ming-bo, *Optoelectron. Lett.*, 2010, **6**, 0108–0111.
50 W. Lisheng and C. Fengxiang, *J. Optoelectron. Adv. Mater.*, 2011, **13**, 81–88.
51 A. Rawat, M. Sharma, D. Chaudhary, S. Sudhakar and S. Kumar, *Sol. Energy*, 2014, **110**, 691–703.
52 L. Zhao, H. L. Li, C. L. Zhou, H. W. Diao and W. J. Wang, *Sol. Energy*, 2009, **83**, 812–816.
53 F. Rana, Semiconductor Heterostructure, in *Semiconductor Optoelectronics*, 2017, ch. 2, https://courses.cit.cornell.edu/ece533/Lectures/handout2.pdf.
54 J. Singh, *Semiconductor devices: Basic Principles*, Wiley India (P) Ltd, New Delhi, 2004.
55 V. K. Thakur and M. K. Thakur, *Chemical Functionalization of Carbon Nanomaterials: Chemistry and Applications*, CRC Press, New York, 2016, p. 918.
56 A. V. Sachenko, V. P. Kostylyov, M. V. Gerasymenko, R. M. Korkishko, M. R. Kulish and M. I. Slipchenko, *Quantum Electronics & Optoelectronics*, 2016, **19**, 67–74.
57 Y. Zhao, C. Liang, M. Sun, Q. Liu, F. Zhang and D. Li, *J. Appl. Phys.*, 2014, **116**, 154506.
58 Y. Zhou and G. Long, *J. Phys. Chem. C*, 2017, **121**, 1455–1462.
59 W. Lisheng, C. Fengxiang and A. Yu, *J. Phys.: Conf. Ser.*, 2011, **276**, 012177.
60 N. Jensen, R. M. Hausne, R. B. Bergmann, J. H. Werner and U. Rau, *Prog. Photovoltaics Res. Appl.*, 2002, **10**, 1–13.
61 X. Liu, J. Gao, G. Zhang and Y. W. Zhang, *Nano Res.*, 2017, **10**, 2944–2953.
62 M. Houssa, K. Iordanidou, A. Dabral, A. Lu, R. Meng, G. Pourtois, V. V. Afanas'ev and A. Stesmans, *Appl. Phys. Lett.*, 2019, **114**, 163101.
63 D. K. Schroder and D. L. Meier, *IEEE Trans. Electron Devices*, 1984, **31**(5), 637–647.
64 Y.-J. Yu, Y. Zhao, S. Ryu, L. E. Brus, K. S. Kim and P. Kim, *Nano Lett.*, 2009, **9**, 3430–3434.
65 H. Wang, L. Yu, Y.-H. Lee, Y. Shi, A. Hsu, M. L. Chin, L.-J. Li, M. Dubey, J. Kong and T. Palacios, *Nano Lett.*, 2012, **12**, 4674–4680.
66 J. Ma, H. Bai, W. Zhao, Y. Yuan and K. Zhang, *Sol. Energy*, 2018, **160**, 76–84.
67 Y. Tsuboi, F. Wang, D. Kozawa, K. Funahashi, S. Mouri, Y. Miyauchi, T. Takenobub and K. Matsuda, *Nanoscale*, 2015, **7**, 14476–14482.
68 K. Jiao, C. Duan, X. Wu, J. Chen, Y. Wang and Y. Chen, *Phys. Chem. Chem. Phys.*, 2015, **17**, 8182.
69 L. Z. Hao, W. Gao, Y. J. Liu, Z. D. Han, Q. Z. Xue, W. Y. Guo, J. Zhu and Y. R. Li, *Nanoscale*, 2015, **7**, 8304–8308.
70 W. Shi and M. Xiying, *Coatings*, 2018, **8**, 2.

Received: 28 October 2019 | Revised: 23 February 2020 | Accepted: 21 March 2020

DOI: 10.1002/ctpp.201900175

Check for updates

Contributions to Plasma Physics

ORIGINAL ARTICLE

Theoretical analysis for transmission of Gaussian and sine time irradiance of electromagnetic beam in collisional dusty plasmas

Ruchi Sharma | **Suresh C. Sharma**

Department of Applied Physics, Delhi Technological University (DTU), Shahbad Daulatpur, Bawana Road, Delhi, India

Correspondence
Suresh C. Sharma Department of Applied Physics, Delhi Technological University (DTU), Shahbad Daulatpur, Bawana Road, Delhi 110042, India.
Email: suresh321sharma@gmail.com

Abstract

A theoretical model of propagation of Gaussian and Sine time irradiance of an electromagnetic beam in collisional dusty plasma has been done in the present analysis. It contains equilibrium of dust charge, particle density, and energy of plasma ingredients having charge neutrality. Ionization of neutral particles, recombination of free electrons with ions, adsorption and emission of electrons from dust grain surface, and binary collisions between plasma components are also considered in this treatment. Time varying behaviour of modified electron temperature and collision frequency has been illustrated numerically as a function of dust densities. Also, the comparative analyses of variation of beam waist parameter with the dimensionless length of transmission for both the Gaussian and Sine time irradiance are involved in this model as a function of distinguishable time width, collision frequencies, and dust densities under the condition that the size of dust nebulous is greater than the electrons mean free path for the adsorption on the dust grain surface. The observed results are significant for the applications in industry and astrophysics.

KEYWORDS

dusty plasma, electromagnetic beam, time irradiance, transmission

1 | INTRODUCTION

Several research[1–11] studies have been performed to show the significant effect of dust grains in plasma because of its exciting applications[12–18] in space, industries, and laboratories plasmas. Particularly, the self-focusing behaviour of an electromagnetic beam propagating in dusty plasma becomes an interesting domain of research promoted by the various investigators.[19–25] Mishra et al.[26] have studied the transportation of complex plasma featured by Mathis, Rumpl, and Nordsieck (MRN) power law grain size distribution with and without electric field. They observed the larger plasma transparency with larger dust grains in dark plasma while the case is reversed in illuminated plasma where the photoemission is predominant. Sharma and Sharma[27] have discussed the kinetics of nonuniform transmission of a Gaussian electromagnetic beam in strongly ionized dusty plasma considering the elastic and charging collisions with photoemission and electrons adsorption on the boundary of dust grains. Jia et al.[28] used the Boltzmann distribution law to study the transportation characteristics of an electromagnetic wave propagating in weakly and strongly ionized dusty plasma. Mishra et al.[29] have used the paraxial ray approximation to study the propagation behaviour of a Gaussian irradiance

 © 2020 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

of an electromagnetic beam in dusty plasma under the effect of elastic and inelastic collisions. Wang et al.[30] have studied the electromagnetic wave propagation characteristics of inhomogeneous dusty plasma. They developed the models of double exponential and hyperbolic electron number density distribution for dusty plasma and also studied the power transmission and reflection coefficient on the basis of their analytical study. Sodha et al.[31] also analysed the cases of thermionic emission, photoemission, and emissionless processes to formulate the problem of fluctuation of charge on account of presence of dust grains.

Several investigations on temporal profile of electromagnetic wave propagating in dust-free plasma have been carried out by various researchers[32–34] but in this paper; analytical model is evolved to access the transmission kinetics of both Gaussian and Sine time irradiance of electromagnetic wave owing to mutual interactions of dust and other plasma components. It is familiar that the electrons in dusty plasma get heated up because of highly irradiated electromagnetic wave to the stationary state temperature of the order of relaxation time $\tau' = 1/(\delta_{comp}\nu_{ef})$, where δ_{comp} is the fractional loss of energy of electrons due to collisions and ν_{ef} is the effective collision frequency under the influence of dust grains. Consequently, with the increasing electron temperature, electron density attains a steady state value during the relaxation time τ'. For the pulse duration $(t_0) < \tau'$, $n_e/n_{e0} = 1$ and for $t_0 > \tau'$, n_e/n_{e0} attains a steady value. Hence, the electron density is considered as a function of temperature. In this communication, for $t_0 > \tau'$, the steady-state electron density and temperature in dusty plasma is obtained on account of ionization of neutrals, recombination of free electrons with ions, net adsorption on dust grain surface, and diffusion due to space charge field and temperature gradient which are ignored in other analytical studies.[19, 22] Power loss due to ohmic heating and collisions between the constituent species have also been taken into account to predict the electron temperature, which causes the nonlinearity. Ohmic heating of ions has been ignored due to their heavier mass. The significance of thermal conduction has been neglected in the present formalism.

For the computation, metallic dust cloud of stainless steel having work function of 7.8 eV as in ionosphere has been considered so that photoemission can take place for spectrum of Lyman α radiations (1215.7 Å) in the extreme UV region.

The left over manuscript is schemed as follows: In Section 2, the coupled sets of differential equations of charge neutrality, electron density, and electron temperature are solved considering the paraxial ray approximation to structure the analytical model for formulating the modified dielectric permittivity under the influence of dust grains. Section 3 includes the solution of propagation of electromagnetic beam for the both Gaussian and Sine time irradiance in dusty plasma following the Akhmanov's et al.[35] and its extension by Sodha et al.[36] in respect of modification of the beam waist parameter f with the dimensionless length of transmission ζ for distinguishable dust parameters. Consequences of the existing model are discussed in Section 4, and, in the end, we reach to the conclusions in Section 5.

2 | ANALYTICAL MODEL

Assume dusty plasma to be made up of dust grain spheres with charge $-Z_d e$ and radius r_d, ions, electrons, and neutral atoms, where Z_d is the state of charge on dust and e is the charge on an electron. The kinetics of dusty plasma include ambipolar diffusion, neutral ionization, neutralization of the ions, adsorption and photoemission of electrons, collisions between plasma constituents, and Ohmic heating caused by nonuniform irradiance of the beam.

The following equations demonstrate the kinetics of dusty plasma as:

2.1 | Ambipolar diffusion term

The ambipolar diffusion term due to density decay of constituent species is given by

$$\prod_e = -n_e\mu_e U - D_{te} n_e(\nabla T_e/T_e) - D_{ne}\nabla n_e, \tag{1a}$$

$$\prod_i = n_i\mu_i U - D_{ti} n_i(\nabla T_i/T_i) - D_{ni}\nabla n_i, \tag{1b}$$

$$\prod_d = -n_d\mu_d U - D_{nd}\nabla n_d, \tag{1c}$$

where Π_g, n_g, U, T_g, $D_{tg}(=5k_BT_g/2m_g\nu_{efg})$, $D_{ng}(=k_BT_g/m_g\nu_{efg})$, $\mu_g(=eD_{ng}/k_BT_g)$, are the flux, number density, space charge field, temperature, thermal diffusion coefficient, diffusion coefficients, and mobility; the subscript g stands for electrons, ions, and dust grains, respectively.

The net rate of diffusion of both the charges per unit area must be balanced in stationary state, so we get

$$\prod_i = \prod_e + \prod_d. \tag{2a}$$

The mobility and diffusion of dust particles are supposed to be zero because of their heavy mass and hence their flux comes out to be zero. Thus, the Equation (2a) gives

$$\prod_i = \prod_e. \tag{2b}$$

Using the Equations (1a), (1b), and (2b), eliminating U, net particle flux or density decay comes out to be

$$\prod_p = -D_{te} n_e (\nabla T_e/T_e) - D_{ne} \nabla n_e - n_e \mu_e \left(\frac{D_{ni} \nabla n_i - D_{ne} \nabla n_e + D_{ti} n_i (\nabla T_i/T_i) - D_{te} n_e (\nabla T_e/T_e)}{n_e \mu_e + n_i \mu_i} \right), \tag{3a}$$

$$\prod_p = - \left(\frac{D_{ni}[c_2 n_e \nabla n_i + c_1 n_i \nabla n_e] + D_{ti} n_e n_i [c_2 (\nabla T_i/T_i) + c_3 (\nabla T_e/T_e)]}{n_e c_2 + n_i} \right), \tag{3b}$$

where $c_1 = D_{ne}/D_{ni}$, $c_2 = \mu_e/\mu_i$ and $c_3 = D_{te}/D_{ti}$.

In this article, axial parameters of diffusion coefficients and mobility are taken into consideration.

2.2 | Charge neutrality

Neutrality of charge may be forwarded as

$$Z_d n_d + n_i - n_e = 0, \tag{4}$$

where dust grains number density is represented by n_d, and n_i and n_e are the ionic and electronic densities under the influence of dust grains, respectively.

2.3 | Dust charge balance

The net charge on dust grains is balanced by charge flux through the surfaces of dust grains, thus

$$\frac{dZ_d}{dt} = n_{ph}(Z_d, T_d) + n_{ic}(Z_d, T_i) - \Upsilon n_{ec}(Z_d, T_e), \tag{5}$$

where $n_{ph}(= \pi r_d^2 n_p)$ is the rate of photoelectric emission by surfaces of charged dust grains, where[37]

$$n_p = \Lambda(\nu)\chi(\nu) = \Lambda(\nu)(4\pi m_e k_B^2 T_d^2/\mathrm{h}^3)\mu(\nu)\Phi_{\mathrm{ph}}(\eta),$$

$\Phi_{ph}(\eta)$ is tabulated integral(Fowler 1936) stated as

$$\Phi_{ph}(\eta) = \int_0^{\exp(\eta)} \frac{\ln(1+\Omega)}{\Omega} d\Omega,$$

$$\eta = (\mathrm{h}\nu - \phi_0)/k_B T_d.$$

n_p indicates the rate of photon density emitted from surface of neutral or negatively charged dust grains, $\Lambda(\nu)$ represent the photon density incident normally to the dust surface, $\chi(\nu)$ is photoelectric efficiency. h represents the Planck's constant, ν indicates the frequency of incident beam, ϕ_0 shows the work function of material of dust grain, T_d represents the temperature of dust grains considered as constant due to large thermal capacity, $\mu(\nu)$ is the probability of absorption of a photon by an electron striking the surface from inside.[38] $n_{ic} = \pi r_d^2 (8k_B T_i/\pi m_i)^{1/2} n_i [1 - Z_d \alpha_i]$ and $n_{ec} = \pi r_d^2 (8k_B T_e/\pi m_e)^{1/2} n_e \exp[Z_d \alpha_e]$ are the ionic and electronic flux charging the dust grain surface, respectively,[29]

$\alpha_e = (e^2/r_d k_B T_e)$, $\alpha_i = (e^2/r_d k_B T_i)$, Υ indicates the adsorption coefficient of electrons on surface of dust grains, r_d is the dust grain radius, k_B is the Boltzmann's constant, T_i and T_e are the ions and electrons temperature, m_i and m_e represent the masses of ions and electrons, respectively.

2.4 | Neutral atoms, electrons, and ion number balance

$$n_i + n_0 = n_{i0} + n_{00} = \mathrm{N}_t, \tag{6}$$

where n_0 is neutral particle density under the influence of dust grains, n_{00}, n_{i0} are the densities of neutral atoms and ions in plasma not having dust grains, and N_t is an invariable, respectively.

$$\frac{dn_e}{dt} = (\beta_i n_0 - \alpha_r n_e n_i) - n_d(\Upsilon n_{ec} - n_{ph}) - \nabla \prod_{\mathrm{p}}, \tag{7}$$

and

$$\frac{dn_i}{dt} = (\beta_i n_0 - \alpha_r n_e n_i) - n_d n_{ic} - \nabla \prod_{\mathrm{p}}, \tag{8}$$

where β_i represent the coefficient of ionization, $\alpha_r = (4.5 \times 10^{-7}(250/T_e)^{-0.7})\ \mathrm{cm^3 s^{-1}}$ is the recombination coefficient of electrons with ions corresponding to O^+ present in ionosphere[1] upto 250 km height, the dominant ionic composition is O^+. Equations (7) and (8) represent the rate of growth of number densities of electrons and ions in plasma under the influence of dust grains, respectively. The initial terms on RHS of Equations (7) and (8) indicate the gain of electrons and ions per unit time per unit volume on account of ionization of neutrals. The succeeding pair of terms in both the equations corresponds to number of neutral atoms formed due to recombination of free electrons into ions, and the net adsorption of electrons and ions on the surface of dust grains.

2.5 | Electrons and ions energy balance

$$\frac{d}{dt}\left(\frac{3}{2} n_e k_B T_e\right) = \beta_i n_0 \varepsilon_e - \alpha_r n_e n_i \left(\frac{3}{2} k_B T_e\right) - n_d(n_{ec}\varepsilon_{ec} - n_{ph}\varepsilon_{ph}) + \frac{n_e e^2 \nu_{ef} U_0^2}{4 m_e(\omega_d^2 + \nu_{ef}^2)} - \nu_{em}\delta_{em}\frac{3}{2} n_e k_B (T_e - T_0) + \nu_{ei}\delta_{ei}\frac{3}{2} n_e k_B (T_e - T_i) \tag{9}$$

and

$$\frac{d}{dt}\left(\frac{3}{2} n_i k_B T_i\right) = \beta_i n_0 \varepsilon_i - \alpha_r n_e n_i \left(\frac{3}{2} k_B T_i\right) - n_d n_{ic}\varepsilon_{ic} - \nu_{in}\delta_{in}\frac{3}{2} n_i k_B (T_i - T_0) + \nu_{ei}\delta_{ei}\frac{3}{2} n_e k_B (T_e - T_i) \tag{10}$$

where ε_e and ε_i are the mean energy of electrons and ions on account of ionization of neutral atoms, $\varepsilon_{ec}(=2k_B T_e - Z_d(e^2/r_d))$ and $\varepsilon_{ic}(=(2 - Z_d\alpha_i)/(1 - Z_d\alpha_i))$ indicate the average energy of adsorption of electrons and ions on the surface of dust grains, $\varepsilon_{ph} = \left(\frac{k_B T_d}{\Theta(\eta - \zeta)} \int_0^\infty 2\zeta \ \ln(1 + \exp(\eta - \zeta))d\zeta\right) - Z_d\alpha_d$ is related to photoemission energy, where Θ $(=Z_d e/r_d)$ is the charged dust grain surface potential, $\alpha_d = (e^2/r_d k_B T_d)$, $\nu_{ef}(=\nu_{ei} + \nu_{ed} + \nu_{en} + \nu_{re} + \nu_{ac})$ is the modified collision frequency owing to ions, dust grains, neutral atoms elastic collisions with electrons and inelastic collisions by virtue of recombination of electrons into ions, and adsorption of electrons on the surface of dust grains, given by $\nu^{[1]}{}_{ei} = \nu_{ei0}(n_i/n_{e0})(T_e/T_{e0})^{-3/2}$, $\nu^{[23]}{}_{\mathrm{ed}} = \nu_{\mathrm{ed0}}(\mathrm{T_e}/\mathrm{T_{e0}})^{-3/2}\mathrm{Z_d^2}$, $\nu^{[1]}{}_{\mathrm{en}} = \nu_{\mathrm{en0}}(\mathrm{n_0}/\mathrm{n_{00}})(\mathrm{T_e}/\mathrm{T_{e0}})^{1/2}$, $\nu_{\mathrm{in}} = \nu_{\mathrm{in0}}(\mathrm{n_0}/\mathrm{n_{00}})(\mathrm{T_i} + \mathrm{T}_d/\mathrm{T_{i0}} + \mathrm{T}_d)^{1/2}$, $\nu_{\mathrm{re}}(=\alpha_{\mathrm{r}}\mathrm{n_i})^{26}$, $\nu_{\mathrm{ac}} = \mathrm{n_d}(\mathrm{n_{ec}}/\mathrm{n_e})$, where $\nu_{ei0} = (5.5 n_{e0}/T_{e0}^{3/2})\ln(220 T_{e0}/n_{e0}^{1/3})$, $\nu_{ed0} = (2.9 \times 10^{-6}) n_{d0} T_{e0}^{-3/2} \ln\Lambda_e$ [$\ln\Lambda_e$(coulomb logarithm) = 10–13], $\nu_{en0} =$

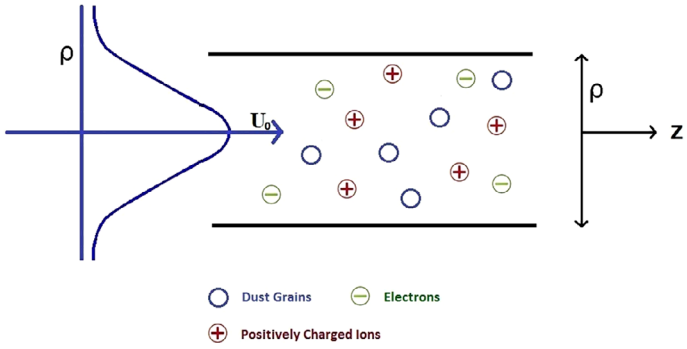


FIGURE 1 Scheme of irradiance of Gaussian electromagnetic beam in dusty plasma

$(8.3\times10^5)\sigma_0 n_{00}T_{e0}^{1/2}$, $\nu_{in0} = (16/3)(2K_BT_d/\pi)^{1/2}\sigma_0 n_{00}(T_{i0}+T_d)^{1/2}$ are the electrons-ions/dust and ion-neutral particles collisions frequencies without dust grains, where n_{e0} is the number density for electrons, T_{e0} is the temperature of electrons without dust grains. $\sigma_0(=6.6\times10^{-15}[(T_e/4)-0.1/\{1+(T_e/4)^{1.6}\}])$ is the electrons and neutral particles collisional cross-section area, where σ_0 is in cm^2 and T_e is in eV. The initial pair of terms of Equations (9) and (10) correspond to average energy gain on account of ionization and recombination processes. The third term in both the equations refers to net energy loss due to adsorption process, and the succeeding term in Equation (9) belongs to power loss due to process of photoemission from the outer layer of dust grains. The fifth term in Equation (9) corresponds to electrons energy gain from the field due to Ohmic heating effect. Ohmic heating of heavier ions can be ignored. The last term in the two describes the rate of energy loss per unit volume attributed to collisions between dusty plasma components. The term regarding thermal conduction can be ignored on account of predominant collisional effect.

2.6 | Evaluation of ionization coefficient and average energy of electrons and ions

By applying electron density and energy balance equations in dust-free plasma, ionization coefficient and average energies of electrons and ions can be evaluated. The initial condition for them are given as

$$\beta_i n_{00} = \alpha_r n_{e0} n_{i0} = \alpha_r n_{e0}^2, \tag{11}$$

$$\varepsilon_e = \frac{d}{dt}\left(\frac{3}{2}k_BT_{e0}\right) + (\alpha_{r0}n_{e0})^{-1}\left[\nu_{en}\delta_{en}\frac{3}{2}k_B(T_{e0}-T_0) + \nu_{ei0}\delta_{ei}\frac{3}{2}k_B(T_{e0}-T_{i0})\right], \tag{12a}$$

and

$$\varepsilon_i = \frac{d}{dt}\left(\frac{3}{2}k_BT_{i0}\right) + (\alpha_{r0}n_{e0})^{-1}\left[\nu_{in}\delta_{in}\frac{3}{2}k_B(T_{i0}-T_0) + \nu_{ei0}\delta_{ei}\frac{3}{2}k_B(T_{e0}-T_{i0})\right]. \tag{12b}$$

2.7 | Paraxial approximation approach

Let us suppose that the electromagnetic beams with Gaussian and Sine time irradiance spreading along their wave front. Cylindrical coordinate system is suitable to study for their propagation, which may be considered along the z-axis. Propagation of Gaussian irradiance is shown in Figure 1.

The spatial irradiance of the beam along z = 0 is described by

$$U^2 = U_0^2(t)e^{(-\rho^2/\rho_0^2)}\ , \tag{13}$$

where U^2 is the electric field intensity of the Gaussian beam at a length ρ from axis, U_0^2 is the axial intensity at $\rho = 0$, whereρ is the radial coordinate of cylindrical system. Major analysis of the theory assumes the time irradiance of Gaussian function of the form.

$$U_0^2(t) = U_{00}^2F'(t/t_0), \tag{13a}$$

where t_0 is the beam time width.

$$F'(t/t_0) = \begin{cases} Sin(\pi - t/t_0) & [0 < t/t_0 < \pi] \quad \text{for Sine beam} \\ 0, \text{orelse} \end{cases}, \tag{13b}$$

$$F'(t/t_0) = \exp(-t^2/t_0^2) \ [-\infty < t/t_0 < \infty] \text{ for Gaussian beam} \tag{13c}$$

For a defined value of z, the irradiance in paraxial approximation is represented as

$$U_0^2 = \frac{U_{00}^2(z,t)}{f^2(z,t)} e^{(-\rho^2/\rho_0^2 f^2(z,t))} F'(t/t_0), \tag{14}$$

where $f(z, t)$ is the beam waist parameter, which depends upon the length of transmission, z and time t, and $\rho_0 f$ represent the spot of the beam.

From above statement, it is suggested that the major intensity of beam survives along its principal axis. Therefore, Gaussian beam can be analysed by the paraxial approximation approach i.e., $\rho/\rho_0 f \ll 1$. So, all the applicable variables are expanded along the principal part ($\rho = 0$) only up to the quadratic expression ρ^2/ρ_0^2 as

$$P(\rho, z, t) = P_a(z, t) - (\rho^2/\rho_0^2) P_\rho(z, t), \tag{15}$$

where P_a and P_ρ are the parts of all the accordant parameters characterized by P(ρ,z), along axis and radius, respectively. Although the spatial irradiance of the Gaussian beams do not modify during its propagation in dusty plasma, by introducing the accordant parameters and the dimensionless variable $t/t_0 = \tau$ in kinetic Equations (5), (7)–(10), and comparing the coefficients of ρ^0 and ρ^2/ρ_0^2 terms, we can write the above equations as

2.8 | Dust charge balance

$$\frac{dZ_{da}}{d\tau} = t_0[n_{ica}(Z_d, T_i) + n_{pha}(Z_d, T) - \Upsilon n_{eca}(Z_d, T_e)] \tag{5a}$$

and

$$\frac{dZ_{d\rho}}{d\tau} = t_0[n_{ic\rho}(Z_d, T_i) + n_{ph\rho}(Z_d, T) - \Upsilon n_{ec\rho}(Z_d, T_e)]. \tag{5b}$$

2.9 | Electrons and ions number balance

$$\begin{aligned} \frac{dn_{ea}}{d\tau} &= t_0[\ \beta_{ia} n_0 - \alpha_{ra} n_{ea} n_{ia} - n_d(\Upsilon n_{eca} - n_{pha})] \\ &\quad - \frac{(2/\rho_0^2)\, t_0}{n_{ia} + c_2 n_{ea}} [D_{1i}(c_1 n_{ia} n_{e\rho} + c_2 n_{ea} n_{i\rho}) + D_{2i} n_{ea} n_{ia}(c_2 T_{ia}^{-1}\ T_{i\rho} + c_3 T_{ea}^{-1}\ T_{e\rho})] \end{aligned} \tag{6a}$$

and

$$\begin{aligned} \frac{dn_{e\rho}}{d\tau} &= t_0[\beta_{i\rho} n_0 - \alpha_{r\rho} n_{ea} n_{ia} - n_d(\Upsilon n_{ec\rho} - n_{pha})] \\ &\quad - \frac{(6/\rho_0^2) t_0}{n_{ia} + c_2 n_{ea}} \begin{bmatrix} D_{1i}(c_1 + c_2) n_{i\rho} n_{e\rho} - \frac{(c_1 n_{ia} n_{e\rho} + c_2 n_{ea} n_{i\rho})(n_{i\rho} + c_2 n_{e\rho})}{n_{ia} + c_2 n_{ea}} \\ + D_{2i}\left[(n_{ea} n_{i\rho} + n_{ia} n_{e\rho})(c_2 T_{ia}^{-1}\ T_{i\rho} + c_3 T_{ea}^{-1}\ T_{e\rho}) \right. \\ \left. - n_{ea} n_{ia}(c_2 T_{ia}^{-2}\ T_{i'\rho}^{2} + c_3 T_{ea}^{-2}\ T_{e\rho}^{2}) \right. \\ \left. - \frac{n_{na} n_{ia}(c_2 T_{ia}^{-1}\ T_{i\rho} + c_3 T_{ea}^{-1}\ T_{e\rho})(n_{i\rho} + c_2 n_{e\rho})}{n_{ia} + c_2 n_{ea}} \right] \end{bmatrix}. \end{aligned} \tag{6b}$$

$$\frac{dn_{ia}}{d\tau} = t_0[\ \beta_{ia}n_0 - \alpha_{ra}n_{ea}n_{ia} - n_d n_{ica}] - \frac{(2/\rho_0^2)}{n_{ia}+c_2n_{ea}}[D_{1i}(c_1 n_{ia}n_{e\rho} + c_2 n_{ea}n_{i\rho}) + D_{2i}n_{ea}n_{ia}(c_2 T_{ia}^{-1}\ T_{i\rho} + c_3 T_{ea}^{-1}\ T_{e\rho})] \quad (7a)$$

and

$$\frac{dn_{i\rho}}{d\tau} = t_0[\beta_{i\rho}n_0 - \alpha_{r\rho}n_{ea}n_{ia} - n_d n_{ic\rho}] - \frac{(6/\rho_0^2)\ t_0}{n_{ia}+c_2n_{ea}}\begin{bmatrix} D_{1i}(c_1+c_2)n_{i\rho}n_{e\rho} - \frac{(c_1n_{ia}n_{e\rho}+c_2n_{ea}n_{i\rho})(n_{i\rho}+c_2n_{e\rho})}{n_{i'a}+c_2n_{e'a}} \\ +D_{2i}\left[(n_{ea}n_{i\rho}+n_{ia}n_{e\rho})(c_2T_{ia}^{-1}\ T_{i\rho}+c_3T_{ea}^{-1}\ T_{e\rho})\right. \\ -n_{ea}n_{ia}(c_2T_{ia}^{-2}\ T_{i\rho}^{2}+c_3T_{ea}^{-2}\ T_{e\rho}^{2}) \\ \left. -\frac{n_{ea}n_{ia}(c_2T_{ia}^{-1}\ T_{i\rho}+c_3T_{ea}^{-1}\ T_{e\rho})(n_{i\rho}+c_2n_{e\rho})}{n_{ia}+c_2n_{ea}}\right] \end{bmatrix}. \quad (7b)$$

2.10 | Electrons and ions energy balance

$$\begin{aligned}\frac{d}{d\tau}\left(\frac{3}{2}n_{ea}k_BT_{ea}\right) &= t_0\left[\beta_{ia}n_0\varepsilon_e - \alpha_{ra}n_{ea}n_{ia}\left(\frac{3}{2}k_BT_{ea}\right) - n_d(\Upsilon n_{eca}\varepsilon_{eca} - n_{pha}\varepsilon_{pha})\right] \\ &\quad - \delta_{en}\nu_{ena}t_0\frac{3}{2}n_{ea}k_B(T_{ea}-T_0) + \delta_{ei}\nu_{eia}t_0\frac{3}{2}n_{ea}k_B(T_{ea}-T_{ia}) \\ &\quad + \delta_{comp}\nu_{efa}t_0\left(\frac{3}{2}k_BT_0\right)\frac{\beta_0U_{00}^2}{f^2}n_{ea}\end{aligned} \quad (9a)$$

and

$$\begin{aligned}\frac{d}{d\tau}\left(\frac{3}{2}k_B(n_{ea}T_{e\rho}+n_{e\rho}T_{ea})\right) &= t_0\begin{bmatrix}\beta_{i\rho}n_0\varepsilon_e - \frac{3}{2}k_B(\alpha_{r\rho}n_{ea}n_{ia}T_{ea} + \alpha_{ra}(n_{ea}n_{ia}T_{e\rho}+n_{ea}n_{i\rho}T_{ea}+n_{e\rho}n_{ia}T_{ea})) \\ -n_d(\Upsilon(n_{ec\rho}\ \varepsilon_{eca}+n_{eca}\varepsilon_{ec\rho}) - (n_{pha}\ \varepsilon_{ph\rho}+n_{ph\rho}\varepsilon_{pha}))\end{bmatrix} \\ &\quad - \frac{3}{2}k_B\delta_{en}t_0[(n_{e\rho}\nu_{ena}+n_{ea}\nu_{en\rho})(T_{ea}-T_0)+n_{ea}\nu_{ena}T_{e\rho}] \\ &\quad - \frac{3}{2}k_B\delta_{ei}t_0[(n_{e\rho}\nu_{eia}+n_{ea}\nu_{ei\rho})(T_{ea}-T_{ia})+n_{ea}\nu_{eia}(T_{e\rho}-T_{i\rho})] \\ &\quad + \left(\frac{3}{2}\delta_{comp}t_0k_B\ T\right)\frac{\beta_0U_{00}^2}{f^2}\left(\frac{n_{ea}\nu_{efa}}{f^2}+(n_{e\rho}\ \nu_{efa}+n_{ea}\nu_{ef\rho})\right)\end{aligned} \quad (9b)$$

where $\beta_0 = \frac{e^2}{3m_e\delta_{comp}k_BT\omega^2}$ and $\delta_{comp}\ (= \nu_{ac}/\nu_{ef})$[22, 39] is the fractional loss of collisional energy of electron with the dusty plasma components,

$$\begin{aligned}\frac{d}{d\tau}\left(\frac{3}{2}n_{ia}k_BT_{ia}\right) &= t_0\left[\beta_{ia}n_0\varepsilon_i - \alpha_{ra}n_{ea}n_{ia}\left(\frac{3}{2}k_BT_{ia}\right) - \Upsilon n_d n_{ica}\varepsilon_{ica}\right] \\ &\quad - \left[\delta_{in}\nu_{ina}t_0\frac{3}{2}n_{ia}k_B(T_{ia}-T_0) + \delta_{ei}\nu_{eia}t_0\frac{3}{2}n_{ea}k_B(T_{ea}-T_{ia})\right]\end{aligned} \quad (10a)$$

and

$$\begin{aligned}\frac{d}{d\tau}\left(\frac{3}{2}k_B(n_{ia}T_{i\rho}+n_{i\rho}T_{ia})\right) &= t_0\begin{bmatrix}\beta_{i\rho}n_0\varepsilon_i - \frac{3}{2}k_B(\alpha_{r\rho}n_{ea}n_{ia}T_{ia} + \alpha_{ra}(n_{ea}n_{ia}T_{i\rho}+n_{ea}n_{i\rho}T_{ia}+n_{e\rho}n_{ia}T_{ia})) \\ -n_d(\Upsilon(n_{ic\text{æ}}\ \varepsilon_{ica}+n_{ica}\varepsilon_{ic\text{æ}}))\end{bmatrix} \\ &\quad - \frac{3}{2}k_B\delta_{in}t_0[(n_{i\rho}\nu_{ina}+n_{ia}\nu_{in\rho})(T_{ia}-T_0)+n_{ia}\nu_{ina}T_{i\rho}] \\ &\quad - \frac{3}{2}k_B\delta_{ei}t_0[(n_{e\rho}\nu_{eia}+n_{ea}\nu_{ei\rho})(T_{ea}-T_{ia})+n_{ea}\nu_{eia}(T_{e\rho}-T_{i\rho})].\end{aligned} \quad (10b)$$

The simultaneous solution of Equations (5)–(10) for the stationary state (i.e., at $\tau \to \infty$) collectively provide the variation of different dusty plasma components along axis and radius, respectively, for both Gaussian and Sine time irradiance.

2.11 | Modified dielectric permittivity

Since pressure gradient of electrons and ionic gas is observed on account of nonuniform distribution of electrons and ionic temperature due to nonuniform irradiance of electromagnetic beam. Therefore, these gradients are balanced by space charge field in steady state. Thus, one can obtain

$$\frac{n_e}{n_{e0}} = \left(\frac{2T_{e0}}{T_e + T_{e0}}\right). \tag{16}$$

the modified dielectric permittivity in dust accompanying plasma may be given by[35,36,40]

$$\varepsilon_{ef} = \left(\frac{\omega_0^2}{\omega_d^2}\right)\frac{\omega_d^2}{(\nu_{ef}^2 + \omega_d^2)}\left(\frac{n_e}{n_{e0}}\right) = \varepsilon_a - \left(\rho^2/\rho_0^2\right)\varepsilon_\rho, \tag{16a}$$

$$\varepsilon_{ef} = \left(\frac{\omega_0^2}{\omega_d^2}\right)\frac{\omega_d^2}{(\nu_{ef}^2 + \omega_d^2)}\left(\frac{2T_{e0}}{T_e + T_{e0}}\right) = \varepsilon_a - \left(\rho^2/\rho_0^2\right)\varepsilon_\rho. \tag{16b}$$

Substituting T_e and ν_{ef} corresponding to the Equation (15) and comparing the coefficient of ρ^0 and ρ^2/ρ_0^2 terms, one obtain,

$$\varepsilon_a = 1 - \left(\frac{\omega_0^2}{\omega_d^2}\right)\left(1 - \frac{\nu_{efa}^2}{\omega_d^2}\right)\left(\frac{2T_{e0}}{T_{ea} + T_{e0}}\right) \tag{16c}$$

and

$$\varepsilon_\rho = \left(\frac{\omega_0^2}{\omega_d^2 + \nu_{efa}^2}\right)\left(\frac{2T_{e0}}{T_{ea} + T_{e0}}\right)\left(\frac{\nu_{efa}\nu_{ef\rho}}{\omega_d^2 + \nu_{efa}^2} - \frac{T_{e\rho}}{T_{ea} + T_{e0}}\right), \tag{16d}$$

where ε_a, ε_ρ, ν_{efa}, $\nu_{ef\rho}$ and T_{ea}, $T_{e\rho}$ are the modified dielectric permittivities, collision frequencies, and temperatures along axis and radius, respectively, under the influence of dust grains. Plasma frequencies represented by the term ω_0 in the absence of dust grains, whereas ω_d represents the vice versa. On introducing ν_{efa}, $\nu_{ef\rho}$, T_{ea}, *and* $T_{e\rho}$ obtained by the coupled solution of Equations (5)–(10), we get the ε_a and ε_ρ as a function of $\beta_0 U_{00}^2$.

3 | PROPAGATION OF ELECTROMAGNETIC BEAM WITH TIME IRRADIANCE IN DUSTY PLASMA

The irradiance of the Gaussian EM-beam in space and time is given by

$$\vec{U} = \hat{j}\, U_0(\rho, z, t)e^{i\omega_d t}, \tag{17}$$

where z is the direction of propagation and y is the direction of polarization represented by $\hat{j}$. The analogous equation corresponds to Equation (13),

$$\nabla^2\vec{U} - \nabla(\nabla\cdot\vec{U}) + \frac{\varepsilon_{ef}}{c^2}\frac{\partial^2\vec{U}}{\partial t^2} = 0. \tag{18}$$

As the transverse component of a beam has zero divergence, the Equation (18) reduces to

$$\nabla^2 U + (\omega_d^2/c^2)\varepsilon_{ef}U = 0. \tag{19}$$

Equation (19) can be represented in the following manner because of cylindrically symmetric nature of beam

$$\frac{\partial^2 U}{\partial z^2} + \frac{1}{\rho}\frac{\partial U}{\partial \rho} + \frac{\partial^2 U}{\partial \rho^2} = -\frac{\omega_d^2\varepsilon_{ef}}{c^2}U. \tag{20}$$

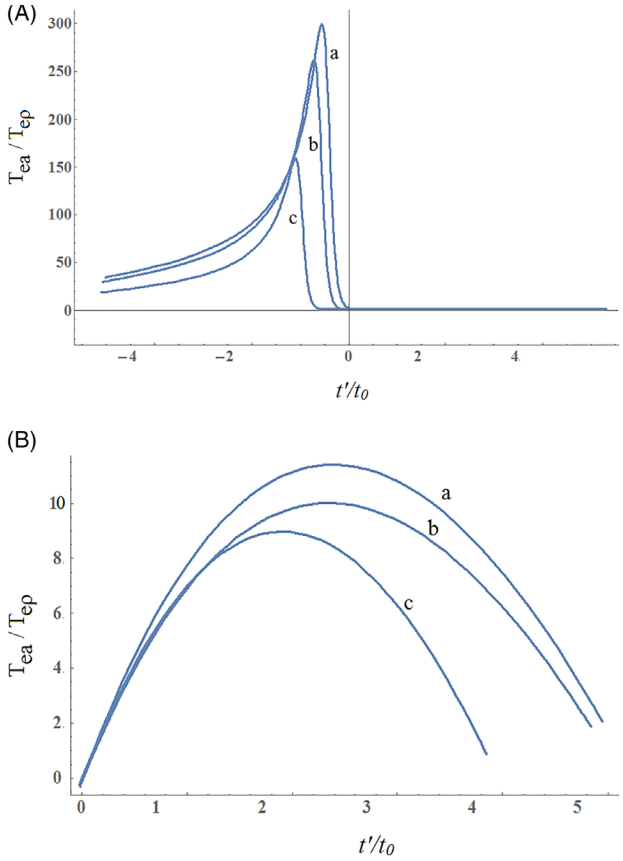


FIGURE 2 (a, b) Dependence of dimensionless electron temperature $T_{ea}/T_{e\rho}$ on dimensionless time width t'/t_0 for distinguishable dust densities for both Gaussian and Sine profile, respectively. The dust densities $n_d = 10, 100, 1000$ in cm^{-3} are labelled by a, b, and c, respectively

The solution of Equation (20) by Akhmanov et al.[35] can be described as

$$U(\rho, z, t) = A_0(\rho, z, t)e^{i(\omega_d t - kz)}, \tag{21}$$

$$A_0(\rho, z, t) = A_{00}(\rho, z, t)e^{-ik\mathbb{S}(\rho,z,t)}, \tag{22}$$

where $k = (\omega_d/c)\,\varepsilon_a^{1/2}$, A_{00} and $\mathbb{S}$ are real quantities, where $\mathbb{S}$ is called eikonal which represents the wave phase.

Measuring the real and imaginary parts of the resulting equations, obtained from putting the values of U and A_0, respectively, from Equations (21) and (22) in Equation (20) and pursuing Sodha et al.,[36] we get

$$2\frac{\partial \mathbb{S}}{\partial z} + \left(\frac{\partial \mathbb{S}}{\partial \rho}\right)^2 + \frac{2}{V_g}\frac{\partial \mathbb{S}}{\partial t} = \frac{1}{\varepsilon_a}\Phi(t) + \frac{1}{k^2 A_{00}}\left\{\frac{\partial^2 A_{00}}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial A_{00}}{\partial \rho}\right\}, \tag{23}$$

and

$$\frac{\partial A_{00}^2}{\partial z} + \frac{\partial \mathbb{S}}{\partial \rho}\frac{\partial A_{00}^2}{\partial \rho} + A_{00}^2\left(\frac{\partial^2 \mathbb{S}}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial \mathbb{S}}{\partial \rho}\right) + \frac{1}{V_g}\frac{\partial A_{00}^2}{\partial t} = 0, \tag{24}$$

where

$$V_g = c{\varepsilon_a}^{1/2}/(\varepsilon_a + \omega_d/2(\partial\varepsilon_a/\partial\omega_d)) \tag{24a}$$

is the group velocity of electromagnetic beam in dusty plasma,
and

$$\Phi(t) = \frac{\varepsilon_\rho e^{-t/\tau'}}{\tau'}\int_{-\infty}^{t} e^{t/\tau'} AA * dt. \tag{24b}$$

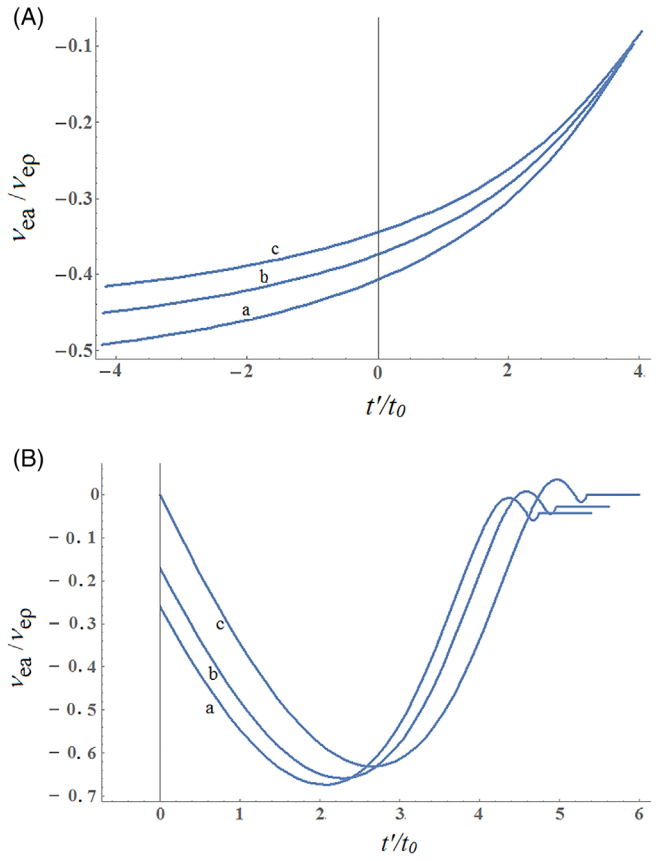


FIGURE 3 (a, b) Dependence of dimensionless electron frequency $\nu_{ea}/\nu_{e\rho}$ on dimensionless time width t'/t_0 for distinguishable dust densities for Gaussian and Sine time profile, respectively. The dust densities, $n_d = 10, 100, 1000$ in cm^{-3} are labelled by a, b, and c, respectively

On transforming the variable as $z' = z$ and $t' = t - z/V_g$, Equations (23) and (24) reduce to the form

$$2\frac{\partial \mathbb{S}}{\partial z'} + \left(\frac{\partial \mathbb{S}}{\partial \rho}\right)^2 = \frac{1}{\varepsilon_a}\Phi(t') + \frac{1}{k^2 A_{00}}\left\{\frac{\partial^2 A_{00}}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial A_{00}}{\partial \rho}\right\}, \tag{25}$$

$$\frac{\partial A_{00}^2}{\partial z'} + \frac{\partial \mathbb{S}}{\partial \rho}\frac{\partial A_{00}^2}{\partial \rho} + A_{00}^2\left(\frac{\partial^2 \mathbb{S}}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial \mathbb{S}}{\partial \rho}\right) = 0. \tag{26}$$

Supporting the earlier observations,[35,40,41] the solution of Equations (25) and (26) can be expressed as

$$A_{00}^2 = \frac{U_{00}^2(t')}{f^2(z', t')} e^{(-\rho^2/\rho_0^2 f^2)} F(t'/t_0), \tag{27}$$

Substituting the expression of $\Phi(t')$ and A_{00}^2 from Equation (24b) and Equation (27), respectively, in Equation (25), we obtain

$$2\frac{\partial \mathbb{S}}{\partial z'} + \left(\frac{\partial \mathbb{S}}{\partial \rho}\right)^2 = \frac{1}{2}\frac{\Phi(t')}{\varepsilon_a} - \frac{2}{k^2\rho_0^2 f^2} + \frac{\rho^2}{k^2\rho_0^4 f^4}. \tag{28}$$

The eikonal function $\mathbb{S}(\rho, z', t')$ in paraxial ray approximation approach, may be expressed as

$$\mathbb{S}(\rho, z', t') = \frac{\rho^2}{2}\mathrm{B}(z', t') + \Psi(z', t'), \tag{29}$$

where $\Psi(z', t')$ is any function of z'.

and t' and $\mathrm{B}(z', t') = \frac{1}{f(z',t')}\frac{\partial f(z',t')}{\partial z'}$.

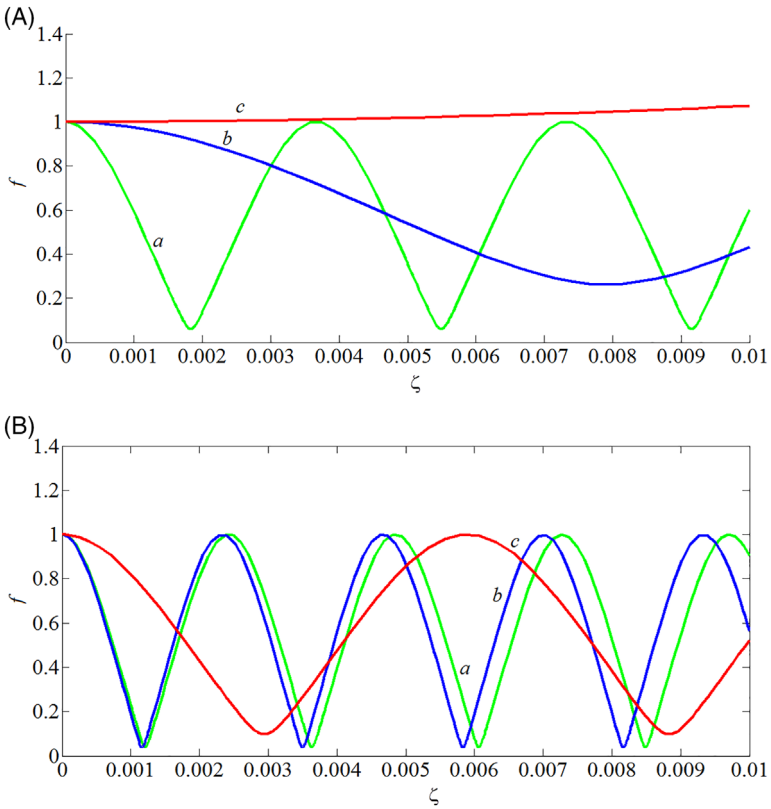


FIGURE 4 (a, b) Relying beam waist parameter f with the dimensionless length of transmission $\zeta(=z'/R_d)$ for the Gaussian and Sine beam for different time width t'/t_0. The labels a, b, and c indicate the dimensionless time width $t'/t_0 = 1$, 2, and 3, respectively

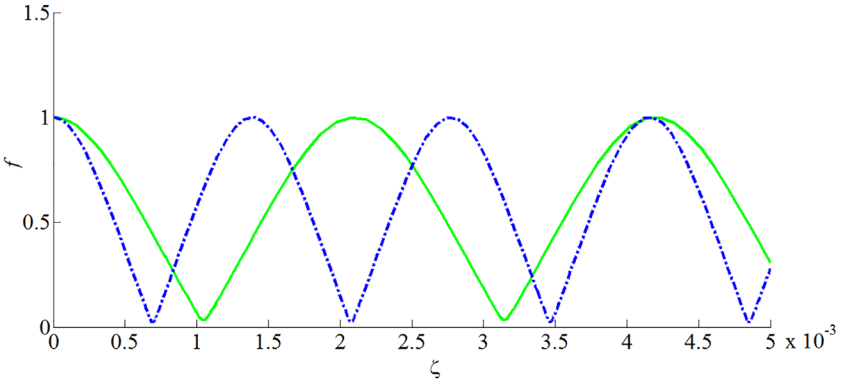


FIGURE 5 Relying beam waist parameter f on the dimensionless length of transmission $\zeta(=z'/R_d)$ for the Gaussian (solid) and Sine (dotted) beam as a function of maximal irradiance at time width $t'/t_0 = -0.5$, for Gaussian and $t'/t_0 = 0.8\pi$ for Sine beam

When beam propagates through dusty plasma, the solution of A_{00}^2 preserves the power and shape of the beam. Employing the expression for B in Equation (29) and substituting $\mathbb{S}$ from Equation (29) in Equation (28) and equating the co-efficient of ρ^2/ρ_0^2 the resulting equation (using the paraxial ray approximation), variation of beam waist parameter with the dimensionless length of transmission can be represented by the equation for both the profiles as

$$\frac{1}{f}\frac{\partial^2 f}{\partial z'^2} = \frac{1}{R_d^2}\frac{1}{f^4} - \frac{1}{2\rho_0^2}\frac{\varepsilon_\rho e^{-t'/\tau'}}{\varepsilon_a}\int_{-\infty}^{t'}\frac{U_{00}^2(t')e^{t'/\tau'}}{f^4}F'(t'/t_0)dt'. \tag{30}$$

Integral sign may be eliminated in the Equation (30), if variation of $F'(t'/t_0)$ is of the order of relaxation time τ, which causes small nonlinearity. Then the Equation (30) can be simplified to

$$\frac{d^2 f}{dz'^2} = \frac{1}{R_d^2}\frac{1}{f^3} - \frac{1}{2\rho_0^2}\frac{\varepsilon_\rho}{\varepsilon_a}\frac{U_{00}^2(t')}{f^3}F'(t'/t_0) \tag{31}$$

or

$$\frac{d^2f}{d\zeta^2} = -\frac{1}{R_{sl}^2}\frac{1}{f^3} + \frac{1}{f^3}, \tag{32}$$

where $R_{sl}\left(=\frac{2\varepsilon_a\rho_0^2}{\varepsilon_\rho U_{00}^2 F'(t'/t_0)}\right)^{1/2}$ is characterized by self-focusing length, $\zeta = \frac{z'}{R_d}$ and $R_d = k\rho_0^2$. On the right-hand side of Equation (32), the initial term corresponds to the nonlinear self-focusing of the beam whereas the succeeding term indicates the divergence phenomenon.

4 | RESULTS AND DISCUSSION

In this analytical model, we present the comparative analysis of both Gaussian and Sine time irradiance of the beam propagating in plasma under the influence of dust grains in collisional plasma. Neutral ionization, recombination of electrons and ions with neutrals, elastic and charging collisions, charging flux via adsorption and photoemission, and ohmic heating due to electromagnetic beam propagating in dusty plasma are the basic energy exchange methods considered in this analysis. Numerical calculations have been carried out in accordance to the metallic dust cloud of stainless steel having high work function of 7.8 eV within close range of space where the photoemission is the dominant mechanism, which corresponds to Lyman alpha radiations with the spectrum of 1215.7 $\mathring{A}$. The simultaneous solution of the coupled Equations (5)–(10) for both the Gaussian and Sine character of electromagnetic beam gives the dependence of modified electron temperature and collision frequency on time for distinguishable dust densities using MATLAB and MATHEMATICA software. Employing the relevant parameter to get ε_a and ε_ρ. Also, at the boundary condition f = 1, $\frac{df}{d\zeta} = 0$ at $\zeta = 0$, we obtain dependence of beam waist parameter upon dimensionless length of transmission, $\zeta(=z'/R_d)$ by substituting ε_a and ε_ρ, for different values of time width, dust densities, and collision frequency for both Gaussian and Sine time irradiance. The set of variables used in this analysis is listed below:$n_{e0} = n_{i0} = 10^6 cm^{-3}$, $n_0 = n_{00} = 2.5\times10^9 cm^{-3}$, $n_d = 10^3 cm^{-3}$, $k_B = 1.38\times10^{-16}\text{ergsk}^{-1}$, $T_{e0} = 1700K$, $m_e = 9.1\times10^{-28}g$, $T_{i0} = 1300K$, $m_i = 16\times1.6\times10^{-24}g$, $\rho_0 = 5.5\times10^{-2}\text{cm}$, $\omega_d = 1.1\times10^8\text{rads}^{-1}$, $e = 4.8\times10^{-10}\text{statcoloumb}$, $T_d = 275\text{K}$, $r_d = 5\times10^{-4}\text{cm}$, $\sigma_0 = 1.5\times10^{-16}\text{cm}^2$, $U_{00} = 10^6$ stat volt cm^{-1}.

The time sensitivity of dimensionless electron temperature has been shown in Figure 2a,b for variant dust densities for both Gaussian and Sine time characteristics, respectively. Peak of temperature corresponding to extremum irradiance at $t'/t_0 = (-0.5)$ for Gaussian pulse and 0.8π for Sine pulse is observed. A shift is observed from the dust-free environment where the maxima is obtained at $t'/t_0 = 0$ (red shift) for Gaussian pulse and $t'/t_0 = 0.5\pi$ (blue shift) for Sine pulse[34] on account of increase in collision frequency due to the presence of dust. Also the maximal temperature decreases with increase in dust density on account of faster electron energy decay due to increase in collisions with the density of dust grains as shown in Figure 3a,b, which represent the variation of dimensionless modified collision frequency with time width for both the Gaussian and Sine time profiles for different dust densities, respectively.

Figure 4a,b show the comparative analysis of propagation of beam waist parameter with the dimensionless length at the same time width for both the Gaussian and Sine time irradiance of the beam. Self-focusing enhances with lesser time width for both the profiles but beam get self-focused faster in sinusoidal irradiance due to boost rate of nonlinearity under the influence of dust grains.

Figure 5 show the reliance of beam waist parameter with the dimensionless length of transmission at the same modified collision frequency and at the point of time width of maximal irradiance under the influence of dust grains shown by the solid (Gaussian) and dotted (Sine) lines. Both the exponential and Sine profiles show the oscillatory convergence but the frequency of Sine profile is greater than the Gaussian one on account of different rate of nonlinearity for both the profiles under the influence of dust grains. Even the waist of the beam becomes narrower as compared to the dust-free environment.[34]

Figures 6a,b describe reliance of beam waist parameter on the dimensionless length of transmission for distinguishable dust densities. Since the collision frequency increases with dust density which further enhances the nonlinearity of the beam causing the enhanced self-focusing for both the profiles of the beam with the dust density but the Sine profile self-focused at faster rate than Gaussian profile.

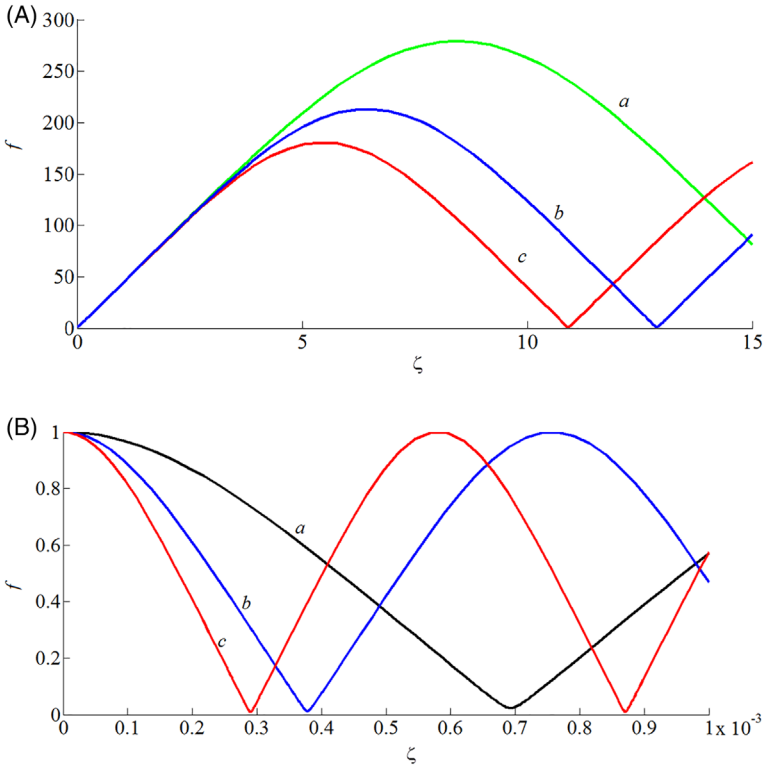


FIGURE 6 (a, b) Relying beam waist parameter f on the dimensionless length of transmission $\zeta(=z'/R_d)$ for the Gaussian and Sine beam for distinguishable dust densities, respectively. The dust densities, $n_d = 10$, 100, 1000 in cm^{-3} are labelled by a, b, and c, respectively

5 | CONCLUSIONS

This analytical model investigates the dust grain impact on self-focusing of both the Gaussian and Sine time irradiance of the beam transmitting through dusty plasma. A comparative analysis of growth of plasma parameters like temperature and collision frequencies modified under the influence of dust grains have been studied numerically to calculate the modified dielectric permittivity for both the profiles of the beam considering the paraxial ray approximation, neutrality of charge, ionization of neutrals, recombination of electrons with ions, charging flux via adsorption and photoemission of electrons, and diffusion. Also, the comparative growth of beam waist parameter along the dimensionless transmission length as a function of time width, point of maximal irradiance, and dust densities have been analysed for both the profiles in the impact of dust grains and compared with the dust-free environment. The consequences of this model are beneficial for the domain of dust particles like astrophysics, manufacturing, fabrication, plasma processing etc.

ACKNOWLEDGMENTS

The authors acknowledge to Prof. M. S. Sodha for his valuable discussions. The author (Ruchi Sharma) is thankful to her colleagues for their supportive approach to accomplish the research article.

REFERENCES

[1] A. V. Gurevich, Nonlinear Phenomena in the Ionosphere, Springer, New York **1978**.
[2] A. Gahlot, R. Walia, S. C. Sharma, R. P. Sharma, *J. Plasma Phys.* **2012**, *78*, 33.
[3] M. S. Sodha, A. Dixit, S. Srivastava, *Appl. Phys. Lett.* **2009**, *94*, 251501.
[4] M. S. Sodha, A. Dixit, *Appl. Phys. Lett.* **2009**, *95*, 101502.
[5] M. S. Sodha, S. Misra, S. K. Mishra, *Phys. Plasmas* **2009**, *16*, 123705.
[6] J. Goree, *Plasma Sources Sci. Technol.* **1994**, *03*, 400.
[7] M. Mikikian, L. Boufendi, A. Bouchoule, H. M. Thomas, G. E. Morfill, A. P. Nefedov, V. E. Fortov, PKE–Nefedov team, *New J. Phys.* **2003**, *5*, 19.
[8] I. Denysenko, J. Berndt, E. Kovacevic, I. Stefanovic, V. Selenin, J. Winter, *Phys. Plasmas* **2006**, *13*, 073507.
[9] D. Samsonov, J. Goree, *Phys. Rev. E* **1999**, *59*, 1047.
[10] P. K. Shukla, V. P. Silin, *Phys. Scr.* **1992**, *45*, 508.
[11] R. L. Merlino, A. Barkan, C. Thompson, N. D'Angelo, *Phys. Plasmas* **1998**, *5*, 1607.
[12] M. Bonitz, C. Henning, D. Block, *Rep. Prog. Phys.* **2010**, *73*, 6.

[13] S. V. Vladimirov, K. Ostrikov, A. A. Samarian, Physics and Applications of Complex Plasmas, Imperial College Press, London **2005**.
[14] V. N. Tsytovich, G. E. Morfill, S. V. Vladimirov, H. M. Thomas, Elementary Physics of Complex Plasmas, Springer, Berlin **2008**.
[15] O. Ishihara, *J. Phys. D: Appl. Phys.* **2007**, *40*, 8.
[16] H. Kersten, G. Thieme, M. Fröhlich, D. Bojic, D. H. Tung, M. Quaas, H. Wulff, R. Hippler, *Pure Appl. Chem.* **2005**, *77*, 415.
[17] F. Verheest, *Plasma Phys. Control. Fusion* **1999**, *41*, 3A.
[18] R. L. Merlino, Plasma Physics Applied, Transworld Research Network, Kerala, India **2006**, **2006**.
[19] S. C. Sharma, A. Gahlot, R. P. Sharma, *Phys. Plasmas* **2008**, *15*, 043701.
[20] M. S. Sodha, S. K. Mishra, S. Misra, *Phys. Plasmas* **2009**, *16*, 123701.
[21] M. S. Sodha, Kinetics of Complex Plasmas, Springer, India **2014**.
[22] G. Sambandan, V. K. Tripathi, J. Parashar, R. Bharuthram, *Phys. Plasmas* **1999**, *6*, 762.
[23] M. S. Sodha, S. K. Mishra, S. Misra, *Phys. Plasmas* **2011**, *18*, 023701.
[24] F. Verheest, P. Meuris, *Phys. Lett. A* **1996**, *210*, 198.
[25] J. Jia, C. Yuan, R. Gao, Y. Wang, Y. Liu, J. Gao, Z. Zhou, X. Sun, J. Wu, H. Li, S. Pu, *J. Phys. D: Appl. Phys.* **2015**, *48*, 465201.
[26] S. K. Mishra, S. Misra, *Phys. Plasmas* **2019**, *26*, 023702.
[27] R. Sharma, S. C. Sharma, *Contrib. Plasma Phys.* **2019**, *59*, 211.
[28] J. Jia, C. Yuan, S. Liu, F. Yue, R. Gao, Y. Wang, Z.-X. Zhou, J. Wu, H. Li, *Phys. Plasmas* **2016**, *23*, 043302.
[29] S. K. Mishra, S. Misra, M. S. Sodha, *Phys. Plasmas* **2011**, *18*, 043702.
[30] H.-b. Wang, Y.-y. Chen, H. Chung, *Optik-Int. J. Light Electron. Opt.* **2019**, *196*, 163148.
[31] M. S. Sodha, S. K. Mishra, S. Misra, S. Srivastava, *Phys. Plasmas* **2010**, *17*, 073705.
[32] M. Faisal, L. Bhasin, M. S. Sodha, *J. Geophys. Res.* **2009**, *114*, A01305.
[33] M. Sodha, A. Sharma, *J. Plasma Phys.* **2009**, *74*, 473.
[34] M. Faisal, M. P. Verma, M. S. Sodha, *Phys Plasmas* **2008**, *15*, 102301.
[35] S. A. Akhamanov, A. P. Sukhorukov, R. V. Khokhlov, *Sov. Phys. Usp.* **1968**, *10*, 619.
[36] M. S. Sodha, R. S. Mittal, S. Kumar, V. K. Tripathi, *Optoelectronics (London)* **1974**, *6*, 167.
[37] R. Fowler, Statistical Mechanics, Cambridge University Press, Cambridge **1936**.
[38] M. S. Sodha, A. Dixit, *Appl. Phys. Lett.* **2009**, *95*, 101502.
[39] J. D. Huba, NRL Physics Formularly, Naval Research Laboratory, Washington, D.C. **1998**.
[40] M. S. Sodha, A. K. Ghatak, V. K. Tripathi, Self-filamentation of Laser Beams in Dielectrics, Semiconductors and Plasmas, Tata-McGrawHills, Delhi **1974**.
[41] M. S. Sodha, A. K. Ghatak, V. K. Tripathi, *Prog.Opt.* **1976**, *13*, 169.

How to cite this article: Sharma R, Sharma SC. Theoretical analysis for transmission of Gaussian and sine time irradiance of electromagnetic beam in collisional dusty plasmas. *Contributions to Plasma Physics*. 2020;e201900175. https://doi.org/10.1002/ctpp.201900175

Rs. 30.00
ISSN- 0566-2257

UNIVERSITY NEWS

A Weekly Journal of Higher Education

Association of Indian Universities

Vol. 58 • No. 28 • July 13-19, 2020

Seema Singh
Theoretical Framework of Collaboration between MSME and Private Engineering Colleges to Co-exist during COVID-19

Preeti Dharmik, Rekha Sharma and C G Dethe
Impact of Orientation Programmes on Knowledge, Attitude and Skills of Senior University and College Teachers

Pranita Gopal
Good Practices in Virtual Presentations for Higher Education

M Venkaiah Naidu
Transformations through Innovative Technological Interventions

Convocation Address

Sajal Dasgupta
Innovative Initiatives Taken for Reforming Conduct of Education During COVID-19 Pandemic Lockdown Period

Communication

Celebrating 90 Years of University News

#Let'sBeatCoronaTogether

UNIVERSITY NEWS
Vol. 59 July 13-19
No. 28 2020
Price Rs. 30.00

A Weekly Journal of Higher Education Published by the Association of Indian Universities

In This Issue

ITEMS PAGE

Articles

Theoretical Framework of Collaboration between MSME and Private Engineering Colleges to Co-exist during COVID-19 3

Impact of Orientation Programmes on Knowledge, Attitude and Skills of Senior University and College Teachers 12

Good Practices in Virtual Presentations for Higher Education 20

Convocation

National Institute of Technology, Tadepalligudem, Andhra Pradesh 25

Communication

Innovative Initiatives Taken for Reforming Conduct of Education During COVID-19 Pandemic Lockdown Period 28

CAMPUS NEWS 30

AIU NEWS 32

Theses of the Month 29

(**Science & Technology**) 38

Advertisements 40

New Subscription Tariff

(Effective April 01, 2020)

	Inland		**Foreign**	
Institutions	Academics/ Students *(at residential address only)*		Airmail	Surface Mail
	Rs.	Rs.	US$	US$
1 year	1250.00	500.00	210.00	170.00
2 years	2200.00	900.00	400.00	300.00

Subscription is payable in advance by Bank Draft/MO only in favour of Association of Indian Universities, New Delhi.

Opinions expressed in the articles are those of the contributors and do not necessarily reflect the views and policies of the Association.

Editorial Committee Chairperson:
Dr (Ms) Pankaj Mittal

Editorial Committee:
Dr Baljit Singh Sekhon
Dr Amarendra Pani
Dr Youd Vir Singh

Editor:
Dr Sistla Rama Devi Pani

#Let'sBeatCoronaTogether

Theoretical Framework of Collaboration between MSME and Private Engineering Colleges to Co-exist during COVID-19

Seema Singh*

In December 2019, a Novel Coronavirus (SARS-CoV-2) emerged in Wuhan, China which became a global pandemic within months (Bouey, 2020). The virus has spread to more than 215 countries with 7.8 million infected and has caused 429,782 deaths. As a preventive measure, Indian Government has imposed a complete lockdown, barring emergency services. This has played havoc on the life and livelihood of a large number of people, more on those who have little reserves to spend. The medical crisis has profound economic and social effects. The Micro, Small and Medium enterprises (MSMEs) are a significant part of India's economic landscape as in many other countries of the globe. The MSMEs generate revenue, enhance employment and provide access to commodities for remote communities (World Bank, 2004; Wiklund, 2005; Satpathy and Rani, 2017). As per the Report of the Working Group on Micro, Small and Medium Enterprises (MSMEs) growth for 12th Five Year Plan (2012–17), the sector accounts for 45 per cent of the manufacturing output and 40 per cent of the total exports of India. Further, it accounted for 30.74 per cent of GDP in the year 2014–15. As per the registration between October 2015 to December' 2019, the micro sector (88.6 per cent) is the largest component of the MSME than the other two, small (10.9 per cent) and medium (0.42 per cent) sectors and provides employment to 61.5 per cent of the workforce against 32.2 per cent and 6.3 per cent in case of small and medium enterprises respectively (GoI, 2020). However, the micro sector has remained comparatively under-researched (Cristian, 2016). Add to this, Indian engineering education system has registered an exponential growth during the last twenty-five years due to enhanced global demand for English-language speaking engineers, who are equally competent, but at lower wages. Moreover, the un-aided private sector is a significant component of the whole spectrum of engineering education systems in India. What is the effect of lockdown on the micro sector as per the anecdotal evidence available? The hundreds of thousands of stranded migrant workers across the country, suggests that MSMEs have been the worst casualty of Covid-19 induced lockdown (Mishra, 2020). Higher education in general and engineering education in particular has also been adversely affected by the pandemic and they are still in search of an alternative sustainable model. In this background, the study tries to investigate whether a collaboration between these two sectors will be a win-win situation for both. The second section of the paper provides relevant literature review. Objective and methodology have been discussed in

*Professor, Department of Economics, Delhi Technological University, Rohini, Delhi- 110 042. E-mail: seemasinghdtu@gmail.com

the third section of the paper. Fourth section provides an analysis. A model has also been developed for cooperation between these two sectors Last section concludes the discussion and provides suggestions.

Literature Review

The literature review has been done in the four groups as i. The Micro Sector and Impact of COVID-19 on it; ii. Private Engineering Education in India and impact of COVID-19 on it; iii. University and Firms collaboration:

The Micro Sector and Impact of COVID-19

MSMEs are perceived as the backbone of the economy and 'an engine of growth' not only in India (Satpathy and Rani, 2017) but in other parts of globe (European Commission, 2016; Audretsch, 2009; Tesfayohannes, 1998). They also have a high export propensity (Claudiu Cicea, Ion Popa, Corina Marinescu & Simona Cătălina Ștefan, 2019), Huong T.X. Nguyen & Viet Le (2019)). However, it needs to be kept in mind that the small firms are not homogenous. They all are unique in nature and not like miniature corporate firms. So, any theory developed for corporate firms cannot be replicated to them (Beaver and Prince, 2004). In spite of being so important a segment of the economy, they have very poor technological endowment and their up-gradation is essential for the economic wellbeing of the nation (Venkatesh and Muthiah, 2012). Their output may be high in quantitative terms, but is very low in quality which can be up-graded by technological up-gradation of the sector. The author further view that advanced technology enables the small firms to create quality employment improving remuneration, duration, and skills (Bhavani T.A.2011; Singh et al. 2012). However, the study conducted by Bargal et.al. regarding performance of small scale industries (SSIs) during the period of nineties vis-à-vis the pre-reform years the annual average growth rate of different parameters have concluded that they have declined during post-liberalisation period (Bargal et al., 2009). According to Lawson and Samson (2001), a firm which is able to link the technological strategy with its business and innovation strategy, are capable of bringing innovation in the firms. Experts as Yang (2012), Martín-de Castro et al. (2013) and Gupta and Nanda (2015) have also highlighted the importance of technological adaptation for enhancing competitiveness. There is consensus among scholars that small firms should concentrate on developing market reach (Wheelen & Hunger, 1999). However, they are unable to raise finances to invest in cost-reducing technology for the firm (Beaver and Prince, 2004).

During COVID-19, countries have taken unprecedented and drastic measures to contain the spread of the virus by closing towns, cities, and national borders, disrupting not only the big corporations but also millions of SMEs around the world (SMEFF, 2020). Though it is too early to assess its full impact on MSME, but given that the COVID-19 is a pandemic has affected humanity, MSMEs would not be immune from it. The world economy is expected to face a very serious crisis besides on the social and political sphere. Over a period of fewer than two weeks, since it was officially declared a pandemic by WHO, COVID-19 has disrupted lives and economies around the globe. According to a briefing issued by the United Nations Department of Economic and Social Affairs, the COVID-19 pandemic will likely shrink the global GDP by almost one per cent in 2020. In such a brief period, many international organizations have analysed, assessed, the impact of the COVID-19 pandemic, and formulated scenarios for world economic growth in 2020. According to Citi Research (published April 7, 2020), the world economy will decline by -2.3 per cent in 2020 (compared to 2.6 per cent in 2019, more than -1.7 per cent in 2009). In particular, China's GDP growth was only about 2.4 per cent (compared to 6.1 per cent in 2019), United States -2.6 per cent(compared to 2.3 per cent in 2019), Japan -1.9 per cent (compared with an increase of 0.7 per cent in 2019), while the Eurozone has regressed -8.4 per cent (compared to 1.2 per cent in 2019) (Dung, Thi Bich and Tho Trần Bá, 2020). About 9-18 million formal jobs may be lost and 100 mn informal jobs vulnerable to income loss in Sub Saharan Africa (McKinsey, 2020). Even studies have been done on the impact in India show that about 30 to 40 percent of the micro sector firms will not be able to survive. However, the impact will be differential depending on the activity of the firm. (ISLE-ILO, 2020).

The pandemic cannot be compared with any other natural disaster due to its magnitude. But, in absence of any other event similar to COVID-19, a comparison to flood can be made. Aruzzir (2018) in her study in Malaysia for another natural calamity, flood, has found that the effect of any natural disaster on SMEs is very severe (Auzzir, 2018). Bouey (2020) has done the immediate impact analysis on the Chinese SME sector. China is the second-largest

economy with international trade with almost all countries of the world which ranges from mineral and food products to highly sophisticated electronic gadgets and automobile parts. So, certainly, it will affect the world's economy (Bouey Jennifer' 2020). With fears of a new recession and financial collapse, Nicola et. al. (2020) have suggested that these times call for resilient and strong leadership in business and government along with wider society. Immediate relief measures need to be implemented and adjusted for those that may fall through the cracks. A broad socio-economic development plan including sector-by-sector plans need to be created, and along with it an ecosystem that encourages entrepreneurship so that those with robust and sustainable business models can be allowed to flourish. COVID-19 will put many small businesses on life support (Nicola et. al. 2020). Karen G. Mills has suggested policymakers on aid options, offers guidance to owners on the brink of ruin (Kost, 2020), and organizations need to develop a trust-based culture with their employees (Beer, 2020). Samuel in a study of 300 randomly sampled SME-Owners in Ghana has discussed the importance of financial literacy for the SME growth (Samuel, 2018).

According to the extant literature, MSMEs are more vulnerable to natural hazards than large firms due to the fact that MSMEs: tend to operate in sub-optimal locations; are smaller and financially weaker; have a more limited, usually local market; tend to implement less DRR measures and be more excluded from recovery programs (Zhang et. al. 2004; Battisti, Martina, and David Deakins, 2012). According to the US Institute for Business and Home Safety, many businesses do not reach the post-disaster stage (Institute of Business and Home safety).

Private Engineering Colleges in India and Impact of COVID-19

History of Engineering Education in India can be traced back to 1794, when the School of Survey was established at Guindy, Tamil Nadu on the recommendation of Michael Topping to train Indians on construction of roads, bridges, docks, etc. With the involvement of British in the Second World War and the first Indian war of independence, mechanical engineering was added with civil engineering to train the lower grade technicians to manage equipment used for the army, navy and other technical establishments for maintenance of British Empire. (Shaha S., 2011). Private higher education institutions are established by philanthropic, private business establishments, religious and for-profit organizations. In 1970, India had a total of 139 engineering institutions, and only four of them were private. (Singh and Singh, 2018).

After Globalisation and use of Information and Communication Technology in the production process, producers have the opportunity to shift a part of the production process to far off places, where it can be done with required quality at low cost. Many large and multinational corporations came to India to use its English-speaking engineers besides many other resources (Singh, 2005). With the country counting on its demographic dividend, the increase in the number of students seeking engineering degrees put up pressure on the higher education sector to expand. Limitation of the government funding in engineering education which was not sufficient to meet the growing demand, led to the historic decision of privatisation. The Private Universities (Establishment and Regulation) Bill was introduced in 1995 in the Parliament. Passing the bill gave a boost to the opening of many private engineering colleges. By the end of 2000, number of engineering institutions rose to nearly 1,400 out of which about 200 only belonged to the government (Singh and Singh, 2014).

As per the Table-1, private engineering colleges were established across India but number was quite impressive in southern states. This is necessary for higher education institutions (especially private engineering institutions) to collaborate or develop partnership with any reputed organization or elite institutions both nationally and internationally rather than depending on students pass out rate and placement only to attract students amidst the fierce competition (Singh & Singh 2018).

University and Institute Collaboration

The Higher Educational Institutions are indulging in market-like behaviour for revenue generation. They are collaborating with business firms for technology transfer all over the world for raising funds. On the basis of the 2008-10 data from National Research Foundation of Korea, Ministry of Education and Technology and Korea Foundation for Promotion of Private Schools, Han and Heshmati (2013) have analysed the financial rewards from university-industry research cooperation. As per the findings, participation of engineering faculty, patent approvals, the volume of research funds, and firms in

Table 1. Regional Presence of Engineering Institutions

Region	State/ Union Territory	Government Aided	Private Unaided	University Managed
North	Chandigarh	6	1	2
	Delhi	18	18	1
	Haryana	46	288	5
	Himachal Pradesh	17	36	1
	Nammu & Kashmir	9	18	1
	Punjab	34	186	6
East	Arunachal Pradesh	2	1	NIL
	Assam	15	5	2
	Jharkhand	16	16	1
	Maniour	2	NIL	NIL
	Meghalaya	3	1	NIL
	Orissa	23	172	2
	Sikkim	2	NIL	1
	Tripura	2	NIL	NIL
	West Bengal	59	93	3
West	Goa	7	3	NIL
	Maharshtra	71	660	8
South	Andaman and Nicobar	1	NIL	NIL
	Karnataka	158	239	3
	Kerala	90	95	13
	Andhra Pradesh	132	763	5
	Pondicherry	8	13	NIL
	Tamil Nadu	78	852	4
	Daman & Diu	1	NIL	NIL
	Goa	7	3	NIL
Central	Chattisgarh	25	47	3
	Gujarat	49	128	3
	Madhya Pradesh	66	213	6

Source: AICTE website, Singh & Singh 2018.

incubators within universities turn out to be significant contributors to externally sourced university revenues. They have given an exhaustive list of study conducted on the same issue (Table-2) which shows positive result out of their collaboration.

It appears that industry-university cooperation foundation programs are likely to play a significant role in private university finances.

In this background, the paper discusses the possibility of collaboration of private engineering colleges and micro sectorto overcome the setback from the COVID-19 pandemic. Whether it will be a win-win situation. The discussion has been made on the basis of secondary sources and small studies done by individual researchers and organisations.

Analysis

Recently, the government of India has tried to remove the ambiguity and clearly define these sectors as discussed in the Table-3. Along with investment, turnover criteria has also been added. The rand has also been increased. However, keeping micro and medium enterprise in one group is akin to comparing cheese and chalk. While the medium sector is qualified and have resources to hire experts and technology to improve their productivity and efficiency, take lone from the bank, the micro sector is not the perspective or if they have, they are not able to develop sufficient resources to accomplish the task.

Coronavirus and following nationwide lockdowns have put the economic future in jeopardy. However, it is a question of survival for micro sector As per various studies about 30 to 40 percent of the micro sector firms will not be able to revive again (ISLE-ILO, 2020). At the same time, many of them are working in such areas that if they become part of a network or ammelgoration, their return on investment will increase which needs strategic planning and technological updardation along with finances. So, in the light of government financial package for revival and mudra loan, micro firms need an institutional support for assist in strategic planning and technological upgradation.

As per AICTE dashboard, in 2018-19, there were 2,678 private unaided engineering institutions in India with 6,09,090 enrolments and 3,14,840 faculty. The intake capacity was 13,20,573. Its means more than fifty percent seats remained vacant during 2018-19. Now, in the post COVID-19 academic session 2020-21, loss of parent's job or cut in salary or delayed payment of salarymay have cascading effect on the demand for seats in such engineering colleges. Universities across the world are expecting shrink in sources of income and as an alternative, many world class universities have started offering on-line

Table 2: Study on University - Firm Collaboration

Researcher and Sample	Findings
Joshua and Patricia (2003) on the basis of study of 108 universities in the USA	Age of Technology Transfer Office, the number of outstanding engineering faculty members and research fund received by external organisations have a positive relation with performance respectively.
Thursby and Thursby (2002) on the basis of 64 universities in USA	Performance is related to the number of experts, faculty members, andpersons working at the Technology Transfer office and their relationship with external firms
Jenson and Thursby (2003) Interview and Regression (2003	Reward system through technology transfer is directly related with performance
Sigel et al. (2003a) Interview and Regression for 98 people who are directly related to TTO in US	Reward systems for interested persons have effects on performance
Power (2003) Interview and Regression for 66 enterprises and 312 university researchers.	Research funds including both public and private have a positive correlation with patent products. But this is not related to revenue through the licensing. ·The more distinguished faculty member a university has, the more patents and licensing a university has.
Kim and Lee (2007) Regression 61 universities in Korea	Research competency such as the number of SCIE papers and the number of patent registrations were significant, but managerial competency such as the scale of technology transfer organizations and the number of specialists was not statistically significant
Kwon and Han (2009) Regression for 169 universities in Korea, Explorative	• Public universities have higher performances than those of private universities in terms of the number of technology transfers and the amount of technology licensing fees. • The characteristics of university such as age, size, number of departments, faculty members, students, and experts, the number of SCIE papers are not statistically related to performances. • Regional universities have higher performance than universities in urban areas in terms of technology transfer.
Seo et al. (2005) descriptive	University's own firms based on university's own technology and holding companies are related with revenue through industry-university collaboration.
Sapsalis et al. (2006) Regression for 89 universities in United States. .	Scientific competency, the number of papers, and patents have a positive effect on revenue through industry-university collaboration.
Kim (2005) Survey for 54 universities and 79 firms in Korea	• Level of education of experts working at technology transfer office has a positive effect on revenue through industry-university collaboration • The number of patents has an effect on revenue through industry-university collaboration
Friedman and Silberman (2003), Jaffe et al. (1993) regression with 12 firms	• The degree of proximity of high-tech firms near school has positive effects on technology transfer. • Geographical location of universities has positive relation with knowledge spill over.
Foltz et al. (2000) Survey for bio-agricultural firms in United States	Research funds received from federal government and university have positive effects on revenue, while research funds received from industries have no relation with performance.
O'Shea et al. (2005) Byun (2004) survey 107 universities in Korea	It shows that university scale, age of professional institutions, and incentive systems for technology developers have positive effects on revenue through industry-university collaboration.

Source: Han Junghee and Heshmati Almas (2013), p.19-20

degrees. A representative list has been discussed in the Table 4.

These efforts may further bleak the future for unaided private engineering colleges in India. However, they have human capital, in the form of qualified faculty, laboratory assistants and young energetic students. Not to forget, the role of higher education institutions (HEIs) towards the development of their surrounding regions is not a new phenomenon. This has been illustrated by the establishment (in the late 1800s) of land-grant colleges in North America and technical and civic universities in Germany and England, respectively (Editorial, 2017).As Borrel-Domial et.al. (2010) have discussed that generally Ph.D. is pursued for following teaching profession but now, about 50 per cent are pursuing for a wide range of careers. On the basis of the experiences reported by stakeholders, the paper suggests working practices and supporting policy measures that can address the (potentially conflicting) requirements for a high quality of education, sound research and adequate preparation for diverse career with participation of private sector firms (Borrel-Domial et.al., 2010). Maya Ziswiler of the UBS Optimus Foundation (UBSOF) has discussed innovative ways the education sector is using both blended finance and outcome-based financing. For example, UBSOF's investment in education social enterprises with Acumen in East Africa provides loans and equity investments coupled with a grant for pre- and post-investment support and an outcome-based incentive for both the investment manager and the investees if outcome targets are reached. Another example mentioned for supporting education social enterprises was impact loans, where the interest rate fluctuates based on the outcomes achieved. social and development impact of bonds (SIBs and DIBs), a form of outcome-based financing where investors provide up-front capital to service providers with the opportunity for a return if agreed-upon outcomes are achieved (Gustafsson- Wright Emily, 2020). In India, firms are collaborating with engineering institutions but they are generally large firms or government organisations. It is not common for

Table 3: MSMEs Classification in India

Existing Criteria	**Investment in Plant & Machinery or Equipment**		
Characteristic	Micro	Small	Medium
Manufacturing Entreprises	Investment< 25 Lac	Investment< 5 Crore	Investment.< 10 crore
Service Entreprises	Investment.<10 Lac	Investment.< 2 Crore	Investment< 5 crore
Revised Criteria	Composite Criteria - Investment and Annual Turnover		
Manufacturing & Services	Investment< 1 crore and Turnover< 5 crore	Investment.< 10 crore and Turnover< 50 crore	Investment< 20 crore and Turnover< 250 crore

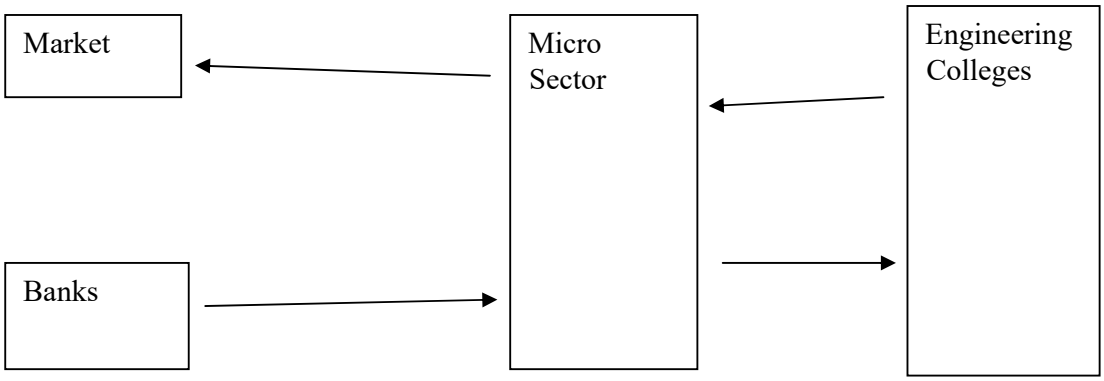
Source: Rao Harshitha, 2020

Table 4: List of On-line Courses

SN	**Course**	**Offered by**
1	MS in Computer Science	Arizona State University
2	Data Science for Engineers	IIT, Madras, July 20-Sept. 11, 2020
3.	Industry Ready Master Classes for Civil Engineering Students	contact@protrainy.com +91 7978817062, +91 8617833413 programme starting from June 7' 2020
4.	Basic Electrical Circuits	IIT Madras [July 20 to Oct 9' 2020
5.	M.Sc in Computer science	Liverpool John Moores University
6.	Master of Computer and Information Technology	University of Pennsylvania [16-40 Months, Fully Online]: Apply by Aug 1
7.	From 0 to N: Architecting on Alibaba Cloud	Alibaba Cloud Academy [Online, 5 Months] Gustafsson-Wright Emily (292

Source: Compiled by author

Fig 1: Model of Engineering Institutions and Micro Sector.



the medium sector, and almost absent in case of micro-sector. However, looking at their capacity to contribute towards the GDP and create employment and for enhanced use of ICT after COVID-19 in the production process, they need use of technology which can be provided by the Private engineering colleges and along with helping them in strategic planning. At present, there is no such platform. It is up to colleges, Industry bodies as CII or FICCI etc. which may work in that area. At Delhi Technological University, B.Tech. students visit slum to assess how knowledge of their branch of engineering can be used to improve life at slum since the academic year 2011-12. (Singh, 2013). The process has been integrated with the examination system as marks is given out of internal marks (Singh, 2016). Similar procedure can be adopted for associating with micro sector. The Unnat Bharat Abhiyan[2]of the Ministry of HRD is similar programme for rural development since 2014.Funds for Awareness, survey, visits etc @₹ 10000/- per village • Technical Intervention ₹ 1.0 lakh per technology for implementation in the village • Customization of Technology ₹ 50000/- per solution for villages.In the Fig-1, a Model for collaboration between Micro sector and Private engineering colleges (unaided) has been represented to achieve a win-win situation for both.

From the financial help under COVID-19 revival package of the government or Mudra loan, the Micro sector can approach engineering institutions for technological help or assistance in strategic planning. The enabled Micro Sector with the help from engineering colleges may go to market alone or may be as part of Cartel. They can become member of a production network for value addition also. This is the time to float loan to Micro Sector and knowledge input from engineering colleges to see the practical implication of the model. Professor ERichard Thaler and legal scholar Cass Sunstein have discussed about NUDGES in their New York Times best seller book entitled Nudge: Improving Decision about Health, Wealth and Happiness. It is crucial to get the necessary response from the beneficiaries, Micro sector firms and Private engineering colleges (unaided) to determine the success or failure of the model. NUDGES stand for incentives, Understand mappings, Defaults, Give feedback, Expect error, Structure complex choices. Through NUDGES, the model will improve further and become sustainable over a period of time.

Conclusion

The lockdown as a preventive measure during COVID-19 has disastrous impact on the micro sector which has employment potential and which contributes significantly to the GDP and exports. Though the government has provided a revival package, but *sans* competitive technological assistance or sustainable business model, survival will be difficult post-lockdown. On the other hand, Private engineering colleges will also have pressure of generating finances to survive. The paper concludes that more empirical research is needed to develop a more unified approach to MSMEs success factors, which informs thinking on this critical matter, to both policy makers and business owners (Lampadarios Evripidis; Kyriakidou Niki; Smith Gordon (2017) and engineering institutions. The paper has real implication for today's world situation which has never been experienced earlier. The model may be used as a win-win situation for both. A secondary source of financing for private engineering colleges and a hand-holder for micro sector.

Notes:

On 13th May 2020, Finance Minister Nirmala Sitharaman added the additional principle of turnover along with the investment. The Union Cabinet had approved the amendment to change the criteria to classify MSMEs from "investment in plant and machinery" to "annual turnover."

Unnat Bharat Abhiyan 2.0, an extension of the programme with improved feature has been stated in 2018.

References

1. Attisti, Martina, and David Deakins (2012). Perspectives from New Zealand Small Firms: Crisis Management and the Impact of the Canter bury Earthquakes, New Zealand Centre for Small & Medium Enterprise Research (Ed.), Business SM Easure. Palmerston North: Massey University.
2. Audretsch, D B (2009). The Entrepreneurial Society. *The Journal of Technology Transfer*, *3*, 245–254.
3. Ballesteros, Marife M, and Sonny N Domingo, (2015). Building Philippine SMEs Resilience to Natural Disasters. No. 2015-20. PIDS Discussion Paper Series, 2015-20, Econstor; https://www.econstor.eu/bitstream/10419/127030/1/pidsdps1520_rev.pdf dated 10.6.2020.
4. Bargal, H, Dashmishra, M., and Sharma, A. (2009), „Performance Analysis of Small Scale Industries - A Study of Pre-liberalization and Post-liberalization period", International Journal of Business and Management, Vol 1, No 2. [3].
5. Bavel Jay J Van, Baicker Katherine et.al.(2020), Using social and behavioural science to support COVID-19 Pandemic Response, Perspective, Nature Human Behaviour, https://doi.org/10.1038/s41562-020-0884-z, pp.
6. Beaver and Prince (2004). Management Strategy and Policy in the UK small business sector: a critical review, *Journal of Small Business and Enterprises Development*, 11(01).
7. Beer Michael (2020). Organizations will Develop Trust-based Cultures with Employees in Gerdman Dina, How The Corona Virus is already Rewriting the future of Business, Working Knowledge, Harvard Business School, March 16 https://hbswk.hbs.edu/item/how-the-coronavirus-is-already-rewriting-the-future-of-business, 14.5. 2020.
8. Bhavani, T.A. (2011). Dynamic Business Environments: What These Mean for Indian Small Enterprises in Micro and Small Enterprises in India: Era of Reforms: Keshab Das (Ed) pp. 27-45.
9. Borrell-Damian, L, Brown, T, Dearing, A. *et al. (2010)* Collaborative Doctoral Education: University-Industry Partnerships for Enhancing Knowledge Exchange. *High Educ Policy* 23 (2), pp**.** 493–514.
10. Bouey, Jennifer (2020). Assessment of COVID-19's Impact on Small and Medium-Sized Enterprises: Implications from China. *Santa Monica*, CA: RAND Corporation,. https://www.rand.org/pubs/testimonies/CT524.html.
11. Claudiu Cicea, *et. at.* (2019). Determinants of SMEs' performance: evidence from European countries, Economic Research-Ekonomska Istraživanja, 32:1, 1602-1620, DOI: 10.1080/1331677X.2019.1636699.
12. European Commission. (2016). Annual Report on European SMEs 2015/2016. Retrieved from http://ec.europa.eu/growth/smes/business-friendly-environment/performance-review-2016_ro
13. Gherhes Cristian, *et. al* (2016). Distinguishing micro-businesses from SMEs: A systematic review of growth constraints *Journal of Small Business and Enterprise Development,* 23(4), November.
14. Gupta, H and Nanda, T, (2015). A Quantitative Analysis of the Relationship Between Drivers of Innovativeness and Performance of MSMEs', *International Journal of Technology, Policy and Management*, Vol. 15, No. 2, pp.128–157.
15. Gustafsson- Wright Emily. (2020). Education In the time of COVID- 19: Reflection from the Charcha 2020 Forum Brookings, https://www.brookings.edu/blog/education-plus-development/2020/05/26/education-in-the-time-of-covid-19-reflections-from-the-charcha-2020-forum/?utm_campaign=Center%20for%20Universal%20Education&utm_medium=email&utm_content=88647927&utm_source=hs_email on 10.6.2020
16. Han Junghee and Heshmati Almas , (2013). Determinants of Financial Rewards from Industry-University Collaboration in South Korea, IZA Discussion Paper No. 7695 October, Bonn, Germany pp. 1-21.
17. Huong T.X. Nguyen & Viet Le (2019). Network Ties and Export Propensity of Vietnamese Small and Medium Enterprises, *Asia Pacific Business Review*, 25:1, 100-122.
18. Jensen. R. and Thursby, M (2003). Disclosure and Licensing of University Inventions: The Best We Can Do, *International Journal of Industrial Organization* 21, 1271- 1300.
19. Kohoutek, J, *et al.* (2017). The Role of Higher Education in the Socio-Economic Development of Peripheral Regions. *High Educ Policy,* 30 (4), 401–403 (2017).
20. Kost Danielle (2020). How Small Business can survive the Coronavirus Outbreak, Working Knowledge, Harvard Business School, March 20https://hbswk.hbs.edu/item/how-small-businesses-can-survive-the-coronavirus-outbreak
21. Lampadarios Evripidis, Kyriakidou Niki, and Smith Gordon (2017). Towards a new framework for SMEs success: a literature review, *International Journal of Business and Globalisation,* 2017 Vol.18 No.2, pp.194 – 232.

22. Lawson, B and Samson, D (2001). Developing Innovation Capability in Organisations: a Dynamic Capabilities Approach, *International Journal of Innovation Management*, Vol. 5, No. 3, pp.377–400.

23. Martín-de Castro, *et.al,* (2013). Linking Human, Technological, and Relational Assets to Technological Innovation: Exploring a New Approach', Knowledge Management Research & Practice, Vol. 11, No. 2, pp.123–132.

24. Metcalf, G., Jenkinson, K. and Johnstone, K A, (2010). Changing Climate for Business. 3rd ed. Oxford: UK Climate Impacts Programme, https://ukcip.ouce.ox.ac.uk/wp-content/PDFs/UKCIP_Business.pdf 12.6.2020

25. Mishre Udit (2020). Explained: Why are Medium, Small, Micro Enterprises worst hit by Covid-19 lockdown?, *The Indian Express*, May 7, https://indianexpress.com/article/explained/coronavirus-india-lockdown-msme-sector-crisis-government-relief-package-6395731/ on 14.5.2020

26. Nicola, M., *et.al*., (2020). The Socio-Economic Implications of the Coronavirus and COVID-19 Pandemic: A Review. *International Journal of Surgery (London, England)*, S1743-9191(20)30316-2. Advance online publication. https://doi.org/10.1016/j.ijsu.2020.04.018 on 11.6.2020

27. Rampal Jeevant, (2020). Corona Virus Crisis: A Behavioural Economist's guide to shaping India's Covid:19 Policy, https://scroll.in/article/958269/a-behavioural-economists-guide-to-shaping-indias-covid-19-policy on 26.5.2020.

28. Rao Harshitha (2020). Turnover to Define MSMEs: The New MSME Definition 2020, https://razorpay.com/learn/new-msme-definition-turnover-2020/ on 27.5.2020.

29. Saha, S.K. and Ghosh, S (2011). Engineering Education in India: Past, Present and Future. *Propagation*, 2(2), pp.114.

30. Samuel Kwaku Agyei, Christian Nsiah (Reviewing Editor) (2018) Culture, financial literacy, and SME performance in Ghana, *Cogent Economics & Finance*, 6:1.

31. Satpathy Saritaand and Rani Sailja (2017). A Study of Micro, Small and Medium Entrepreses: the Back Bone for Economic Development of the Indian Economy, International *Journal of Research and Scientific Innovation,* 04(06) June, pp 75-82.

32. Singh Antra and Singh Seema, (2018). Expansion of Private Engineering Institutions: Challenges Ahead, *Higher Education for the Future*, 5(1) pp. 20–39.

33. Singh Antra and Singh Seema, (2014). Private Engineering Education in India: Past, Present and Future, *International Journal of Management and Social Sciences Research, ISSN: 2319-442; 03(11),* November; pp. 39-46.

34. Singh, R, Verma, O P, and Anjum, B (2012). Small Scale Industry: An Engine of Growth, *Zenith International Journal of Business Economics & Management Research*, Vol.2 Issue 5.

35. Singh Seema, (2016). Integrating Social Responsibility of University and Corporate Sector for Inclusive Growth in India, *Higher Education for Future* (Sage Publication), 03(02) July 2016, pp. 183-196.

36. ----------, (2013). Social Responsibility of Engineering Institutions for Inclusive Growth, *Universities News, 2013,* 51(03), pp. 01-04-12.

37. ----------, (2005). MNCs Effect on Employment of Engineers, *The Indian Journal of Technical Education,* 28(2), April- June 2005, pp. 38-44.

38. SME Finance Forum (2020). Five Lessons From COVID-19 Crisis and SME, https://www.smefinanceforum.org/post/5-lessons-we-learned-about-covid-19-crisis-and-smes on 14.5. 2020.

39. Tesfayohannes, Mengsteab, (1998). The Promotion of Small and Medium-Scale Enterprises Financing as a Contribution to Sustainable Industrial Development- The Case of Eritrea. Aachen: Shaker Verlag, Germany.

40. Venkatesh, S and Muthiah, K, (2012). SMEs in India: Importance and Contribution, *Asian Journal of Management Research*, Vol. 2, No. 2.

41. Yang, J (2012), Innovation Capability and Corporate Growth: An Empirical Investigation in China', *Journal of Engineering and Technology Management*, Vol. 29, No. 1, pp.34–46.

42. Zairol Auzzira*, Richard Haighb , and Dilanthi Amaratungab, (2018). Impacts of Disaster to SMEs in Malaysia, *International Conference on Building Resilience using scientific Knowledge to informpolicy and practice in disaster risk reduction*, ICBR 2917 27-29 November; 2017.

43. Zhang, Yang, (2004). Modeling and Managing the Vulnerability of Community Businesses to Environmental Disasters, *45th Annual Conference of the Association of Collegiate Schools of Planning*, Portland, OR, October. Vol. 23. ❑

Impact of Orientation Programmes on Knowledge, Attitude and Skills of Senior University and College Teachers

Preeti Dharmik*, Rekha Sharma and C G Dethe*****

India's higher education system is the world's third largest in terms of students, next to China and the United States. In future, India will be one of the largest education hubs. India's Higher Education sector has witnessed a tremendous increase in the number of Universities/University level Institutions & Colleges since independence (Sheikh Younis Ahmad, 2017). Higher education is becoming a major driver of economic competitiveness in an increasingly knowledge-driven global economy. The imperative for countries to improve employment skills calls for quality teaching within educational institutions. In order to identify and promote good teaching practices, institutions may implement evaluation mechanisms. The environment of higher education institutions can enhance the quality of teaching through various means, for example, by issuing a national policy or recommendations by quality assurance agencies (Hénard, 2010).

The National Policy on Education (NPE) 1986 recognized the urgent need for creating effective systems to provide opportunities for professional and career development of teachers, necessary for quality education and research so that they are well equipped and motivated to accept new challenges emerging from growth of new knowledge, international competitiveness and changing requirements of learners, especially in the institutions of higher education. Therefore, it was proposed to enhance their motivation skills and knowledge through systematic orientation in specific subjects, techniques and methodologies, and thereby inculcate in them the right kind of values that would in turn encourage them to take initiatives for innovative and creative work (Guidelines for Regional Centre for Capacity Building and Human Resource Development Centre, 2015).

** Assistant Professor, UGC- Human Resource Development Centre, Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur- 440033. E-mail: preeti.dharmik5@gmail.com*
***Associate Professor, UGC- Human Resource Development Centre, Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur- 440033. E-mail: drrekha_sharma2000@yahoo.com*
****Director, UGC- Human Resource Development Centre, Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur-440033.E-mail: cgdethe@gmail.com*

In order to achieve the above objectives, the Academic Staff orientation scheme, later known as Academic Staff College (ASC) scheme was initiated by the University Grant Commission in 1987 based on the NPE recommendations. There are 66 ASCs in India spread over 26 states of our country. In 2015, these Academic Staff Colleges were renamed as Human Resource Development Centres. The Guidelines for UGC—Academic Staff College from 2002-07 and 2007-12 and UGC- Human Resource Development Centre, 2019 mentioned the four components to be included in orientation courses viz.,

Component A: Awareness of linkages between society, environment, development and education

Component B: Philosophy of education, Indian education system and pedagogy

Component C: Resource awareness and knowledge generation and

Component D: Management and personality development

After renaming of UGC-Academic Staff College to UGC-Human Resource Development Centres in 2015, the topics to be included in Orientation Programmes were redefined as follows:

- 10 per cent weight for topics in higher education such as issues of ethics, gender, marginalized communities, plagiarism etc.
- 10 per cent weight for issues related to environment.
- 10 per cent weight for issues concerning service matters of teachers.
- 20 per cent for broad cross discipline topics to motivate the trainees for development of interdisciplinary understanding and interest including basic legal awareness.
- 10 per cent weight for Research Methodology.
- 15 per cent weight for Communication Skills and Information Technology.
- 10 per cent weight for Microteaching

- The remaining 15 per cent of content of the First Level Programme will be focused on broad understanding of various subjects with recent global trends and developments.

Though the weightage for each domain were redefined, but the main thrust areas were similar to the UGCs Guidelines issued during 2002-07, 2007-12 and 2019. UGC-Human Resource Development Centre of Rashtrasant Tukadoji Maharaj Nagpur University was established in 1998 and till date it has organised 105 orientation programmes and catered to about 4363 teachers from university and colleges. Several researches have been carried out to study the impact of Orientation Programme on the professional development of teachers (Dhawan (2000), Rastogi Savita, (2001), Joshi Suresh and Pareek Sushila (2006), Pawar I. Anand and Mouli. S. Chandra, (2008), Survey and Bagul (2012), Satsangi *et al.* (2012), Sonawane and Survey (2015) and Khadilkar (2016)); however, there is a paucity of researches in terms of impact of different domains of the Orientation Programmes viz., teaching, ICT, research and soft skills and economic, social and environmental knowledge on teachers' knowledge, attitude and skills. Hence, the present study was carried out to assess the impact of Orientation Programmes on knowledge, attitude and skills of senior college and university teachers. The objectives of the present study were to assess the improvement in teachers in following domains after attending the Orientation Programme viz.,

1. Knowledge and practice of teaching skills and methodology;
2. Attitude of teachers towards students;
3. Research aptitude and skills;
4. ICT skills;
5. Life skills;
6. Knowledge of social, economical and environmental issues.

Materials and Methods

About 113 university and college teachers who attended Orientation Programme at UGC-Human Resource Development Centre of Rashtrasant Tukadoji Maharaj Nagpur University during the year 2001 to 2017 (20th to 98th Batches) were selected on the basis of judgemental sampling. The study was carried out during 2017. Survey method was used. A structured questionnaire was developed having questions related to personal characteristics of teachers and the core areas of Orientation Programmes. During the Orientation Programme, lectures, seminars, group discussion, panel discussion were organised on the topics stated in Guideline of UGC Academic Staff College (2002-07), (2007-12), and Guidelines for Regional Centre for Capacity Building & Human Resource Development Centre (2015) and Guidelines for UGC- Human Resource Development Centres (2019). Hence the questions included in the questionnaire were focused in the core areas of Orientation Programme. The teachers were asked to rate themselves on improvement in domains viz., knowledge and practice of teaching skills and methodology, attitude of teachers towards students, research aptitude and skills, ICT skills, life skills and knowledge of social, economical and environmental issues on five point Likert scale. The five point Likert scale had options viz., 'very low', 'low', 'neither low nor high', 'high' and 'very high'. The data was analysed for frequency, percent and chi square test. The confidence interval was set at 95 per cent.

Results and Discussion

Personal Characteristics of Teachers

The personal characteristics of the teachers studied were age, gender, educational qualification, faculty, location of the college and the teaching experience. The results are discussed here.About 67.9 per cent teachers were male and 32.2 per cent were female. About 42.7 per cent teachers were from the age group of 41 to 50 year and 39.1 per cent from 31 to 40 years. Only 7 per cent teachers were below 30 years. Majority (52.2 per cent) teachers were from Arts and Humanities whereas 16.8 per cent and 26.5 per cent were from Commerce and Science faculties respectively. Rest of 4.5 per cent teachers were from other disciplines. The majority teachers were from urban areas (54.8 per cent) and only 32.2 per cent from rural areas however 13 per cent teachers were from semi urban areas. There were 45 per cent and 43.5 per cent UG and PG teachers having teaching experience of up to 10 years whereas 20 per cent and 49.6 per cent UG and PG teachers having 10-20 years teaching experience respectively. Only 7.3 per cent teachers had teaching experience of more than 20 years.

Teaching Related Skills

During Orientation Programmes lectures and practical sessions were organised on effective teaching, teaching and learning styles, microteaching,

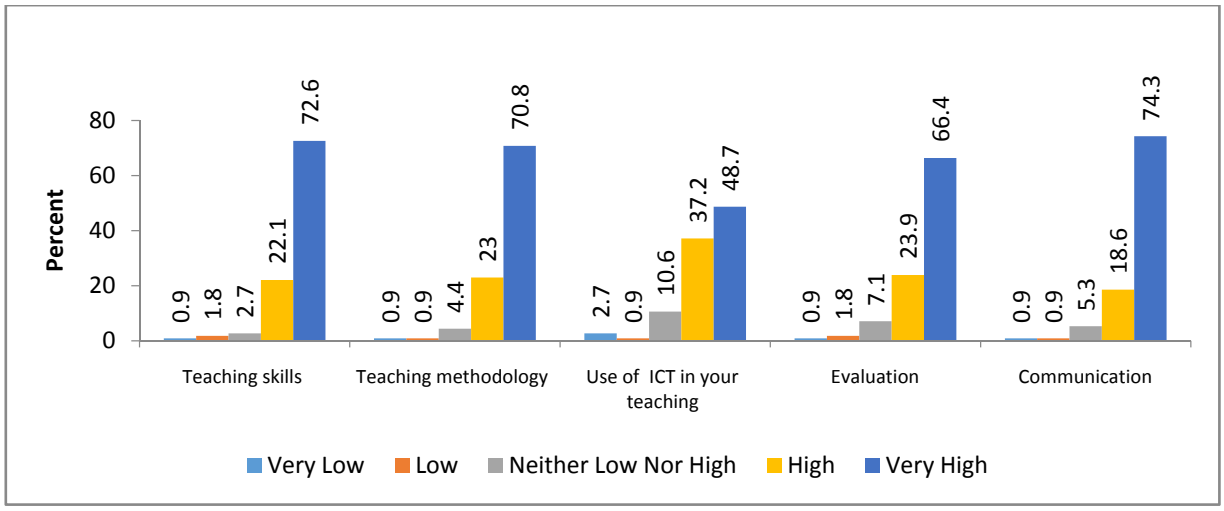
communication skills, use of ICT during teaching, learning and evaluation. The teachers were asked to rate themselves for the improvement in teaching skills, teaching methodology, use of ICT in teaching, evaluation and communication skills after attending Orientation Programme and results are presented in Figure 1.

The data presented in Figure 1 shows that the majority teachers rated themselves for 'high' and 'very high' for improvement in teaching related skills after attending Orientation Programmes. The highest number of teachers rated themselves 'very high' for improvement in communication (74.3 per cent) followed by teaching skills (72.6 per cent), teaching methodology (70.8 per cent) and evaluation (66.4 per cent). Slightly lesser percent of teachers (37.2 per cent and 48.7 per cent) rated themselves for high and very high for the use of ICT in teaching which may be because of those participants who attended Orientation Programme before 2005 and did not study computers during the course. The teachers, who rated themselves low and very low, ranged between 0.9 to 2.7 per cent. About 2.7 to 10.6 per cent teachers opted for neither low nor high for the improvement in their teaching skills.

The statistical analysis showed that teachers rating for their improvement significantly differs with respect to teaching skills (χ^2 = 212.796, p=0.000), teaching methodology (χ^2 =201.292, p=0.000), use of ICT in teaching (χ^2 = 105.717, p=0.000), evaluation (χ^2 = 171.204, p=0.000) and communication skills (χ^2 = 220.407, p=0.000). The study showed higher impact of Orientation Programme as compared to the reported studies. Satsangi *et al.* (2012) studied the impact of orientation and refresher courses on knowledge up gradation, motivation, leadership traits, technological information, communication skills, personal attributes, health awareness, stress management and national and social values on the participants. Authors reported that ASCs functioning of teaching in higher education in orientation and refresher programme were effective and its help participants in improving their teaching abilities. Similarly, Sonawane and Survey (2015) studied the impact of Orientation Programmes among the teachers. The study reported that orientation helped 10 per cent of the faculty to increase their abilities as a teacher, 9.33 per cent to improve their lecture delivery skills, 9 per cent to achieve improvement in practical and vocational skills, 8 per cent to improve their contents of teaching, only 7 per cent to improve their skills regarding student evaluation and assessment practice and 6.67 per cent of the student and classroom management skills.

Khadilkar (2016) carried out a study on role of faculty development programme and reported that out of total sampled faculty members from Maharashtra, 44 per cent of the faculty members strongly agree about improving communication skills, 50 per cent in improving teaching learning process, 44 per cent in improvement of lecture delivery, 50 per cent in improvement in contents of teaching, 36.67 per cent of the faculty members them in using innovative teaching pedagogy and 44 per cent in improving use

Figure 1: Percent Distribution of Teachers According to Improvement in Teaching Related Skills after Attending Orientation Programmes



of ICT in teaching, after attending the orientation program.

Understanding Student's Issues

Student counselling and guidance is an important core area which is covered during the Orientation Programme. In this section, the teachers were asked to rate themselves for understanding students' attitude towards learning, towards teachers, their psychological problems, peer group pressure, financial and employment problems as well as employment opportunities for the students and results are presented in Figure 2.

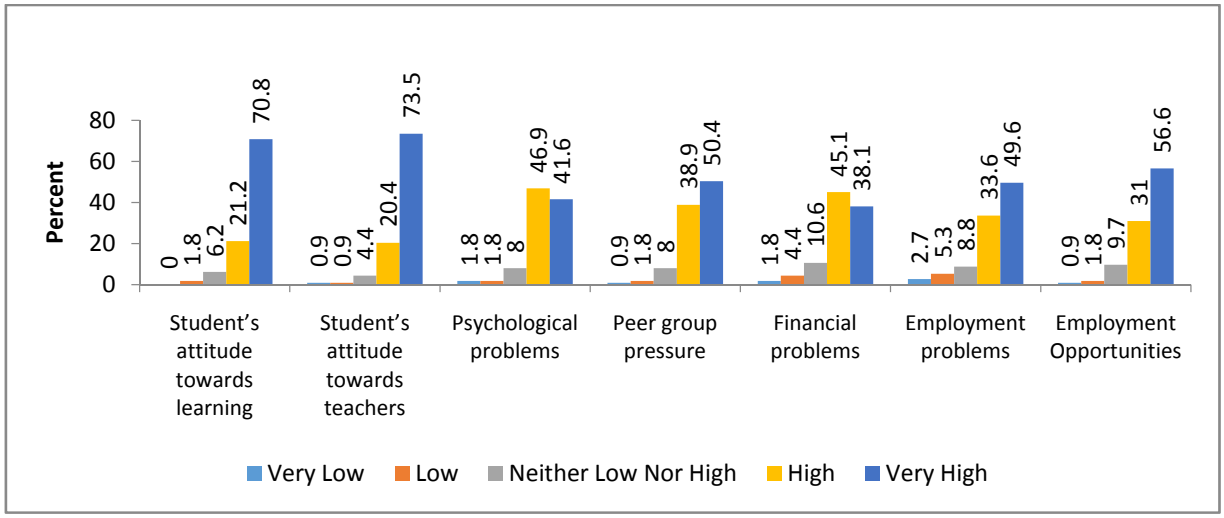
The data presented in Figure 2 indicates that the majority teachers rated 'very high' for themselves for their improvement in understanding student's attitude towards learning (70.8 per cent) and teachers (73.5 per cent), employment opportunities for students (56.6 per cent) and employment problems (49.6 per cent) after attending Orientation Programmes. The teachers understanding for peer group pressure were rated 'very high' by only 50.4 per cent teachers. In case of teachers' understanding of students psychological and financial problems the majority teachers rated themselves for 'high' (46.9 per cent and 45.1 per cent respectively) followed by 'very high' (41.6 per cent and 38.1 per cent). About zero to 2.7 per cent teachers rated for 'very low' whereas 0.9 to 5.3 per cent rated for 'low' scores. There were 4.4 to 8.8 per cent teachers who were indifferent.

The statistical analysis showed that teachers rating of themselves for improvement significantly differs with respect to understanding students' attitude towards learning (χ^2 = 135.814, p=0.000), towards teachers (χ^2 = 216.425, p=0.000), their psychological problems (χ^2 = 112.973, p=0.000), peer group pressure (χ^2 = 120.230, p=0.000), financial (χ^2 = 91.558, p=0.000) and employment problems (χ^2 = 96.071, p=0.000) as well as employment Opportunities for the students (χ^2 = 128.018, p=0.000). Several studies have been carried out on the impact of Orientation Programme on the professional development of teachers (Dhawan (2000), Rastogi Savita (2001), Joshi Suresh and Pareek Sushila (2003), Pawar I. Anand and Mouli S. Chandra, (2008), Survey and Bagul (2015), Satsangi *et al.* (2012) and Sonawane and Survey (2015)), however, none of the studies reported impact of Orientation Programmes on teachers with respect to understanding problems of their students.

Research Aptitude and Skills

Research in higher education help in discovery dissemination and interpretation of new knowledge, ideas, facts and motivate the teachers. All teachers should normally engage in research because research is essential to their professional identity (Satsangi et al.,2012). Research methodology and statistical analysis is an important core area which is covered during the Orientation Programme. The teachers are acquainted with research process, writing review of literature, bibliography, analysing

Figure 2: Percent Distribution of Teachers According to Improvement in Teachers' Understanding on Issues of Students after Attending Orientation Programmes



data and interpreting results, writing research proposals publishing research papers at national and international levels besides plagiarism, citation, statistical analysis and statistical tools, during the Orientation Programmes. In this domain, the teachers were asked to rate themselves for their improvement in research aptitude and skills after attending the Orientation Programme and the results are presented in Figure 3.

The data presented in Figure 3 shows that the majority teachers (73.5 per cent) benefitted 'very high' for searching review of literature whereas 67.3 per cent and 66.4 per cent teachers rated 'very high' for writing research proposals and research papers at national levels respectively. About 59.3 per cent, 54 per cent and 49.6 per cent teachers rated themselves very high for writing bibliography, writing research papers at international levels and analysing and interpreting data respectively.

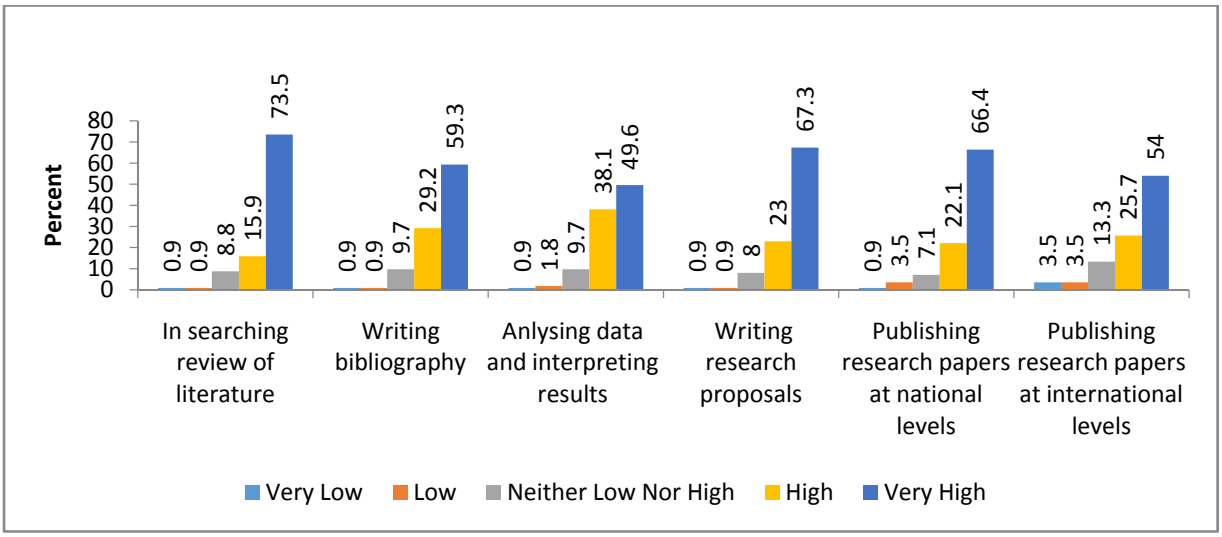
The statistical analysis showed that teachers rating of themselves for improvement significantly differs with respect to knowledge and skills of teachers viz., writing review of literature (χ^2 = 210.673, p=0.000), bibliography (χ^2 = 139.257, p=0.000), analysing data and interpreting results (χ^2 = 113.150, p=0.000), writing research proposals (χ^2 = 176.159, p=0.000), publishing research papers at national ((χ^2 = 167.133, p=0.000) and international levels (χ^2 = 100.230, p=0.000). Satsangi *et al.* (2012) studied the impact of orientation and refresher courses on teachers with respect to research aptitude of teachers viz., knowledge up-gradation, art of thinking, writing skills, technological information and research skills. It was concluded that, ASCs contribution towards inputs in research in higher education in orientation and refresher programme were effective and its help participants in improving their research skills.

Use of ICT Skills in Teaching

Information Communication Technology (ICT) is one of the important buzzwords of today's IT world. It has changed the society into information society and our way of life (Sinha, 2008). During Orientation Programmes, the teachers are acquainted with Microsoft Office. The teachers were given hands on in preparing power point presentation, using M.S. Excel for data analysis, using M.S. word for writing papers, using internet for searching reviews and communicating through emails as well as introduction to web 2.0 tools viz., Blogs/ You tube/ Face book. In this section, the teachers were asked to rate themselves for their improvement in use of ICT skills after attending the Orientation Programme and results are presented in Figure 4.

The data presented in Figure 4 shows that about 63.7 to 66.4 per cent teachers rated 'Very high' for improvement in preparing power point presentation, using M.S. word for writing papers, using internet for searching reviews and communicating through emails. Only 40.7 per cent and 47.8 per cent teachers rated "very high' for improvement in using MS

Figure 3: Percent Distribution of Teachers According to Improvement in Research Knowledge and Skills after Attending Orientation Programmes



Excel for data analysis and knowing web 2.0 tools viz., Blogs/ You tube/Face book respectively. The statistical analysis showed that teachers rating for improvement significantly differs with respect to preparing power point presentation (χ^2 = 172.973, p=0.000), using M.S. Excel for data analysis (χ^2 = 89.611, p=0.000), using MS word for writing papers (χ^2 = 104.876, p=0.000), using internet for searching reviews (χ^2 = 156.336, p=0.000) and communicating through emails (χ^2 = 87.389, p=0.000), and knowing web 2.0 tools viz., Blogs/ You tube/Face book (χ^2 = 59.080, p=0.000) after attending Orientation Programme.

Sinha (2008) studied Information Communication Technology (ICT) and internet awareness amongst the college and university teachers attending refresher course in ICT. The results showed that 90.24 per cent teachers were using IT in their classroom activities whereas only 9.75 per cent are not using IT for this purpose. 63.41 per cent teachers were comfortable in using MS-Office for their class work whereas 36.58 per cent were not able to use this. MS-Office was used frequently by 42.30 per cent teachers which are followed by 38.46 per cent for moderately use whereas 19.23 per cent were using MS-Office often for their class work. Internet literacy was very much poor among the respondents. Out of 41 participants only 36.59 per cent were Internet literate whereas maximum respondents (63.41 per cent) were Internet illiterate. While assessing the rating of Internet /e-mail services, e-mail ranks first which is followed by WWW (2nd), e-journals access (3rd), INFLIBNET Databases Search (4th) and Chatting (5th). For e-mail service, Yahoo (33.33 per cent) emerged as most preferred e-mail service which is followed by Indiatimes mail service (26.66 per cent), BSNL and Rediffmail (13.33 per cent) and 6.66 per cent for NICNET and Google mail service. 53.33 per cent respondents prefer www.goggle.com search engine which is followed by www.yahoo.com (26.66 per cent) whereas www.msn.com, www.rediffmail.com and www.indiatimesmail.com are preferred by only 6.66 per cent each.

Soft Skills

The Component D of the guidelines of ASC/HRDC states inclusion of topics related to management and personality development. During these Orientation Programmes, teachers were given training in soft skills like time management, team work, self analysis, self confidence, interpersonal skills, change management, personality development, communication skills, attitude, motivation level and stress management. Teachers were asked to rate the improvement in these soft skills after attending the Orientation Programme and the results are presented in Figure 5.

The data presented in Figure 5 shows that about 72.6 to 77.9 per cent teachers rated 'Very high' for improvement in soft skills viz., team work, communication skills, motivational level, personality development and self confidence. About 64 to 69 per cent teachers rated 'very high' for stress management,

Figure 4: Improvement in ICT Skills of Teachers after Attending Orientation Programmes

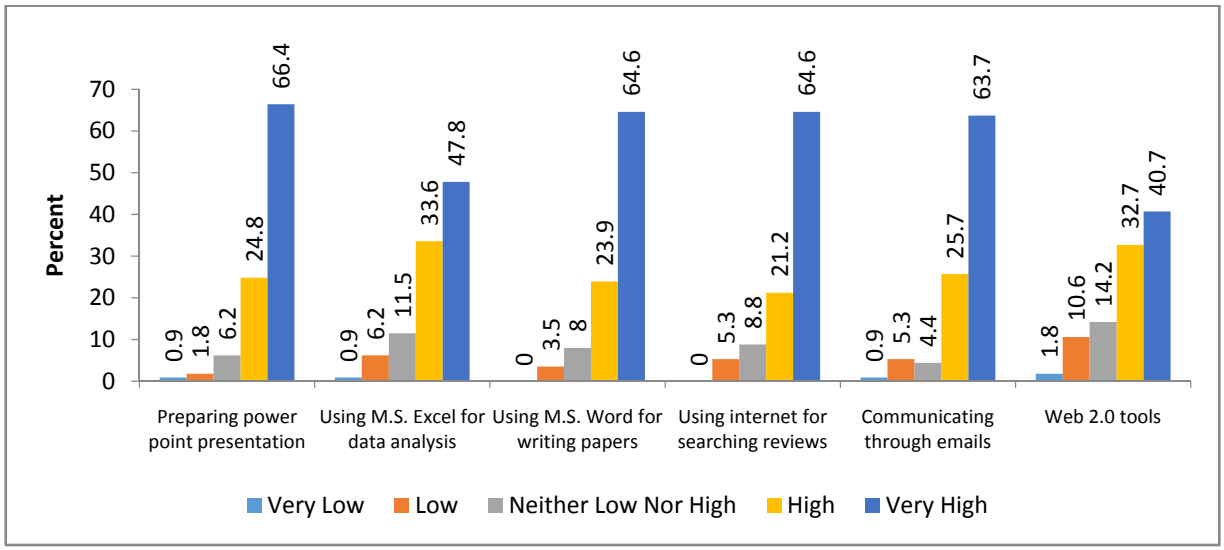
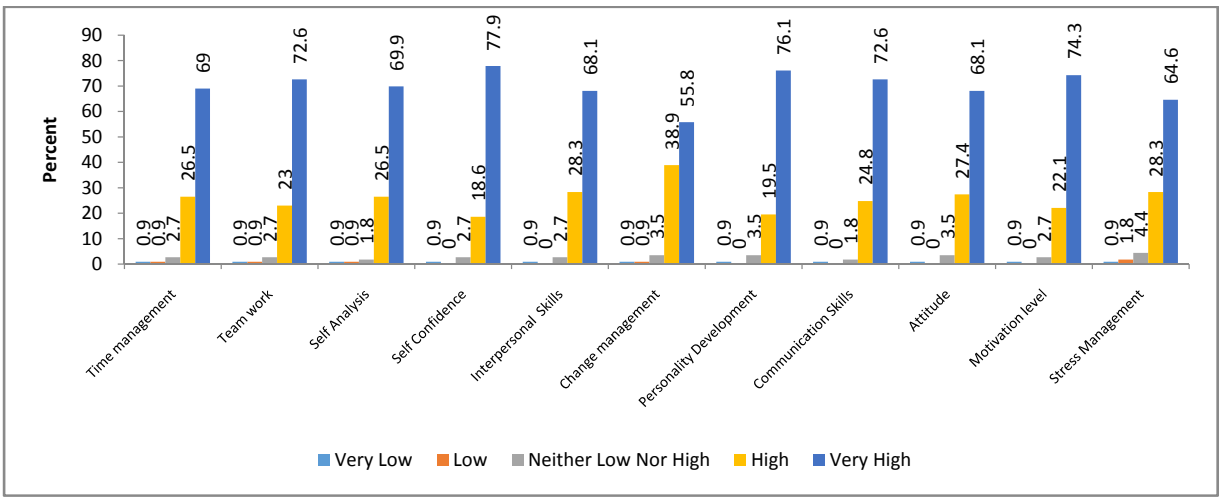


Figure 5: Percent Distribution of Teachers According to Improvement in Soft Skills after Attending Orientation Programmes



attitude, interpersonal skills, time management and self analysis. Only 55.8 per cent teachers rated for 'very high' for change management. About 0.9 to 1.8 per cent teachers rated for low and very low for improvement in all the soft skills after attending Orientation Programme. There were 1.8 to 4.4 per cent teachers who rated for 'neither high nor low' for improvement in all the skills after attending the course.

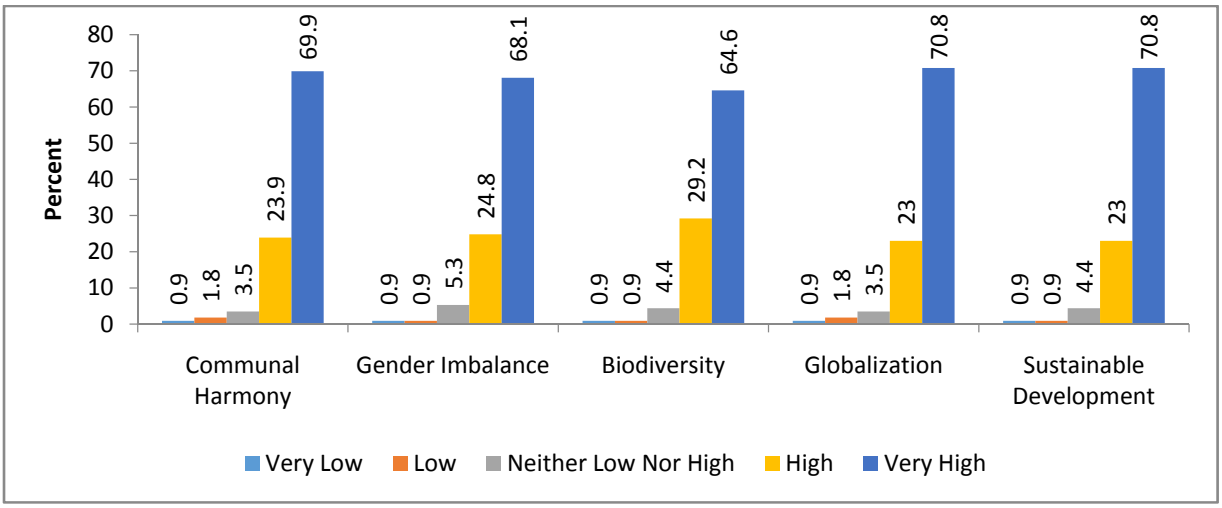
The statistical analysis showed that teachers rating for improvement significantly differs with respect to soft skills like time management (χ^2 = 196.513, p=0.000), team work (χ^2 = 214.920, p=0.000), self analysis (χ^2 = 203.239, p=0.000), self confidence (χ^2 = 177.088, p=0.000), interpersonal skills (χ^2 = 133.478, p=0.000), change management(χ^2 = 149.080, p=0.000), personality development (χ^2 = 166.540, p=0.000), communication skills (χ^2 = 152.947, p=0.000), attitude(χ^2 = 131.496, p=0.000), motivation level (χ^2 = 159.248, p=0.000), and stress management (χ^2 = 169.434, p=0.000) after attending Orientation Programme.

The extensive review of literature showed the paucity of data on improvements in soft skills of the teachers after attending Orientation Programme.

Knowledge on Social, Economical and Legal Aspects

During Orientation Programmes teachers are

Figure 6: Scores for Improvement in Teachers' Knowledge on Social, Economical and Environmental Issues after Attending Orientation Programmes



informed about social, economical, environmental and legal issues. The teachers were asked to rate themselves for improvement in their knowledge about communal harmony, gender imbalance, sustainable development, biodiversity and globalization and the results are presented in Figure 6.

The knowledge gained for social, economical and environmental issues during Orientation Programmes were rated 'very high' by 64.6 to 70.8 per cent teachers. About 70.8 per cent teachers each rated 'very high' for gain in knowledge in sustainable development and globalization whereas for communal harmony, gender imbalance and biodiversity 69.9 per cent, 68.1 per cent and 64.4 per cent teachers rated 'very high'(Figure 6). The statistical analysis showed that teachers rating for improvement significantly differs with respect to communal harmony (χ^2 = 196.336, p=0.000), gender imbalance (χ^2 = 185.717, p=0.000), sustainable development (χ^2 = 172.177, p=0.000), biodiversity (χ^2 = 201.027, p=0.000), and globalization (χ^2 = 201.292, p=0.000).

Satsangi *et al.* (2012) also studied the impact of Orientation Programmes on personal attributes, health awareness, environmental awareness, stress management and national and social values and concluded a positive impact on teachers. Sonawane and Survey (2015) reported that 5.67 per cent of the faculty members were positive about the statement that the orientation helped them in improving the interdisciplinary subject knowledge. Khadilkar (2016) also reported that out of total sampled faculty members from Maharashtra, 11.33 per cent of the faculty members strongly disagree, 8.67 per cent of the faculty members disagree, 4 per cent of the faculty members are neutral, 26 per cent of the faculty members agree and 50 per cent of the faculty members strongly agree about the orientation program helped them in improving social awareness.

Conclusion

It can be concluded from the present study that Orientation Programmes helped the teachers to improve their knowledge and practice of teaching skills and methodology, attitude of towards students, research aptitude and skills, ICT skills in imparting education, life skills and knowledge of social, economical and environmental issues. The improvement in all domains was found to be significantly (p=0.000) "very high" in likert scale in all the domains. Imparting knowledge, and skills through Orientation Programme may go a long way in improving the quality education through well trained teachers.

References

1. Dhawan Rani, (2000). Impact of Academic Staff College's programmes on Teachers and Students, *University News*, 38 (16) April 17, 2000.
2. Guidelines for Academic Staff Colleges, University Grants Commission (2002-2007) X Plan.
3. Guidelines for Academic Staff Colleges (2007-12) XI Plan.
4. Guidelines for Regional Centre for Capacity Building (RCCB) & Human Resource Development Centre (HRDC) (Implementable from 01 February, 2015) .
5. Guidelines Human Resource Development Centre (HRDC), (2019).
6. Hénard, Fabrice (2010). Learning our lesson. *Review of Quality Teaching in Higher Education.* https://www.researchgate.net/publication/44841258_Learning_our_lesson_Review_of_quality_teaching_in_higher_education
7. Joshi Suresh and Pareek Sushila, (2006). Impact of Orientation Courses on teachers in higher education. *Indian Journal of Adult Education*, New Delhi, Jan-June 2006.
8. Khadilkar Sujay Madhukar, (2016). Role of Faculty Development Program, *Scholarly Research Journal for Interdisciplinary Studies*. 3(22), 1655-1671
9. Pawar I Anand and Mouli S Chandra, (2008). Impact of training on university and college teachers: An Empirical Study. *University News*, 46 (49), 08-14, 2008.
10. Rastogi Savita (2001). Teachers Development in Higher Education in Academic Staff Colleges in India, *www.Google.co.in*. September 2009.
11. Sachin Sadashiv Surve and Dhananjay Bagul, (2015). The Orientation Programmes - Challenges Ahead. *Scholarly Research Journal for Interdisciplinary Studies*. 3(19), *375-39.*
12. Satsangi Falguni, Bhatnagar Sanjeev and Kaushik S P, (2012). An Evaluative Study of Functioning of Academic Staff College in India and Its Impact, Ph.D. thesis, Department of Management, Faculty of Social Sciences, Dayalbagh Educational Institute, Dayalbagh, Agra, https://shodhganga.inflibnet.ac.in/handle/10603/14183
13. Sheikh Younis Ahmad, (2017). Higher Education in India: Challenges and Opportunities. *Journal of Education and Practice*. 8(1), 39 - 42.
14. Sinha, Manoj (2008). Information Communication Technology (ICT) and Internet Awareness amongst the College and University Teachers. 6th International CALIBER-2008, University of Allahabad, Allahabad, February 28-29 & March1, 2008 © INFLIBNET Centre, Ahmedabad .https://www.researchgate.net/publication/269334809
15. **S**onawane Sanjeev A, and Surve Sachin Sadashiv, (2015). Professional Development of Academic Staff in Higher Education: Impact of Orientation Programmes, *Scholarly Research Journal for Interdisciplinary Studies*, 3(20), 1149-1157. ❑

Good Practices in Virtual Presentations for Higher Education

Pranita Gopal*

The COVID pandemic has forced educators tap into the potential of technology. On the face of it, technology appears to be quite useful in breaking the barriers of location and time but it steals away the infectious enthusiasm that one experiences in a live classroom. Educators today are relying heavily on virtual modes of presentation to help students and colleagues around move past these strange times. This paper attempts to understand virtual presentations and how educators can benefit from delivering engaging virtual presentations for their students.

Purpose of a Presentation

All presentations fall into three categories:

- *Inspirational:* The best example to illustrate an inspirational presentation is to think of TED Talks (www.ted.com). The inspirational presentations focus on certain problems and thereafter leave the audience with ideas and questions to think and ponder about. Using the audience's emotions and intellect, inspirational presentations put forth viewpoints that have the ability to inspire.
- *Informative:* Informative presentations on the other hand aim at sharing information – they are expected to communicate instructions, or deliver a point or share a research or simply educate the audience regarding areas of interest to either the organization (business or institution) or the audience. Educative presentations, training presentations could fall under this category.
- *Influential:* These presentations could also be called persuasive presentations where the presenter expects to the audience to accept the proposal or idea at hand or accept the solution presented to a particular problem. Sales pitch or business proposals could be an example of an influential presentation.

In education, the presentations draw heavily from the genres of inspirational to the informative presentation.

In higher education we have another lens to view a presentation. Presentations could be considered as opportunities:

- *to contribute*: where the presenter has the ability to contribute towards a discussion or a problem-solving process or the learning journey of a fellow learner.
- *to connect*: where the presenter has the ability to make connections with others present. It helps the presenter network with like-minded people and grow in their respective fields.
- *to change*: the perception of the audience towards issues or problems.

Face to Face Presentations Vs Virtual Presentations

Gendelman (2010) explains the difference between face to face presentations and virtual presentations under the following four heads:

- *Location & Time*: In face to face presentation, the location and time are fixed and the audience and the presenter are expected to be at a certain physical location at the same time, while in the case of a virtual presentation there are opportunities for the presentation to be happening synchronously (where the audience and the presenter are present virtually at the same virtual place & at the same time (irrespective of the time zones) or asynchronously (where the audience and the presenter can break away from the barriers of time and place).
- *Size of the audience*: Face to face presentations have a physical limitation with regard to the size of the audience and the logistics involved thereafter. In virtual presentations, the size of the audience is not gauged by the physical constraints of space but that of the internet bandwidth and technological platform.
- *Content delivery tools*: Often in face to face presentation, presenters are limited by the tools they can use to deliver the content (projector, LCD, video projector) as often it is difficult to create a complete multimedia experience in all presentation venues. With virtual presentations the choice of content delivery tools increases with options of screen sharing and using the internet simultaneously within the presentation. Using break-out rooms within the presentation framework allows for short group discussions without disturbing the entire group.

**Instructional Coach & Educational Consultant, O 1090 Devender Vihar, Sector 56, Gurgaon, Haryana. Email: pranitagopal@gmail.com*

- *Modes of Communication*: In face to face presentation, the mode of communication is mostly audio, that can create accessibility issues. Using closed captions in a virtual presentation can help break down the accessibility barriers.

Challenges in a Virtual Presentation in Higher Education

In spite of having advantages like, inclusion of large number of participants, lower economic costs for organizers, virtual presentations come with their own challenges. Gendelman (2010) list down the following challenges:

- *Absence of visual/ facial cues*: As a face to face presenter, we all rely on facial cues to give us reinforcements while we are presenting. In the world of virtual presentation, there is generally a lack of visual cues with regard to the audiences' mood and receptivity and these effects the enthusiasm of the presenter.
- *Struggle in building rapport with the audience:* Virtual presenters often struggle in building rapport with their audience due to internet bandwidth issues visible via the poor-quality audio or video transmission.
- *Internal distractions:* As virtual presentations are conducted using the laptop or mobile device, the presenters are always facing challenges of participants getting distracted by the emails or chats or other websites.
- *External distractions:* The audience of virtual presentations could be either logging in from their homes or office spaces where simultaneously other activities could created is tractions for the presenter and the audience. Also, another example of an external distraction could stem from the unmuting of the audio function in the device leading to distractions from the barking of pets to squabbles amongst family members.
- *Struggle in encouraging students:* In face to face presentations, the presenter can use visual cues and facial expressions to create a rapport with the participants. In a virtual world, often the audience has the ability to login to a presentation and switch off the video feed, leaving very little room for the presenter to engage with such participants.

Minimizing the Challenges while Presenting Virtually

These challenges could affect the communication channel, but with a bit of planning and conscious efforts, one can overcome the challenges, like for example:

- *Keeping the presentation short*: Unlike face to face presentations that could go on for an hour or more, virtual presentations need to be short with ample room for interaction. Although there is no fixed formula to ensure the distribution of presentation time vis a vis the question-and- answer time, a general rule of thumb followed by good presenters is to use only 60 per cent of the allotted time for presenting and the rest 40 per cent of the time for questions.
- *Using a brisk pace*: Unlike face to face presentations, in the virtual presentation, maintaining a brisk pace ensures the presentation reflects the enthusiasm of the presenter. A brisk pace doesn't mean to speak fast, instead it means that the presenter should be so well versed with the content that the presentation flows like an oft repeated story.
- *Keeping the interaction alive during presentation*: Interactivity is a key element to create memorable virtual presentations. Interaction keeps the audience actively engaged away from the various distractions.
- *Using more visuals than texts*: Using more visuals than texts helps keep the attention of the students. They ensure the presentation doesn't become dull. Free copy right images and images licensed under the creative commons can be found in sites like Unsplash, Pixabay, Wikicommons.
- *Add FIRES to the presentation:* Virtual presentations that have the necessary FIRES (F-Fresh, I – Informative, R- Relevant, E – Enthusiastic, S-Story) have the ability to engage the participants. In the case of higher education, the presenter can mix and match audio and visual clips to add freshness to the topics.

Planning & Delivering a Virtual Presentation

Every presentation has a certain purpose and that purpose is built on understanding the elements of a good presentation. Research has shown there are four elements to effective presentation:

- *Understanding the audience*: Concepts across disciplines and grade levels often overlap. Understanding the audience – both students and teachers- includes understanding the age, the role of the audience, socio-economic background, linguistic diversity, technological capability, accessibility issues etc. Understanding the audience provides the presenter a perspective to deal with the content and leads to the next important element of the presentation.

- *Preparing the content*: If we wish that our presentation is interesting and our audience is attentive during our presentation, the content needs to mirror audience needs and interests, along with the coherence of content. Students have access to all material that is available on the internet, what they lack is the ability to string the content into a tapestry that is meaningful and age appropriate.
- *Delivering confidently*: Confidently delivered presentation have "hooks" that engage the participants and create room for inter activity between the students and the teacher. One of the hallmarks of a confidently delivered presentation is linking what is being discussed to the larger context of the subject and demonstrating how small pieces of concepts have shaped the domain knowledge.
- *Controlling the environment*: Creating ground rules for the presentation, giving clear instructions on acceptable behaviors in an online environment, letting the participants know when their questions would be answered are some ways in which the presenter can control the presenting environment.

Durate (2012) describes three phases of planning and delivering a presentation. This holds true for even virtual presentations.

Phase I: Conceiving the presentation: Conceiving a presentation includes finding answers to the following questions:

- Who is the audience?
- What do I want the audience to know after hearing the presentation?
- What do I want the audience to do after hearing the presentation?
- What do I want the audience to feel after hearing the presentation?
- How does the presentation answer the question of "*WIFM – What's in it for me*?" for the audience?
- How does the presentation organize the content? Does it follow a chronological order or causal-effect or does it build on research or does it present a logical argument?
- Does the presentation flow like a story?
- What are the media modes that have been chosen for communicating the message?
- Are there alternatives for images so that accessibility issues in the audience is catered for?

Phase II: Visualizing the presentation: It is in this stage the presentation takes the form of a narrative with the help of slides or any presentation modes. While visualizing the presentation it is useful to plan answer the following questions:

- Are the visuals used free from cultural, gender or socio-economic barriers?
- Are the visuals such that they easily convey the point that the presenter is trying to make?
- Does the image help to make the presentation memorable?
- Does the image supplement the content and the topic that is being presented?
- Are the visuals too bulky in terms of their size?

It is important to remember that unlike face to face presentations, in virtual presentations, the presenter has to plan for more slides so as to keep the audience engaged.

Phase III: Delivering the presentation: There are two aspects that Durate (2012) suggests we consider at this stage – one pertains to delivery and the second pertains to the impact of the presentation. It is important to consider the following points while delivering the presentation:

- Is there enthusiasm in the presentation?
- Is the voice modulated for driving the maximum impact?
- Depending on the content and audience is the presenter's voice communicating assertiveness or caution or is it critical or is it motivational?
- Is there a way to measure the impact of the presentation?

Virtual Presentation-Good Practices for Higher Education

There is an analogy that one often reads about when a presenter's face to face and virtual presentation skills are compared. The analogy comes from the world of aviation – where pilots are not only trained but expected to have dual proficiency – first is the proficiency to fly with sight and what is visible in front of them (akin to face to face presentations where the audience is visible) and the second is the proficiency to fly with just the instruments in the cockpit – that is without any visibility of the world outside (akin to virtual presentations – where the presenter is facing just a camera and the computer screen). And therefore, like how pilots gain competency and dual proficiency following certain good practices can help a presenter gain proficiency in presenting virtually as well as in face to face environments.

- *Have a wingman*: A wingman is a supporter; a person who will assist the presenter in many ways – like to present the introduction, to monitor the chat, to monitor the questions, to steer the flow of questions to the presenter, to be have a copy of the presentation. The wingman can also keep providing useful summaries at the end of each section, giving the presenter a break from continuous speaking. If the planned presentation is of a longer duration, then the wingman can help break the monotony of a single presenter.
- Giving students the opportunity to be a wingman during the presentation, helps boost the student's confidence and creates a sense of responsibility within the student towards her learning.
- *Understand the audience*: A good practice to deliver effective virtual presentations revolves around understanding the audience composition – which institutes they belong to, what is their educational background, what is their work experience – all this information can be obtained when the audience registers for the presentation. This information also helps the presenter plan and share specific examples pertaining to certain specific areas of interests of the audience – automatically increasing the interaction amongst the participants.
- *Create a presentation title & description that will intrigue/ appeal the audience*: Presenters need to plan their topic to center around areas of interest of the audience or current need. Like for example, in today's pandemic times, any presentation that helps final year students increase their chances of getting employed will automatically appeal to a large majority of the students. In the same manner, presentations that deal with the mental health issues that students undergo due to lockdown will help students cope better as human beings.
- *Understand the presenting environment*: Virtual presenting environments have the following common features: a *presentation area* (where the presentation is available for display to the audience) , a *picture area* (where the presenter's video feed is available), a *chat area (*where the audience can chat either with the presenter or amongst themselves), a *break-out room (*where the larger groups can be broken into smaller groups for small group interactions), a *polling area (*where the presenter can create polls to understand the audience), a *screensharing area* (where the presenter can share the computer screen – this area could be overlapping with the presentation area), a *whiteboard* (which can be used during the presentation), the *audio/ video controller (*which is used to control the audio levels and the video clarity of the presentation.) With so many areas under one screen, it is important to understand the presentation environment as all apps and products have the same functions in different areas of the screen. \
- *Test the platform once before going live*: As presentation environments could vary across operating systems and amongst themselves, it is a good practice to test the presentation on any new platform before the final presentation. Also, when the presentation is tested before the final presentation, the presenter is able to gauge how the presentation fonts and images will be visible when the final presentation is screened across devices. When a presenter tests the presentation before delivering, it displays professionalism and behaviors that we wish the next generation imbibes.
- *Remember -Lag time*: Due to the internet bandwidth issues a virtual presenter needs to be aware of the lag time that could occur while presenting. The lag time is often witnessed by the audience as a slow-motion picture.
- *Avoid scroll quickly*: As many times, a lag time occurs during virtual presentations, it is important to remember that while viewing websites or programs via screen sharing, one must not scroll through the page quickly.
- *Avoid Animations*: In face to face presentations, animations if used correctly can help hold the attention of the audience. In a virtual presentation, they generally do not provide the same assistance as they could add to the lag during the presentation. Animations during virtual presentations could create unwanted distractions while moving between slides.
- *Present information in short logical chunks*: Presenting information in short logical chunks will help the presenter maintain the momentum of the presentation. It will assist the presenter move from one topic to another with confidence and agility that are hallmarks of a good presentation.
- *Using visuals*: The visuals in the presentations act as anchors that help the presenter and the audience. In a virtual world, the images help carry the talk forward. Visuals are great tools for creating transitions within the presentation topic and help the presenter support learners with diverse learning styles.

- *Promote interaction*: Polls are very useful tools to promote interaction. Polling tools can also transform into tools for checking the perception of the audience, understand the conceptual clarity of a topic. These questions can then be used as hooks to grab the attention of the audience.

  The wing man can also play an important role in adding to the interaction of the presentation by asking planned questions that help propel the discussion/ presentation ahead.

  Integrating live results or direct poling sites within a presentation can also help add to the inter activity within the presentation.

  The polling option can also be used as a tool to see the perception of the audience before and after the presentation.
- *Answer questions*: A good practice is to clearly set the stage as to when the questions would be answered. Many presenters prefer the questions are answered at the end of the presentation. Also, another good practice is to repeat the question and the name of the person who asked the question to help others understand the question in perspective. At times to add to the interactivity within the presentation, a good idea could also to be ask leading / investigative questions to the audience and then break them into smaller groups to think about the answer.
- *Keep a glass of water handy*: Like all presentations, when the speaker speaks for a longer duration, having a sip of water helps maintain the clarity of speech.
- *Do some vocal exercises*: In Julian Treasure's TED Talk (How to speak so that people want to listen) there are some exercises that help the speaker warm up her voice. These simple exercises add to the clarity of speech that helps the speaker keep the attention of the listener.
- *Reboot the computer*: Often we do not realize the number of background activities that take place in the computer and many at times the computer could update itself and start rebooting right when the presentation is on. Therefore, it is generally suggested that the computer/laptop be rebooted once before a presentation.
- *Upload supporting material*: It is also recommended not to share the presentation handouts before the presentation, instead the material shared once the presentation is over. Many presenters also share the presentation handouts/ supporting material once the presentation is over but the question and answer session is still on, giving the audience and students an opportunity to ask their questions before the event is wrapped up.
- *Gather feedback*: In the face to face situations, one is able to gather feedback based on the interaction after the presentation is over. In virtual presentations, the presenter has the provision to solicit specific feedback via forms and questionnaires. The feedback always assists the presenter in either collaborating with other like-minded colleagues or helps improve on their own presentation.

To Sum Up

In the UGC Guidelines on Examinations and Academic Calendar for the Universities in View of COVID-19 Pandemic and Subsequent Lockdown, issued in April 2020 a case for blended learning model has been put forward. The guideline also emphasizes that the responsibility of the teacher doesn't end by just sharing the e-content, instead it lays emphasis for interaction with the students.

Presenting in the virtual world and working with students and colleagues via a virtual medium necessitates that interactivity and engagement are deeply rooted within all presentations and discussions. With adequate planning and following the good practices from world over virtual presentations can be a great scaffold for in the learning journey of individuals.

References

1. Campbell, G M, (2002). Bulletproof Presentations. Career Press. New Jersey.
2. Duarte, N (2012). HBR Guide to Persuasive Presentations. Harvard Business Review Press.
3. Duarte, N (2019). Data Story: Explain Data and Inspire Action Through Story. Ideapress Publishing.
4. Gendelman, J, (2010).Virtual Presentations That Work, McGraw-Hill Education, New York.
5. Reynold, Garr (2012). Presentation Zen: Simple Ideas on Presentation Design & Delivery. 2nd Edition. Peachpit.
6. Reynold, Garr (2011). Naked Presenter. Peachpit. California.
7. Treasure, Julian (2013). TED Talk - https://www.ted.com/talks/julian_treasure_how_to_speak_so_that_people_want_to_listen.
8. University Grants Commission, (2020). UGC Guidelines on Examinations and Academic Calendar for the Universities in View of COVID-19 Pandemic and Subsequent Lockdown. From: https://www.ugc.ac.in/pdfnews/5369929_Letter-regarding-UGC-Guidelines-on-Examinations-and-Academic-Calendar.pdf. ❑

Transformations through Innovative Technological Interventions

M Venkaiah Naidu, Hon'ble Vice President of India delivered the Convocation Address at the First Convocation of National Institute of Technology, Tadepalligudem, Andhra Pradesh on December 24, 2019. He said, "The challenge I place before all of you as young technocrats today, as you embark on the next phase of your life, is to find technological solutions to combat the pressing problems. Incremental improvements and frugal innovations can be carried out constantly to bring about advancements. By solving these problems, you will create a virtuous cycle which will help pull millions out of poverty and improve the quality of their lives. Please remember that that the ultimate aim of science and technology is to make lives of the people comfortable and happy." Excerpts

I am delighted to be with all of you today at the first convocation ceremony of NIT Andhra Pradesh, a young and promising institution nestled in the pristine land of Tadepalligudem. A visit to Andhra Pradesh is like a homecoming to me. I was born here and was nurtured and nourished by this great land.

National Institutes of Technology (NITs) that have evolved from Regional Engineering Colleges (RECs) have been at the forefront of technical education in the nation. NIT AP had the privilege of being mentored by NIT Warangal, the first REC in the country. I am happy to know that NIT Andhra Pradesh is the 31st NIT in the nation. I am told that it started functioning with the support of its mentor institute, NIT Warangal, in Sri Vasavi Engineering College Campus, Tadepalligudem. I fondly remember the foundation laying ceremony of NIT Andhra Pradesh which was held on 20 August 2015 when I was the Union Minister for Urban Development and Parliamentary Affairs. It is truly noteworthy that the construction work of the campus which started in October 2018, was completed in thirteen months. The transformation of an erstwhile runway to an awe-inspiring campus of an institution of national importance is indeed remarkable. It is clear from the report presented by the Director that the youngest NIT of the country has made great strides in all areas.

My Dear Young Students, let me take this opportunity to congratulate each and every one of you for this tremendous accomplishment. You have the proud privilege of being the first batch of students to graduate from NIT, Andhra Pradesh. NIT Andhra Pradesh was born with you and will grow with you and prosper with you. As you leave the portals of this Institution today, please remember that you are amongst the brightest young minds of our country. You live in a very exciting era of scientific and technological disruptions.

Technological breakthroughs in the form of Automation, Artificial Intelligence, Internet of Things, Big Data and analytics have truly transformed the way we live and work. The existing frontiers of science are being constantly challenged through interaction between various disciplines ranging from arts to humanities to engineering to Biotechnology. Young engineers like yourself have the never-before opportunity to further revolutionize these technologies and find new ways to use them for the benefit of humanity. Let me stress that the ultimate aim of all these technological advances must be the betterment of the life of the common man. They must lead to the discovery of solutions to the most pressing problems of our time.

There is no doubt in my mind that innovation is the watchword for the 21st Century. Institutions such as the IITs and the NITs must become hubs of innovation. They must recognize potential and nurture it to bring out the best in each student. The curricula and teaching methods in these institutions must never be set in stone. They must constantly evolve to suit the times. Institutes that render technical education must never hesitate to teach and experiment with the latest of technologies. They must allow the free flow of ideas and must never stifle innovative thinking through rigid protocols. Let me focus on a few areas that I think can be transformed through innovative technological interventions.

The first is agriculture. There is a vast scope to introduce new and innovative technologies to spur agriculture growth and improve the lot of the farmers. This is the time for young engineers to come up with out-of-box solutions to the problems faced by agriculturists from climate change to increasing crop productivity. I urge young technocrats to understand the real problems faced by Indian agriculture, including fragmented land holdings and insufficient market access.

I want you to think:

- Can we understand soil better?
- Can we carry out more efficient irrigation by overcoming shortages and avoiding surpluses?
- Indian agriculture is greatly dependent on weather. Keeping the changing climate and precipitation change in mind, can we create better weather prediction systems and make agriculture become more resilient?
- Can we facilitate better price discovery and market access to farmers?

It is of paramount importance that we boost our food production to overcome hunger and feed a billion people. I have always maintained that imported food security is never a solution. The second is the arena of clean and renewable energy. Our climate patterns and weather systems are changing. Global Warming is a reality and it is happening now. We cannot be in denial anymore.

India is one of the most carbon-efficient economies in the world today. The government of India under the leadership of our visionary Prime Minister is all set to surpass the target we had set for ourselves of 175 GW of renewable energy by 2022. Our Solar capacity has increased by eight times between FY14-18 and we have spearheaded the International Solar Alliance (ISA), a testimony to India's commitment to the development of 'zero emission sources'. I want engineers and technocrats like you to champion the cause of clean energy.

I strongly believe that a balance between environment and development can be achieved through good technology. It is essential that we strike that balance. Our development has to be sustainable. Let conservation be the cornerstone of every single one of your innovations.

The third is urban development. India's urban population is expected to rise to 517 million by 2020 and will cross the 700 million mark by 2050. Our urban spaces are our engines of growth. Unfortunately, our towns and cities have exceeded their carrying capacity and are choking. Our most prosperous cities are also home to large slums. Our cities must grow inclusively and sustainably. We must find solutions for sustainable housing, drinking water, mass mobility and provide economic opportunities to bridge the huge income gaps in cities. We must address concerns of pollution and congestion and create sustainable waste management solutions and waste-to-wealth initiatives. We must make our urban spaces safe for women and children and accessible to the differently-abled.

We must check distress migration by providing quality healthcare, education, other amenities, and economic opportunities in rural areas as well. Urban development is possible only if we achieve rural development, a cause that is very close to my heart.

Mahatma Gandhi also believed that self-sufficient villages are necessary if we are to build a prosperous nation. Our rural areas too need technological interventions that are tailor-made to suit their unique requirements. I also urge you to find solutions to challenges of water management. Many parts of India are facing an acute water crisis. This crisis will only worsen in the time to come.

According to the Composite Water Management Index (CWMI) report released by the Niti Aayog in 2018, 21 major cities might reach zero groundwater levels by 2020. The Union government recently formed a new Jal Shakti (water) ministry, which aims at tackling water issues with a holistic and integrated perspective. The ministry has announced an ambitious plan to provide piped water connections to every household in India by 2024. The government alone will not be able to solve the entire problem. We must augment the efforts of the government.

We must use technology for better rainwater harvesting and to improve water use efficiency. We must urgently transit from this 'supply-and-supply-more water' provision to measures which lead towards improving water use efficiency. The challenge I place before all of you as young technocrats today, as you embark on the next phase of your life, is to find technological solutions to combat these pressing problems. I am hopeful that you will come up with solutions to these challenges through hard work and creative thinking. You do not have to wait for big-ticket innovations. Incremental improvements and frugal innovations can be carried out constantly to bring about advancements. By solving these problems, you will create a virtuous cycle which will help pull millions out of poverty and improve the quality of their lives. Please remember that that the ultimate aim of science and technology is to make lives of the people comfortable and happy.

My Dear Youngsters, In spite of temporary global slowdown, the Indian economy has the potential to become a $ 5 economy in the coming years as the macroeconomic fundamentals are strong and resilient. Government is also taking various reforms from time to time to improve GDP growth, cut down red tape and attract more investments.

A bright future awaits all of you. All the stakeholders in the country, particularly youngsters like you, must strive to transform the nation and the lives of the people in tune with the three-word mantra of the Prime Minister, Shri Narendra Bhai Modi to *'Reform. Perform and Transform'.*

Even as our technical education institutions such as NITs equip young technocrats with such skills, also take care to nurture life skills and moral values in them. Always remember the words of Mahatma Gandhi. He said that education without character is a sin. Every graduate passing out of the portals of this institution must not only be academically proficient, they must also be ethical, compassionate and honest individuals.

Before I conclude, I must raise one more issue. I read with mounting concern reports about young students resorting to extreme measures because of the inability to handle stress and cope with challenges. Our campuses must be the safest spaces in our nation, for in our campuses, our future is cast. I urge our higher educational institutions to take extra care to teach our youngsters to manage stress. I understand that the curriculum is heavy. But students must be taught to take time out. We must make sure that our students are physically fit and mentally happy.

My Dear Young Friends, never hesitate to reach out to your friends, teachers or loved ones when you are in need of help or support. Be optimistic and embrace the challenges that await you in your life and career with a balanced mind. The practice of Yoga and meditation will greatly help you to deal with stressful situations with poise and equanimity. I am confident that you will make the impossible possible with your grit and determination.

Let me congratulate all of you once again. Go forth and create a bright, beautiful future for you, our nation and the world. ❑

ATTENTION : SUBSCRIBERS UNIVERSITY NEWS

The NEW RATES of Subscriptions effective April 01, 2020 shall be as per following:

	Institutions Rs.	Teachers/Students/Individuals* Rs.
1 year	**1,250.00**	**500.00**
2 years	**2,200.00**	**900.00**

***AT RESIDETNAIL ADDRESS ONLY**

The payable amount is required to be remitted **in advance** by any of the following modes of payments:

a) AIU WEB Portal (b) Cash Deposit (c) Demand Draft/At Par Cheque and (d) NEFT/RTGS/Net Banking/G-Pay/ BHIM APP, etc.

1	**Bank Account No.**	**0158101000975 (Saving)**
2	Beneficiary Name and Address	**ASSOCIATION OF INDIAN UNIVERSITIES**
		16, Comrade Indrajit Gupta Marg, New Delhi – 110 002
3	Bank & Branch Name	CANARA BANK, DDU MARG
4	Bank's Address	"URDU GHAR", 212, Deen Dayal Upadhyaya Marg, New Delhi – 110 002
5	Branch Code	0158
6	IFSC Code	CNRB 0000158
7	Contact No. & E-mail ID	(011) 23230059 Extn. 208/213 (M) 09818621761

THE NEFT/RTGS/ONLINE PAYMENT TRANSACTION/UTR NUMBER MUST BE SENT BY MAIL IMMEDIATELY WITH COMPLETE MAILING ADDRESS & PIN CODE FOR LINKING AND ITS SETTLEMENT AT OUR END.

For further information/enquiries, ***send Mail at : subsun@aiu.ac.in / publicationsales@aiu.ac.in***
Website : ***https://www.aiu.ac.in***

COMMUNICATION

Innovative Initiatives Taken for Reforming Conduct of Education During COVID-19 Pandemic Lockdown Period

Sajal Dasgupta*

We are going through a very difficult situation now due to this Corona virus invasion, COVID 19 pandemic. Regular traditional teaching learning process (face to face) has been completely halted. We do not know when things will normalize. Fortunately, we had taken lots of exemplary initiatives in last two years to go digital in line with Digital India Initiatives launched by Government of India. In the month of Feb end 2020 when we realized the seriousness of the situation and anticipated a nationwide preventive drive is going to be implemented shortly to encounter the outbreak , our statutory bodies decided to confront the threat, crisis and panic created due to COVID pandemic with a positive attitude of transformational leadership to avail its opportunities and become self reliant through a paradigm shift from offline to online mode of education in such a manner so that there is no loss of available academic hours as per our academic calendar being followed strictly before the pandemic period. The very objective of these initiatives taken was not to reinvent education but to continue it uninterruptedly with a different methodology.

As per the directive of the Government we had notified to suspend all face to face classes from 16th March 2020 and closed down the University on and from 23rd March 2020. Since then all our employees including faculty are locked down in home as per the decision of the govt. to protect against Covid-19 pandemic. We firmly believed that if we delay the teaching learning process and examination and declaring results, enormous damage will be done to the families who are waiting for their child to financially supplement family income. About 90 percent of our final year students have got job offers. Families of these students are looking forward eagerly for time when their ward will start supplementing family income. Employer company will not wait further.

**Sajal Dasgupta, Vice Chancellor, University of Engineering and Management, Kolkata, West Bengal- 700160, E-mail: dasguptasajal@yahoo.com.*

In order to maintain our academic calendar unaffected and to protect the career of our students, we decided to continue our teaching learning process through digital platform. Around 2000 lectures done by our teachers were videotaped. Our teachers and students collectively practiced mock examinations on digital platforms. University authority was committed to ensure no students loose time due to corona virus invasions.

Accordingly, since 16th March 2020 all our faculty were sincerely engaged in classroom teaching online and session / lab classes in virtual mode following our daily academic class routine. The syllabi for even semester in 2019 -20 academic session completed satisfactorily. The daily attendance of faculty, progress of syllabi and % attendance of students were monitored daily through online mode. An academic audit was conducted online by the IQAC of our university to oversee the quality of online education and timely completion of courses which was found to be satisfactory in nature. We were confident to come out of the crisis period with complete victory.

During this lockdown period all over India due to COVID-19 pandemic, our group of institutions have taken lot of innovative initiatives to continue academic as well as students/ faculty development activities through digital platform. Some of the initiatives are mentioned hereunder

- Digital Attendance of Faculty- Online morning and afternoon meeting daily with all faculty(which is one of the good practice being followed in all our group of institutions since inception) through youtube live, Goto Webinar- Everyday all faculty meet via Youtube Live, Goto Webinar in the morning and afternoon, comprising of all 500+ members of the group and everyday departmental meetings and faculty assignments reported via Google meet are monitored and considered as a part of compliance for virtual attendance of all faculty and officers.

- Online classes through Google Hangout meet, YouTube live , every day as per routine were held. The digital tools used are - Youtube Live, Google Classroom, IEMCRP (our proprietary educational ERP software), Simulators, like, MATLAB, NS-2, GridSim, Cloud Sim etc., Online Coding platforms like, Hacker Rank, Code Chef, Hacker Earth etc.
- Online Labs by virtual mode using MATLab, Hacker Rank - online coding platform
- Online lecture by corporate experts- With the continuous persuasion and strong initiatives of various departments corporate connect programme are being conducted through online sessions in the evening for our students, thereby not hampering their everyday WFH schedule, and much to our delight, many of them have responded spontaneously and a good number of sessions have been done which are immensely beneficial for students making them aware and ready for the upcoming battle in their career.
- University have conducted online mid-term exam for students as a part of continuous assessment and prepared for end semester exam. These exams were held where students sitting in their home with pen and paper answered the. descriptive questions sent to them digitally . This was not an open book system of examination. Questions taken from question banks were framed in such a manner so that no two question papers were identical. Answer scripts were scanned by the students with mobile phone and uploaded digitally for evaluation. Teachers took help of zoom and later google meet for proctoring (a group of students under one mentor) so that students cannot take help of others during examination. This hybrid on line system with 4500 students during our 2nd term online exam was introduced and found to be very effective . Almost 95 percent students attended the examination without any problem. Students will later submit the hard copies of answer script for keeping record.
- With this hybrid on line system of examinations tested and proven to be very effective , our university has recently completed end semester examinations using the same model for final year students successfully as per the approval of our statutory bodies and in accordance with UGC/ AICTE guidelines . Almost 100 percent students have appeared for the examination. Rest of the students will get another opportunity to appear for examination after the situation normalizes.
- University has facilitated development of faculty members during this testing time by organizing online FDPs, which are taken by eminent academicians of India.
- Online Admission- New batch admission process has been made completely online by making entrance test, IEMJEE online, followed by online counselling-cum-interview session.
- Online Placement- Many recruiters have resorted to online recruitment process and our current final year students have coped up with those opportunities well and grabbed the job offers. ❑

CAMPUS NEWS

National Webinar on Education in Post-pandemic Era

A ten-day National Webinar on 'Education in Post-pandemic Era : Challenges and Opportunities' was organized by the School of Education, Ravenshaw University, Cuttack during June 4-13, 2020. A total of 2452 participants registered from almost all the states and Union territories of the country and from more than 100 universities. The nature of participants was PG/ UG/B.Ed. students, research scholars, school teachers, NGO officials, Government officials, faculty members, *etc.*

In the inaugural session, Dr. Sudarshan Mishra, Head, Department of Education and Organizing Secretary of the Webinar read the theme of the seminar and introduced the guest. Prof. Ishan K Patro, Vice Chancellor, Ravenshaw University, in his inaugural address deliberated upon the problems and opportunities during post-pandemic context. He encouraged participants to rely on self-study which will help in developing higher order thinking skills. He also encouraged the faculty members to motivate the students for self-study. He said that during this pandemic context, the role of the teacher has been changed from mere instructor to a facilitator who will play supportive role for motivating students for self-study. He also emphasized upon continuous assessment of learners which will also help the learners for self-study as well as self-assessment. He cited some practical examples which the universities will going to face during post-pandemic time such as, sanitization of hostel, university campus, departments, maintaining social distancing norms, resuming teaching and examination, *etc.*

Every day one resource person is invited to deliberate upon a topic related to the theme of the webinar. Dr. Sudarshan Mishra delivered a talk on 'Open Educational Resources, Its Need, History and Development, Advantages and Challenges'. Dr. Manas Ranjan Panigrahi, Sr. Programme Officer, Commonwealth Educational Media Centre for Asia (CEMCA), New Delhi delivered a talk on 'SWAYAM MOOCs: An Opportunities for Self-Learning' who focused on teaching-learning through SWAYAM, development of course proposal for SWAYAM, *etc.* Prof. G C Nanda, Senior ICSSR Fellow, Department of Education, Ravenshaw University delivered a talk on 'Autonomy and Accountability in Higher Education' who emphasized upon levels of autonomy, how to promote autonomy and ensure accountability. Prof. Dhaneswar Harichandan, Former Professor cum Director, Institute of Distance and Open Learning, University of Mumbai, Mumbai delivered a talk on 'ICT and MOOCs through SWAYAM'. He gave a conceptual orientation about ICT and step by step procedure for development of MOOCs on SWAYAM. Dr. Amita Puri, Director, Optimus Centre for Wellbeing, Citizen Hospital and De addiction Centre, Gurgaon delivered a talk on 'Developing Resilience and Positive Mental Health during COVID Times'. She discussed about stress and practicing skills in developing unique styles of reacting to stress. Prof. Bharti Dogra, School of Education, Indira Gandhi National Open University, New Delhi delivered a talk on 'Blended Learning and UGC Guidelines, Blending Models, Credit Equivalency, Online Activities in Blended Learning', and how to adopt blended learning. Prof. C P Nanda, Department of History, Ravenshaw University delivered a talk on 'State, People, and Technology: A Regime of New Normality in Post-COVID World Order'. He said that the crisis that neoliberalism faces as a result of COVID-19 is much more deep rooted. He said it is time to reset the certitudes of neoliberalism, including that being pedaled in education. While technology is an important component of life today, the larger issues of life and life world must be through by educationist today. This requires planetary rather than an anthropogenic perspective, a reset of view from body and matter to perhaps soul, vitality and joy. Dr. T K Pany, Associate Professor, Department of Commerce, Ravenshaw University delivered a talk on 'Challenges of Services Sectors with Reference to Higher Education'. Prof. Sanjay Satapathy, Dean, School of Commerce and Business Studies, Ravenshaw University delivered a talk on 'Nourishing Personality'. He highlighted on effective time management, study skills and future classroom.

In the valedictory session, Dr. Sudarshan Mishra read the webinar Report. Prof. Nityananda Pradhan, Principal, Regional Institute of Education, Bhopal delivered the valedictory speech on 'Pedagogy 2.0 in the Era of Education 4.0'. In his talk, he highlighted about history of Revolution in Education, Framework of Pedagogy 2.0, key pedagogical practices, flipped classroom, critical pedagogy, and capacity building of

teachers. In Valedictory Address, Prof. C B Sharma, School of Education, IGNOU, New Delhi said that digital medium can only change the Post-COVID-19 education system. He said that we could not use the technology optimally in education in the first twenty years of twenty-first century. This is the right time to use digital technology in education. In future, teachers will be the most celebrated personality in the media and they will be the highest paid professionals. The world will listen to them. Prof CP Nanda, Department of History, Ravenshaw University, in his presidential address recognized the challenges and opportunities that come with the future education system and hoped for a rapid change in the model of education reform and the distribution of educational processes in the coming days.

Conference on Creativity, Innovation and New Business Models

A two-day Conference on 'Creativity, Innovation and New Business Models for Reviving the Economy' is being organised by TAPMI School of Business, Faculty of Management and Commerce, Manipal University Jaipur in association with Manipal Instituite of Management, Manipal Academy of Higher Education, Manipal and Atal Incubation Centre, Manipal University, Jaipur during August 27-28, 2020 at Jaipur.

Economies all over the world are witnessing prolonged slowdown beset by structural, global, and cyclical factors resulting in of job losses; trade wars and the subsequent reduction in trade flows across the international borders, record-high unemployment rates, liquidity crisis, crashing demand for consumer goods, and pessimistic investor sentiments. A large chunk of businesses is coming under severe stress and strain to sustain or stay afloat in the market. It even requires sustainable business innovations and reforms to withstand the economic storm. Economic revival in developing countries would undoubtedly require pro-growth initiatives, evolving business models and development initiatives to boost the economy. The role of creativity and innovation in recognizing hitherto unexplored opportunities for the introduction of new techniques, new products, and new services needs no emphasis on rejuvenating entrepreneurs are critical components of society which take calculated risks, manage uncertainty and engage in revamping economic activities in explored terrains. Sustainable growth combined with a progressive distribution of income is achievable through inclusive and innovative business models. The themes of the event are:

- Finance and Accounting.
- Economics and Health Care.
- Marketing.
- Human Resource Management.
- Operation Management.

For further details contact, Organising Secretary, TAPMI School of Business, Manipal University Jaipur-303007, Rajasthan, Phone: 0141-3999100 Ext.: 317, 868, 378, E-mail: samita.sharma@jaipur.manipal.edu. For updates, log on to: www. jaipur.manipal.edu

Online Workshop on Basics of C and C++ Programming

A five-day Online Workshop on 'Basics of C and C++ Programming' is being organized by the Department of Mathematics, Rajiv Gandhi University Arunachal Pradesh in collaboration with Spoken Tutorial Project, IIT Bombay during July 15-19, 2020. The students of universities/colleges, research scholars, and academicians may participate in the event.

C is middle level programming language which combines the features of both high level and low level languages. It was developed by Dennis Ritchie during the early 1970s while working at AT&T Bell Labs in USA. The need to redesign the UNIX operating system to enable it to be used on multiple computers led to the development of C. C is a highly portable structured general-purpose programming language which is used for scripting system applications which form a major part of Windows, UNIX and Linux operating system. It also efficiently works on enterprise applications, games, graphics, and applications requiring calculations. C++ is a highly portable language and is often the language of choice for multi-device, multi-platform app development. In fact, C++ is an extended object-oriented programming language which includes classes, inheritance, polymorphism, data abstraction and encapsulation. C++ also allows exception handling, and function overloading which are not possible in C. Being a powerful, efficient and fast language, it finds a wide range of applications - from GUI applications to 3D graphics for games to real-time mathematical simulations. On account of the wide significance of C and C++ programming languages, it has become an integral part of course syllabus in most of the higher educational institutions around the world. Now a day, it helps a great deal in research oriented works in the field of mathematics, Physics, Computer Science, *etc.*

For further details, contact Convener, Dr. Nipen Saikia, Assistant Professor, Department of Mathematics,

Rajiv Gandhi University, Arunachal Pradesh-791112, Mobile: +91 9435507484, E-mail: nipen.saikia@rgu.ac.in.

International Web Lecture Series on Qualitative Research Methodology

A Ten-day Online Faculty Development Programme on 'Crafting Qualitative Research: How can We Study the World as it is Lived?' is being organised by the University of Science and Technology, Meghalaya from July 15-24, 2020. The Programme is particularly for faculties, research scholars especially interested to know about qualitative research. Independent researchers and senior master degree students may also participate in the event.

For further details, contact Convenor, Amit Choudhury, Professor and Dean, School of Business Sciences, University of Science and Technology, Meghalaya-793101, Mobile: +918761007133/ + 919854453170, E-mail: sbuswebls@gmail.com. For updates, log on to: www.ustm.ac.in.

AIU NEWS

West Zone Vice Chancellors' Meet–2019-20
on
Internationalization of Higher Education and Global Rankings

The West Zone Vice Chancellors Meet (2019-20) of the Association of Indian Universities (AIU), was hosted by IIS University, Jaipur on December 11th-12th, 2019. Shri Kalraj Mishra, Hon'ble Governor of Rajasthan graced the Inaugural Session as Chief Guest. Prof. Bhushan Patwardhan, Vice Chairman, UGC was the Guest of Honour and Prof. M M Salunkhe, President AIU, presided over the Meet. Dr. Pankaj Mittal, Secretary General, AIU Convened the Meet. Prof. Ashok Gupta, Vice Chancellor of the IIS University was the Co-convener of the Meet. Dr Alok Mishra, Joint Secretary, AIU was the Nodal Officer of the Meet and Mr Vijendra Kumar took care of logistics and coordination. Dr S Rama Devi Pani, Editor University News was the Nodal Officer for Academics. Dr. Raakhi Gupta, Registrar IIS University was the Nodal Officer of the Meet at host University. The Meet was attended by 80 Vice Chancellors from Gujarat, Maharashtra and Rajasthan apart from international experts from World Bank, Shastri Indo-Canadian Institute, Times Higher Education, British Council, South Asian University & IIM Ahmadabad. Main theme of the Meet was *'Internationalization Of Higher Education and Global Rankings'***.** Under the main theme, following three technical sessions were conducted in the Meet: ***'Promoting Indian Higher Education Abroad and Building International Collaborations'***; *'Improving International Rankings of Indian Universities through Internationalization'; 'Attracting International Students and Faculty to India--Issues and Challenges'*. In addition to Inaugural Session, Valedictory Session and three Technical Sessions, the Meet had AIU Business Session and newly induced Session on Interaction with Apex Bodies. The University had arranged Cultural Evening.

The Chief Guest; Hon'ble Governor of Rajasthan, Shri Kalraj Mishra initiated his in his inaugural address reminded the audience of their constitutional duties and appealed for taking pledge to abide by them. In his peech, he highlighted the role of academia in sensitizing students towards constitutional duties. Teachers should inculcate in students strong value system and must ensure that students not only understand their fundamental duties but also perform them religiously. He emphasised that concerted efforts are required by the Indian universities to be competitive in this era of globalization. Higher education should be more employment oriented and steps should be taken to prevent brain drain from India, he said. Chief Guest released the Special Issue of the University News edited by Dr S Rama Devi Pani was released to commemorate the occasion.

Prof. Bhushan Patwardhan stressed on the pressing need for internationalization of Indian HEIs to survive in this globalized world. He deliberated on various initiatives taken by UGC and AIU to give impetus to internationalization of higher education. He recommended an indispensable shift in Indian Higher Education System; to be student-centric rather than teacher-centric. Prof. M M Salunkhe, in his presidential address focused upon the need for internationalization of education. He envisioned the transformation of Indian higher education through

"Study in India" programme of MHRD.Dr. Pankaj Mittal, Secretary General, AIU, New Delhi outlined AIU's accomplishments in providing value added services to the member universities, apart from bringing visibility to AIU by creating Twitter handle, digitization of the AIU Library, creating portals for admissions, jobs for the students, and also facilitating collaboration with the overseas universities to encourage global partnerships. Dr. Ashok Gupta expressed confidence that the deliberations during the meet will go a long way in redefining operational strategies in Indian higher education within the national boundaries and beyond.

Session on Interface with Apex bodies

The Interface with officers from Apex bodies was chaired by Dr. Bhushan Patwardhan, Vice Chairman, UGC and Dr. Pankaj Mittal, Secretary General, AIU was the Co-Chair. Prof. Amiya Kumar Rath, Adviser represented NAAC. Discussing about the roles AIU, UGC and other apex bodies of education, Dr. Patwardhan suggested that consistent communication between AIU and other apex bodies is very significant in maintain the relevance and standard of higher education. He also shared information about the proposed National Academic Credit Bank.

Prof. Rath during the discourse, shared success stories of NAAC focusing on two major aspects i.e. changes during more than two decades of NAAC's existence and Revised Accreditation Framework. He deliberated on the IIQA application process, SSR submission readiness guidelines, distribution of metrics, key indicators across each criterion, reforms such as data based qualitative indicator evaluation; data validation by external professional agencies, instituting key indicators on alumni engagement and student satisfaction survey. Dr. Mittal threw light on various schemes of UGC like ARPIT, IMPRESS, MOOC's and emphasized that community engagement programmes should be undertaken by the universities to fulfill their social responsibilities. The important take aways from the session are:

1. The universities should undertake curriculum revision based on learning outcomes as revised under the UGC guidelines. Focusing on relevant, updated and learning outcome-based curriculum will prepare the HEIs as well as students for futuristic roles and competing in an international system.
2. Expert committee should be appointed by the universities to assess the quality of research papers published by the faculty. Research profile of a faculty members should not be adjudged on the basis of number of research papers rather it should be based on the quality of research work.
3. NBA accredits only few selected programs; NAAC accredits all the programs but there is dichotomy in the mandates of accreditation agencies which creates confusion. There should be a single accrediting agency.
4. NAAC manuals should provide examples with reference benchmarks.
5. UGC search webpage should be more informative with detailed information about each scheme.
6. IIMs & IITs should handhold the general universities, including self-financing universities which are not having good grades in NAAC to enable them to learn and adopt their best practices. AIU may facilitate interactions and signing of MoUs between the aforesaid institutions.
7. Many rankings such as ICAR, NIRF provide well laid down and transparent benchmarking. Accordingly, NAAC benchmarks be made available to help in self-assessment of university, evaluate their SWOT and make directional efforts for improvement in their grading.
8. More international accreditations should be introduced in India.
9. The fees of some self financing universities is too high, causing financial hardship to parents. To ease this out, institutions may provide avenues to 'Earn-while-learn'. 'Earn-while-learn' should be introduced in all universities. The data of students' earning during degree programs should be one of the criteria for NAAC assessment. It should be essential for getting A+ grading.

Session on Promoting Indian Higher Education Abroad and Building International Collaborations

The first Technical Session; on the theme 'Promoting Indian Higher Education Abroad and Building International Collaborations' was chaired by Prof. A. M. Pathan, former President, AIU and former Vice Chancellor, Maulana Azad

National Urdu University and Central University of Karnataka. The speakers of the session were Prof. Francisco Marmolejo, World Bank's Lead Tertiary Education Specialist, Prof. Manjula Rao, Director, Higher Education and Society, British Council and Dr. Ajit Motwani, Head, Development Office, IIM Ahmedabad, former Director EdCIL, and former India Director, Institute of International Education, IIIE, New York. Prof. Pathan introduced the topic, defining the concept of Internationalization of higher education and outlined the place that India holds in pan-national education.

Prof. Francisco emphasized that India's education system is extremely dynamic and increasingly getting complex. He discussed key trends prevalent in global higher education i.e. uneven expansion, relevance and internationalization of education. Speaking on the need for comprehensive Internationalization, he also conversed on the possible solutions, the most important of which was internationalization of the curriculum. He opined that India's position in terms of capacity, innovation and large contributions might help in making room for international collaborations. His presentation concluded by expressing need for universities to be future ready in terms of constantly changing focus and paradigms in accordance with the challenges being faced by entire tertiary education system in the years to come.

Prof. Manjula Rao discussed the practicalities of internationalization of higher education in India. She stressed on the need for universities to explore each other, forge partnerships and sustain these corporations for mutual benefits. Dr. Rao also emphasized the need for understanding the perception of international audience about the Indian education, brand building and accordingly universities being proactive. Dr. Ajit Motwani opined that collaborations cannot be over emphasized. He stated that the roadmap for internationalization must be different for each university / Institution. The recommendations of the session are:

1. Regional Committees of universities located in an area should be formed to facilitate collaborative regional studies, universities' strengths, contribution, USP etc. and should explore; how their internationalization can be promoted with respect to mobility, course content, cultural exchange, projects and research.
2. Universities must create international facilities in the campuses to ensure comfortable and harmonious stay of Indian and foreign students together.
3. Universities should essentially invite experts from foreign universities for workshops, seminars, examinations and various other activities to promote Indian education.
4. Government should ensure proper distribution of foreign students among universities.
5. Government should invite and arrange some free trips to foreign students and faculty to India to promote Indian higher education.
6. Dual degree at postgraduation level in collaboration with foreign universities other than parent university should be facilitated. AIU should ensure equivalence of dual degrees.
7. Regulatory authorities should act as facilitators instead of controllers or inspectors.
8. Internationalization requires a lot of support from the Government and embassies for visa & FRRO with police department. Therefore, Government should make the process easy, fast and smooth.
9. AIU should popularize the initiatives of the Government of India towards promoting Indian higher education abroad and international collaboration through workshops, seminars and University News.
10. Indian Government should create permanent country-wise platform involving Indian diaspora to promote Indian higher education abroad.
11. Government should launch promotional campaigns like 'Incredible India' to promote best Indian universities/institutes of higher education abroad.
12. Barriers/challenges need to be identified that hinder the inward mobility of foreign students to Indian universities.
13. The skills and perspectives of international students on campus need to be understood and shared with other students.
14. Alumni to be leveraged significantly so that Institute builds its brand in other countries; alumni being the spokesperson.

Session on 'Improving International Rankings of Indian Universities through Internationalization'

The Technical Session on 'Improving International Rankings of Indian Universities

through Internationalization' was chaired by Prof. Sandeep Sancheti, former President of AIU and Vice Chancellor, SRM Institute of Science and Technology, Tamil Nadu. The panelists of the session were Mr. Ritin Malhotra, Regional Director (South Asia) Times Higher Education and Prof. Kamlesh Mishra, Vice Chancellor, Rishihood University, Bahalgarh. In his opening remarks, Prof. Sancheti enumerated various advantages and disadvantages of rankings.

Mr. Ritin Malhotra discussed the framework of Times Higher Education (THE) ranking and parameters for internationalization of colleges. He stated that the alumni are the goodwill ambassadors of academic excellence of an institution and the reputation of an institution is built by its alumni. Mr. Kamlesh Mishra shared his experience in evolution of ranking systems and the challenges faced by the system. Following recommendations emerged from the session:

1. Industry & Education sector are lacking in research activities. There is need to encourage the research initiatives in both the sectors which complement each other. Emphasis should be laid on research collaboration, academic reputation and citation impact
2. Yoga courses should be introduced in universities as a science with both theoretical & practical component and not just as a casual practice for international acceptance.
3. While trying to compete with other players in the world, we need to have an established system with defined guidelines, vision and mission. The standards, goals and priorities need to be set and be followed in principle accordingly.
4. Ranking and ratings do not necessarily indicate quality in real terms. More emphasis should be laid on quality of the professionals that are being produced in the universities. Balance should be done by universities on the priorities and strengths not only for ranking but to elevate the overall standards of institutions.
5. Universities should look at the changing education dynamics in international scenario and align their syllabus in line with emerging challenges with a definite industry interface.
6. In order to raise the standards of the universities, there is an urgent need to groom and mentor the skills of faculty members for their overall development through faculty exchange programmes and tie-ups with national and international universities. Student exchange programs within the country and with different nations also need to be organized.
7. There is a need to hire thinkers and innovators with enhanced skills at par with global standards. National programmes like 'Make in India' need to expand their horizons further, to engage indigenous professionals, provide employment and curb brain drain. Close collaboration with international media is also the need of the hour.
8. The mapping of expertise and expectations is very important. Students need to be equipped with proper resources of knowledge and skill. Universities should be proactive in communication, specifically involving research activities and research outputs.
9. The prime objective of the universities should be addressing local problems and align themselves with the development of surrounding areas and resident communities. The reputation of a university is largely impacted by their community outreach programme, wellbeing of neighboring area and its inhabitants. Internationally, universities should strengthen their public relations and media presence.
10. There should be well defined criteria on attracting international students to India. Serious and scaled efforts need to be made to help Indian Higher Education brand building and accelerate inward students' mobility.
11. HEIs must concentrate on quality research and low impact factor in research needs to be improved. HEIs need to focus on entrepreneurship and being job creators instead of job seekers.

Session on Attracting International Students and Faculty to India: Issues and Challenges.

The third technical session on 'Attracting International Students and Faculty to India: Issues and Challenges' was chaired by Prof. A V S Ramesh Chandra, President, South Asian University, New Delhi. The speakers of the session were Dr R.L. Raina, VC, JKLU, Jaipur and Dr. Prachi Kaul, Director, Shastri Indo- Canadian Institute, New Delhi.

The session began with insightful observations from Prof. Chandra who remarked at the changing perceptions about Indians in foreign countries. He stated that Indians professionals represent values such as hard work, trust, teamwork and loyalty. He opined that comfortable and affordable campus life and universal acceptability of grades are some of the important factors that play a crucial role in attracting foreign students. He remarked that Indian institutions are striving hard to improve the quality of education. Prof. Raina reckoned multiple factors; playing vital role in improving the quality of education; for it to compete internationally. He opined that enriching classroom experience with focus on learning, pedagogy, curriculum, quality of faculty, evaluation process, and innovative courses can significantly improve the quality of Indian Higher Education.

Dr Prachi Kaul, Director, Shastri Indo Canadian Institute forwarded suggestions that can help the universities in attracting foreign students and faculties. She stressed on the need to outreach at a global level by marketing abroad. She reiterated that flexibility in grading, curricula and training can attract foreign students. She also recommended that HEIs need to work on employability factor and bring in required change in academic processes and systems. Major recommendations of the session are:

1. Internationalization of higher education can be augmented by creating Non-Resident Indians (NRI) and foreign students' quota on supernumerary basis in HEIs.
2. Academic credit exchange should be systematized and seamlessly aligned so that uniform standards of credit transfer can be developed among countries. This will facilitate International students' inward and outward mobility.
3. Research domains should be prioritized for international collaboration. The priority areas must be identified keeping in view the international trends of academic research.
4. Internationalization of higher education mandates developing world class infrastructure, amenable academic environment enabling exchange of students, teachers, academic programmes, high quality research projects and conducive research ambience.
5. Encouraging regulations are essential to facilitate cross border movement of academic stakeholders in higher education. The draft education policy must clearly spell out regulations covering all aspects of Internationalization.
6. There is an immediate need to inculcate cultural sensitivity among academia to seamlessly amalgamate international stakeholders. Adequate safety and security provisions should be ensured on the campus. Home Stay, involving international students in various cultural activities and familiarizing them with indigenous culture are some of the desirable steps to encourage inward mobility. The counseling bureaus of the universities must organising workshops and orientation programmes time to time for cultural orientation of foreign students.
7. Institutional arrangements are mandatory to address various concerns of out-bound and in-bound students. The policy must include standardized, tangible modalities; compulsory for all universities admitting international students. An effective and speedy grievance redressal mechanism must be in place to address the concerns of international students.
8. Internationalization of higher education can be achieved by striking a balance between indigenous and global knowledge systems. New academic programs must have international orientation without losing sight of the indigenous values. Existing programmes may be redesigned to accommodate both indigenous and international components.
9. Policies should be formulated to support institutional collaborations for comprehensive advancement of higher education as a universal public good.
10. Indianisation of higher education should be first step toward internationalization. Indianisation can only be feasible by removing regional barriers and state reservations in the country.
11. At present around, five lac students go out of India to study in foreign HEIs. As compared to this; only 54000 foreign students are admitted in Indian HEIs. The inflow of foreign students needs to be maximized by building world class infrastructure, simplifying visa rules, easy entry and exit provisions, good hostel

facilities, curricular revision and international orientation, etc.

12. There is a need to attract foreign faculty and incentivizing them not only financially but also stimulating their intellect by interacting with cross-cultural, bright students and colleagues.
13. It is high time for Indian higher education institutions to achieve desired credibility and visibility among national and global circles. Collective efforts need to be made by the Government, apex regulatory bodies and universities.
14. Enriching experiences by students and faculty inside as well as outside the classrooms are critical factors to attract foreign faculty and students to HEIs.
15. Apart from the educational infrastructure and faculty upgradation programs, India should consider sensitizing Indian diplomatic mission abroad to make concerted efforts to attract foreign students; looking for educational opportunity in India.
16. MHRD, apex regulatory bodies and other ministries should work in tandem for ease of policies, regulations, providing required support, infrastructure as well as human resource support to foreign students coming to study in Indian universities.
17. AIU should provide platform to the universities to share best practices and experiences; worth emulating from leading global universities. University News may have column of best practices in which every week best practices adopted by one best university can be brought out.
18. Universities should rope in and actively involve Indian embassies in highlighting strengths of Indian HE system.
19. It is evident that the number of students admitted in Indian HEIs and supported by Indian Government scholarships is not impressive. In fact, more than half the students come from Afghanistan. Entire foreign students' policy needs to be revised taking into consideration prevailing geopolitical scenario, availability of higher education system in the respective countries and India's relationship with different countries.
20. Nowadays, education has also become commercial entity. Hence, in order to compete in international market, role of policy makers is as important as that of universities. At university level, infrastructure, faculty, hostel facilities and research structure should be enriched to attract, adapt and sustain foreign students in Indian conditions. Cost of education should also be competitive.
21. Academic hubs should be developed in India to attract foreign students.

Valedictory Session

Dr. Subhash Garg, Minister of Technical Education, Rajasthan graced the valedictory session as the Chief Guest. The session was chaired by Prof. Sandeep Sancheti, former President AIU. Ms. Shuchi Sharma, Secretary Higher Education, Government of Rajasthan, Prof. Vivek Saoji, Vice Chancellor, KLE University, Belgam & Governing Council Member, AIU were the Guests of Honour. Dr Pankaj Mittal, Secretary General, AIU convened the Session and Dr. Ashok Gupta, Vice Chancellor, IIS University felicitated the Guests and Officers of AIU. Dr S Rama Devi Pani presented the Report of the Meet. The Meet ended with vote of thanks by Dr Alok Mishra, Joint Secretary, AIU.

. ❑

THESES OF THE MONTH

SCIENCE & TECHNOLOGY

A List of doctoral theses accepted by Indian Universities (Notifications received in AIU during the month of Feb-March, 2020)

AGRICULTURAL & VETERINARY SCIENCES

Agricultural Economics

1. Kausadikar, Harsha Hemant. **Economics of production and marketing of sweet orange in Marathwada Region of Maharashtra**. (Dr. S R Nagargoje), Department of Agricultural Economic, Vasantrao Naik Marathwada Agricultural University, Parbhani.

2. Srikanth, Bandi. **Econometric analysis of production and marketing of maize in Telangana State**. (Dr. K V Deshmukh), Department of Agricultural Economic, Vasantrao Naik Marathwada Agricultural University, Parbhani.

Biotechnology

1. Malik, Palvi. **Genome wide selection for rapid introgression of productivity traits from *Oryza rufipogon* into O sativa**. Department of Biotechnology, Punjab Agricultural University, Ludhiana.

2. Verma, Annu. **Gene networks involved in cold stress response in chickpea (Cicer arietinum L) anthers**. (Dr. H K Chaudhary), Department of Agricultural Biotechnology, CSK Himachal Pradesh Krishi Vishvavidyalaya, Palampur.

Entomology

1. Bokan, Sanjog Chandrakant. **Seasonal incidence, crop loss assessment and bioefficacy of insecticides against major insect pests of soybean (*Glycine max* (L) Merrill)**. (Dr. P R Zanwar), Department of Agricultural Entomology, Vasantrao Naik Marathwada Agricultural University, Parbhani.

Genetics & Plant Breeding

1. Mohd Shamshad. **Genetic and physiological basis of nitrogen uptake and utilization in wheat (*Triticum aestivum* L)**. Department of Plant Breeding & Genetics, Punjab Agricultural University, Ludhiana.

Horticulture

1. Damodhar, Vijay Pundlik. **Effect of nutrients and plant growth regulators on yield, quality and post-harvest life of banana Cv grand naine**. (Dr. G M Waghmare), Department of Horticulture, Vasantrao Naik Marathwada Agricultural University, Parbhani.

2. Ingole, Revati Hemantrao. **Studies on nutrient management in response to growth, yield, quality, shelf life and value addition in papaya (*Carica papaya* L)**. (Dr. G M Waghmare), Department of Horticulture, Vasantrao Naik Marathwada Agricultural University, Parbhani.

3. Zoremtluangi, Jenny. **Nutrient management in Khasi mandarin (Citrus reticulata blanco) under subtropical agroclimatic condition in Mizoram**. (Dr. Debashis Mandal), Department of Horticulture, Mizoram University, Aizawl.

Veterinary Science

1. Kondampati, Kasturi. **Effect of different anaesthetic protocols on the plasma levels and disposition kinetics of ampicillin-cloxacillin combination in horses**. Department of Veterinary Pharmacology and Toxicology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana.

2. Mir, Insha. **Aflatoxin contamination in cattle feed value chain and its impact on milk quality in Punjab**. Department of Veterinary and Animal Husbandry Extension Education, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana.

3. Sharma, Deepti. **Ultrasonographic evaluation of canine liver and kidney with special reference to shear wave elastography in health and disease**. Department of Veterinary Surgery and Radiology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana.

4. Srinu, B. **Epidemiological study on the anthropogenic disseminated environmental pollutants in the aquatic ecosystem and their association with antimicrobial resistance emergence**. Department of Veterinary Public Health & Epidemiology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana.

BIOLOGICAL SCIENCES

Bioinformatics

1. Tripathi, Swetakumari Pauharisharan. **Identification and characterization of potential drug targets for glioblastoma: An *in-silico* approach**. (Dr. Rakesh M Rawal), Department of Bioinformatics, Gujarat University, Ahmedabad.

Biotechnology

1. Gangadasu, Seetharavamma. **Development of monoclonal antibody based ELISA and Lateral Flow Strip (LFS) for detection of cry 1 Ac Bt protein in genetically modified cotton plants**. (Dr. P Sudhakar), Department of Biotechnology, Acharya Nagarjuna University, Nagarjuna Nagar.

Botany

1. Patel, Svetaben Bharatbhai. **Genetic diversity assessment of *Acmella paniculata* (Wall Ex Dc) R K Jansen collected from different regions**. (Dr. H A Solanki), Department of Botany, Gujarat University, Ahmedabad.

Life Science

1. Raval, Nidhi Pankaj. **Identification and validation of Y-sperm specific markers for semen sexing in dairy animals**. (Dr. Linz-Buoy George), Department of Life Science, Gujarat University, Ahmedabad.

2. Sharma, Bhagwati Sharan. **Comparison of Locus Control Regions (LCRs) and designing of a system to understand regulatory potential in complex gene expression**. (Dr. R J Verma), Department of Human Genetics, Gujarat University, Ahmedabad.

Zoology

1. Umrigar, Namrata Rajesh. **Biodiversity and ethological study of water birds in and around Surat City**. (Dr. Arunkumar H Dholakia), Department of Zoology, Veer Narmad South Gujarat University, Surat.

EARTH SYSTEM SCIENCES

Environmental Science

1. Parmar, Payal Vinod. **Biodiversity analysis and habitat suitability studies of certain important wetlands of Gujarat using aquatic macroinvertebrates as the niche indicators**. (Dr. Rita N Kumar), Sophisticated Instrument Center for Applied Research and Testing, Sardar Patel University, Anand.

2. Shakti. **Preparation characteriazation and pharmacological evaluation of polyphenolic nanoformulations from camellia sinensis**. (Dr. Neeraj Dilbaghi and Dr. Dinesh Dhingra), Department of Bio and Nano Technology, Guru Jambheshwar University of Science & Technology, Hisar.

ENGINEERING SCIENCES

Civil Engineering

1. Amitha, N R. **Some studies on strength behaviour of concrete due to improper casting sequence**. (Dr C Sashidhar), Department of Civil Engineering, Jawaharlal Nehru Technological University Anantapur, Ananthapuramu.

Computer Science & Engineering

1. Bhavsar, Shreyaben Bhupendrabhai. **Multilingual multimedia based crops disease management system using mobile technology**. (Dr. Dipti B Shah), P G Department of Computer Science & Technology, Sardar Patel University, Anand.

2. Naralasetti, Neelima. **Cost and profit analysis of web application deployment on cloud**. (Dr. Bobba Basaveswara Rao), Department of Computer Science & Engineering, Acharya Nagarjuna University, Nagarjuna Nagar.

3. Preeti. **Metaheuristic techniques for feature selection in face recognition**. (Prof. Dinesh Kumar), Department of Computer Science & Engineering, Guru Jambheshwar University of Science & Technology, Hisar.

Electrical & Electronics Engineering

1. Dhamecha, Vishal Dineshkumar. **Studies on compound semiconductor ($ZnGa_2Se_4$) thin film electronic devices**. (Dr. Basumati H Patel), P G Department of Electronics, Sardar Patel University, Anand.

Electronics & Communication Engineering

1. Kanna, T Venkatesh. **Analysis of wavelet extensions for epileptic seizure detection using support vector machine**. Department of Electronics & Communication Engineering, Hindustan Institute of Technology and Science, Chennai.

Mechanical Engineering

1. Mishra, Santosh Kumar. **Numerical studies on non-Newtonian fluid flow through pipe using CFD**. (Dr. Arun Arora and Dr. H S Chandra), Department of Mechanical Engineering, Chhattisgarh Swami Vivekanand Technical University, Bhilai.

MATHEMATICAL SCIENCES

Mathematics

1. Rabish Kumar. **Spectral theory and Eigen functions expansions for a pair of second order matrix differential equations**. (Dr. Amar Kumar), Department of Mathematics, T M Bhagalpur University, Bhagalpur.

2. Sunita Rani. **Deformation of a stratified poroelastic media due to surface loads and buried sources**. (Prof. Sunita Rani), Department of Mathematics, Guru Jambheshwar University of Science & Technology, Hisar.

MEDICAL SCIENCES

Biotechnology

1. Salappa, Arun Ramling. **Solubility enhancement of poorly water soluble simvastatin and fenofibrate**. (Dr. E Jayachandran), Department of Biotechnology, Acharya Nagarjuna University, Nagarjuna Nagar.

2. Zodinpuii, Doris. **Characterization of clinically significant mutations associated with breast cancer in Mizo population**. (Prof. N Senthil Kumar), Department of Biotechnology, Mizoram University, Aizawl.

Nursing

1. Kori, Asha Deepak. **Community Initiatives of corporate social responsibility in engineering organizations in Mumbai: A study of women beneficiaries**. (Prof. Vandana Chakrabarti), Department of Nursing, S.N.D.T. Women's University, Mumbai.

2. Waghe, Aakanksha Ashok. **Effect of planned teaching on KAP of mothers and nutritional status of infant/children**. (Dr. Shakuntala Prabu), Faculty of Science and Technology, S.N.D.T. Women's University, Mumbai.

Pathology

1. Panicker, Soumya Mohanan. **A study on liquid-based cytology-new possibilities in the diagnosis of cervical lesions in India**. (Dr. Menka H Shah), Department of Pathology, Sardar Patel University, Anand.

Pharmaceutical Science

1. Hadia, Jayeshkumar Jasubhai. **Development and evaluation of colloidal drug delivery system containing**

poorly water soluble drug. (Dr. Ashvin V Dudhrejiya), Department of Pharmacy, Saurashtra University, Rajkot.

2. Pandey, Ajit Kumar. **Synthesis and biological evaluation of some thiadiazole derivatives and validation by UV spectrophotometric method**. (Dr. Pranita Kashyap and Dr. Chanchal Deep Kaur), Faculty of Pharmacy, Biomedical Engineering and Biotechnology, Chhattisgarh Swami Vivekanand Technical University, Bhilai.

3. Samir Odhavajibhai, Rabadiya. **Effect of selective histone deacetylase inhibitors on cardiac and renal complications associated with typeI I diabetes mellitus in rats**. (Dr. Ashvin V Dudhrejiya), Department of Pharmacy, Saurashtra University, Rajkot.

4. Sawarkar, Hemant Arunrao. **Phytochemical and pharmacological investigation of traditionally used herbs for treatment of Buccal ulcer**. (Dr. Pranita Kashyap and Dr. Chanchal Deep Kaur), Faculty of Pharmacy, Biomedical Engineering and Biotechnology, Chhattisgarh Swami Vivekanand Technical University, Bhilai.

5. Sheokand, Sneha. **Development of nanocrystalline solid dispersion of hesperetin and its oral formulation, with enhanced oral bioavailability for prevention of breast cancer**. (Dr. Arvind K Bansal), Department of Pharmaceutics, National Institute of Pharmaceutical Education and Research, Mohali.

Physiology

1. Vaidya, Lopa Madhusudan. **A comparative study of audiogram in ear sound devices users and non-users in age group of 18-40 years**. (Dr. Nirupama J Shah), Department of Physiology, Gujarat University, Ahmedabad.

PHYSICAL SCIENCES

Chemistry

1. Jasminder Singh. **Study on CO_2 capture using nanostructure carbon adsorbents**. (Dr. Soumen Basu and Dr. Haripada Bhunia), School of Chemistry and Bio-Chemistry, Thapar Institute of Engineering and Technology, Patiala.

2. Kamani, Ronakkumar Dhanjibhai. **Isatin and pyrazole clubbed heterocycles: Synthesis, characterization and biological studies**. (Dr. K H Patel and Dr. D K Raval), P G Department of Chemistry, Sardar Patel University, Anand.

3. Kanthecha, Darshana Natvarlal. **Synthesis, characterization and biological investigations of some novel gold (III) and ruthenium (III) complexes**. (Dr. M N Patel), P G Department of Chemistry, Sardar Patel University, Anand.

4. Prakash, Ajay. **Transition metal complexes of schiff bases and their biological studies**. (Prof. Rajesh Malhotra), Department of Chemistry, Guru Jambheshwar University of Science & Technology, Hisar.

5. Savaliya, Ravikumar Bhagvanjibhai. **Synthesis, characterization and application of graft copolymers and its blends with thermoplastics polymers**. (Dr. M M Raj), P G Department of Chemistry, Sardar Patel University, Anand.

Physics

1. John, Jinya. **Study of cloud microphysical properties over the Indian Region using satellite data**. (Dr. V Sathiyamoorthy), Department of Physics, Gujarat University, Ahmedabad.

2. Patel, Hiral Timin. **A comprehensive study of physical properties of some simple and non-simple metals by using an effective and local pseudopotentials**. (Dr. P R Vyas), Department of Physics, Gujarat University, Ahmedabad.

3. Singh, Rachana. **AB-initio modeling and performance study of light emitting electro chemical cells**. (Dr. Mohan L Verma), Department of Applied Physics, Chhattisgarh Swami Vivekanand Technical University, Bhilai.

4. Tiwary, Prem Pyari. **Laser and plasma interaction at high power laser flux**. (Prof. Vibha Rani and Prof. R P Sharma), Department of Physics and Computer Science, Dayalbagh Educational Institute, Agra. ❑

Bishop Chulaparambil Memorial College (BCM College)

Kottayam, Kerala -686001 | Tel. 0481-2562171

www.bcmcollege.ac.in

Affiliated to Mahatma Gandhi University, Kottayam

Applications are invited from qualified candidates for appointments to the permanent posts of **Assistant Professor** in the following disciplines, subject to the approval from Mahatma Gandhi University, Kottayam/Govt. of Kerala.

Subject	Category & No. of Posts
Zoology	Community - 1
Botany	Community - 1
	Open Merit - 1

Qualification, Age and Approval etc as per Kerala Government/UGC/University norms. Apply to the Manager within one month from the date of publication of this advertisement in the prescribed form which can be had from the college office on payment of Rs. 1000/- (Rs. 1100/- by post).

Sd/-, Manager

ASSOCIATION OF INDIAN UNIVERSITIES

ADVERTISEMENT TARIFF

UNIVERSITY NEWS JOURNAL

GST AT PRESENT RATE OF 5% IS PAYABLE FOR PUBLICATION OF ALL TYPES OF ADVERTISEMENTS IN UNIVERSITY NEWS W.E.F. APRIL 01, 2020 IN ADDITION TO THE PAYABLE CHARGE FOR EACH CATEGORY AS GIVEN BELOW)

A. FOR EDUCATIONAL INSTITUTIONS, GOVT. ORGANIZATIONS, PUBLISHERS, BOOK SELLERS & DISTRIBUTORS DISPLAY

(Amount in Rupees)

Categories of Advertisement	1 Insertion	4 Insertions	8 Insertions	12 Insertions
Full Page	15000	45000	85000	120000
Half Page	8000	28000	50000	68000
Quarter Page	5000	16000	28000	40000
Cover (Inside)	16000	55000	100000	144000
Cover (Back)	20000	65000	120000	165000

B. TARIFF FOR SPECIAL NATURE OF MATTERS/ITEMS (DOUBLE THE RATES)

TARIFF FOR SUPPLIERS OF COMPUTERS, COMPUTER STATIONERY & PERIPHERALS, SCIENTIFIC & SURGICAL INSTRUMENTS, SPORTS GOODS AND OTHERS (*NOT COVERED IN ANY FORM OF THE TARIFF)* WILL BE AT DOUBLE THE RATES AND TARIFF CAN BE HAD ON REQUEST.

C. CONCESSIONAL TARIFF (For Publishers/Book Distributors- Exclusively for Books)

Per Square Cm (Display)	1 Insertion	4 Insertions	8 Insertions	12 Insertions
	30.00	28.00	26.00	24.00

MECHANICAL DATA OF JOURNAL

Size of Page 21 cms x 27 cms

PRINT AREA

Full Page 23 cms (Height) x 16.5 cms (Width)
Half Page 12 cms (Height) x 16.5 cms (Width)
Quarter Page 11 cms (Height) x 8 cms (Width)

The Art Work/CRC IN PDF as per above Print Area (in BLACK & WHITE ONLY) or the OPEN FILE IN MS WORD may be sent accordingly. The text is required to be sent positively on E-Mail IDs as shown below. *MATTER FOR ADVERTISEMENT MUST REACH SEVEN (07) DAYS IN ADVANCE FROM THE DATE PUBLICATION OF EACH ISSUE OF UNIVERSITY NEWS*, WHICH IS EVERY MONDAY.

ADVERTISEMENT AGENCIES (INS ACCREDITED) ARE ALLOWED 15% DISCOUNT.

Full advance payment could be sent by Demand Draft/At Par Cheque favouring "ASSOCIATION OF INDIAN UNIVERSITIES", payable at New Delhi. Also, the details for payment via NEFT/RTGS are available in AIU Website. Also, the required data could be provided on request.

***For further information write to* :-**

Publication & Sales Division
Association of Indian Universities
AIU House, 16, Comrade Indrajit Gupta Marg, New Delhi - 110 002
EPABX : 011-23230059 (Extn. 208/213), FAX : 011-23232131
E-mail IDs : advtun@aiu.ac.in /publicationsales@aiu.ac.in
Website : http://www.aiu.ac.in

Towards Social & Engaging Peer Learning: Predicting Backchanneling and Disengagement in Children

Mononito Goswami*, Minkush Manuja*, Maitree Leekha*

arXiv:2007.11346v1 [cs.HC] 22 Jul 2020

***Abstract*—Social robots and interactive computer applications have the potential to foster early language development in young children by acting as peer learning companions. However, studies have found that children only trust robots which behave in a natural and interpersonal manner. To help robots come across as engaging and attentive peer learning companions, we develop models to predict whether the listener will lose attention (Listener Disengagement Prediction, LDP) and the extent to which a robot should generate backchanneling responses (Backchanneling Extent Prediction, BEP) in the next few seconds. We pose LDP and BEP as time series classification problems and conduct several experiments to assess the impact of different time series characteristics and feature sets on the predictive performance of our model. Using statistics & machine learning, we also examine which socio-demographic factors influence the amount of time children spend backchanneling and listening to their peers. To lend interpretability to our models, we also analyzed critical features responsible for their predictive performance. Our experiments revealed the utility of multimodal features such as pupil dilation, blink rate, head movements, facial action units which have never been used before. We also found that the dynamics of time series features are rich predictors of listener disengagement and backchanneling.**

***Index Terms*—Educational Data Mining, Time Series Classification, Backchanneling, Engagement, Permutation Feature Importance, Interpretability, Machine Learning, Deep Learning**

I. Introduction

MANY intervention studies have shown that engaging in narratives can help children prepare to succeed in formal education in school [Curenton, 2010]. Conversational activities like storytelling play a pivotal role in children's early language development. Storytelling, for instance, nurtures children's narrative skills by forcing them to combine multiple ideas in the form of sentences into a coherent whole. Moreover, early conversational exchanges that children have with their parents and peers significantly influences their narrative skills. Peterson and Jesso [Peterson et al., 1999] revealed that children whose mothers encouraged longer narratives through backchannel responses and open-ended questions, had a better vocabulary and overall narrative skills in the long term [Peterson et al., 1999]. Narrative conversations with peers can also cultivate children's communication skills much in the same way. Furthermore, peer learning (and tutoring) is not only beneficial in theory [Topping and Ehly, 1998], and many studies have reiterated the positive outcomes of peer learning for children belonging to low-income households and urban areas [Tymms et al., 2011].

Given the benefits of peer learning, can we design peer learning companions to support early language learning in children? Studies in the past have shown that interactive computer applications and social robots indeed have the potential to educate young children [McReynolds et al., 2020, Mostow et al., 2003, Aist and Mostow, 1997, Aist et al., 2001, Movellan et al., 2009]. Children not only trust social robots and believe that they have emotions [Kahn Jr et al., 2012], but also readily learn new information from them [Movellan et al., 2009]. However, recent studies found that children not only trust '*social*' robots more [Breazeal et al., 2016], attend to them and find them more attentive, but are also distracted by non-socially contingent robots [Park et al., 2017a].

For robots or tablet applications to be socially-contingent, they must interact in a 'natural' and interpersonal manner. Having a natural conversation is a product of a successful, continuous and back-and-forth process of speaker cues and listener backchannels. Backchanneling responses are subtle verbal or non-verbal signals that indicate the speaker that the listener is following the conversation [Yngve, 1970]. In this paper, we focus on making robots more natural and engaging. While our long term goal is to create social and engaging peer learning companions to foster *children's* early language development, our methods are reasonably generic and can be readily adapted for adult-adult interactions.

In this work, we introduce two models to help robots come across as active listeners and engaging storytellers. An active listener must consistently communicate with the speaker through backchanneling responses [Park et al., 2017a]. On the other hand, a good storyteller must be able to predict when a listener will lose attention and prevent disengagement by making an engaging gesture [Sidner et al., 2005]. To this end, we develop two models: a (*i*) listener disengagement prediction (LDP) to anticipate listener disengagement, and a (*ii*) backchanneling extent prediction (BEP) model to predict the extent to which a robot should generate backchanneling responses in the next few (3) seconds. We also carry out multiple statistical and machine learning experiments to shed light on the influence of socio-demographic and economic factors on the amount of time children spend listening and backchanneling. Finally, in order to explain the decisions made

*All authors contributed equally and wish that they be considered as Joint First Authors.

This work was carried out when all authors were with the Department of Computer Science & Engineering, Delhi Technological University, New Delhi INDIA

by our models so that they are more useful for educational technologists and educators, we extensively examine critical features responsible for their predictive potential.

This paper extends our prior work [Jindal et al., 2020] in several directions. First, we examine the predictive utility of a wide range of time series characterization techniques as inputs to our Random Forest models. Second, we analyze socio-demographic features responsible for differences in the amount of time children spend listening to their partners. Third, we generalize the Permutation Feature Importance algorithm to examine features relevant to our ResNet models. Fourth, we view our rich multimodal feature set in the context of features used by prior work through multiple feature ablation experiments. Fifth, we devise an algorithm to quantify the importance of time series features for our models.

The rest of the paper is organized as follows. The next section reviews existing work on backchanneling, disengagement and learning technology for children. Section III describes the dataset used, feature engineering and the LDP and BEP models. Section IV discuss our experiments and their results. Finally, Section V concludes our work with a discussion of our contributions and findings.

II. Relation to Prior Work

A. Peer tutoring and Robots as peer learning companions

Peer tutoring usually refers to two children working in a tutor-tutee dyad, where each individual has well-defined roles and protocols of interactions [Tymms et al., 2011]. Topping and Ehly [Topping and Ehly, 1998] had theoretically modelled the cognitive advantages of peer tutoring and learning. In the course of peer tutoring, a tutor manages activities such that they remain in the 'Zone of Proximal Development' through adequate support and scaffolding. This is a consequence of co-construction, wherein a peer tutor acts as a co-learner and minimizes their partner's frustration resulting from excess challenge. Since peer tutors do not hold a position of authority, tutees may also develop a trusting relationship with them and voluntarily disclose their ignorance or misconception. Many studies have reported encouraging outcomes of peer tutoring young students and those from low-income households and urban areas [Tymms et al., 2011].

Interactive tablets and social robots have shown great promise in educating young children worldwide and have the potential to become excellent peer tutors. While a majority of learning technologies are developed for children of WEIRD (Western, Educated, Industrialized, Rich and Democratic) countries, a recent large-scale intervention study showed that carefully designed educational software can provide quality basic education to all children regardless of their cultures and socio-economic status [McReynolds et al., 2020]. In the context of early language education, learning agents such as Project LISTEN's Reading tutor have been shown to improve children's reading comprehension significantly [Aist and Mostow, 1997]. An intervention study compared the Reading Tutor with regular classroom activities and found that children using the tutor outgained their classmates in comprehension [Mostow et al., 2003]. Another surprising study [Aist et al., 2001] found that students tutored by certified teachers did no better than children using a shared Reading tutor in tasks like word identification, word & passage comprehension and fluency. Among other factors, learning technologies dramatically improve learning gains by steering children towards their 'Zone of Proximal Development'.

However, for a learning agent to be a successful peer learning companion, it must be able to leverage social cues and our means of communication (facial expressions, gestures, verbal expressions etc.) to interact with us more 'naturally'. A recent study [Park et al., 2017b] had shown that children not only direct their attention towards social robots and find them more attentive, but are also distracted by non-contingent robots. To this end, in this paper, we develop models to enable peer learning companions to be social and engaging.

B. Social and Engaging communication

To communicate effectively, both speakers and listeners must go beyond merely issuing utterances and hearing & understanding, respectively. They must not only actively coordinate on the content of the communication, but also its process. Both the speaker and listener must establish a mutual understanding of the subject of discussion (content coordination). Moreover, the speaker must speak only when the listener is paying attention and understanding what the speaker says. Likewise, the listener must clearly indicate to the speaker that s/he is doing just that (process coordination) [Clark et al., 1991]. Thus, effective communication is a continuous, dynamic and collective process of the first order.

Listeners communicate through subtle backchannel responses such as gaze locking, nodding, gestures and short verbal expressions (*'yeah'*, *'ohh'*, *'hmm'*). These responses serve several cognitive functions such as indicating the state of engagement, understanding, clarification of ideas and sentence completion. For learning agents to be perceived as better listeners, they must backchannel adequately and at appropriate times. Few studies have been able to predict backchanneling opportunities successfully using the speaker's vocal prosodic features such as the pitch and energy [Morency et al., 2010, Noguchi and Den, 1998, Ward and Tsukahara, 2000]. In their seminal work, Ward and Tsukahara [Ward and Tsukahara, 2000] proposed a rule-based model which associated backchanneling responses with a region of low pitch lasting a few milliseconds. A decade later, Morency *et al.* [Morency et al., 2010] demonstrated how sequential probabilistic models such as Hidden Markov Models can automatically learn how to predict backchanneling opportunities from a database of adult-adult interactions. They also demonstrated that using multimodal features derived from the speaker (prosody, spoken words and eye gaze) resulted in a significant improvement over Ward and Tsukahara's unimodal model. This study not only opened up the exciting possibility of scaling up the development of backchanneling opportunity prediction models by automatically analysing human interactions, but also established the predictive utility of multimodal features. Poppe *et al.* [Poppe et al., 2010] also used the speaker's speech and gaze and multiple rule-based strategies to determine the timing of

backchanneling. A major drawback of all these models is that they were trained on adult interactions and voices, and thus unlikely to generalize to child-child interactions. Recently, Park *et al.* [Park et al., 2017a] developed a backchannel prediction model based on the non-verbal behaviour they observed in children. They identified both speaker cues (gaze, pitch, energy etc.) which listeners respond to and backchanneling responses (leaning towards, smiling etc.) which indicate engagement, and developed a backchanneling opportunity prediction model based on their findings.

Furthermore, to enable learning companions to be engaging speakers, they must be able to predict when their listeners are likely to be disengaged. Moreover, experiments conducted by Sidner *et al.* [Sidner et al., 2005] confirmed that human listeners were more likely to attend to robots with engagement gestures. They had used connection events like directed & mutual facial gaze, conversational adjacency pairs and backchannels to predict listener disengagement. In another study, Michalowski *et al.* [Michalowski et al., 2006] devised a system that used spatial information from a laser tracker along with head movements to predict engagement. However, the deployment of such a system is not practical, especially in underserved regions of the world. Instead, our work only makes use of a camera which is an easily available and cheap device. In [Bohus and Horvitz, 2009], Bohus *et al.* designed policy-based computational models for managing engagement in a multiparty environment. With multiple agents interacting with the system at a time, they made use of the head pose of the engaged agents, along with several other features to infer the focus of attention and the relationship between the agents to manage engagement. Our LDP model, however, is limited to dyadic interactions at present. Later, Park *et al.* [Park et al., 2017a] studied the impact of annotated visual and prosodic features in analysing listener disengagement. Lastly, Lee *et al.* [Lee et al., 2017] were amongst the first to show that both listener and speaker behaviours must be taken into account to best infer the attentive state of the listener. Drawing inspiration from their work, we use features from both the speaker and listener to predict the attentive state of the latter.

Our work is different from the existing literature in the following ways. First, we use a rich set of automatically extracted features (with the exception of a few hand-annotated features) across visual and vocal modalities. We incorporate features such as Facial Action Units (FAUs) [Sariyanidi et al., 2014], gross head and eye movements, pupil dilation and blink rate, which to the best of our knowledge, have never been used to predict the extent of backchanneling and listener disengagement. These features have, however, been used in other fields of study. For instance, FAUs [Sariyanidi et al., 2014] that detect phenotypical facial expressions, have been widely used in the affective computing literature. Similarly, studies such as [Goswami et al., 2020, D'Mello et al., 2012, Bhatia et al., 2019] have used gross body movements for a variety of tasks such as identifying emotions, cognitive states of children and predicting the severity of depression. The blink rate of the eye has been used to infer alertness [Haq and Hasan, 2016]. Moreover, many behavioural studies have shown that the pupil size is an important social signal which can influence social impressions [Kret, 2018].

Second, most of the existing approaches use a few time-series characterizations, such as the mean and standard deviation of pitch or gaze to predict backchanneling opportunities and disengagement. However, in this study we not only experiment with a wide range of time series characterizations such as approximate entropy, quantiles etc. (Section IV), but also utilize dynamic time series features to predict backchanneling and listener disengagement. In our earlier work [Jindal et al., 2020], we had hypothesized that dynamic time series features are a rich source of information which can help us better predict listener disengagement and backchanneling. In this paper, we devise an algorithm to quantify the importance of these time series features over their aggregates.

III. Methodology

A. Dataset

Datasets of social interactions play an essential role in gaining a deeper understanding of verbal and non-verbal human communication, and consequently in informing the design of social technologies. Multiple datasets exist that capture social interactions between adults [Malisz et al., 2016, de Kok and Heylen, 2011], and adults & children [Rehg et al., 2013, Nojavanasghari et al., 2016]. However, the social interactions of children have only been studied through the lens of adult–child interactions [Singh et al., 2018]. It is reasonable to believe that children interact with their peers differently from adults. Thus, while prior literature offers rich insights into adult–adult and adult–child social interactions, there is only a limited understanding of peer-to-peer interactions in children.

The P2PSTORY dataset is among the first attempts at capturing and analyzing peer-to-peer storytelling interactions in young children of five to six years of age. Like our prior work [Jindal et al., 2020], we use the P2PSTORY dataset to develop engaging and social agents to support early language learning in children. Below we present a brief description of the dataset and the data collection methodology. We urge the interested reader to refer to [Singh et al., 2018] for further details.

In order to collect data from children in a natural environment, the authors recruited eighteen children of a single kindergarten classroom from a public elementary school in Boston. The participants were 5.22 years old on an average and belonged to diverse socio-cultural and economic backgrounds. Most of the children were reported as having "typical development" with a few exceptions. The participants had an average Ages and Stages Questionnaire (ASQ) score of 30. ASQ is a standardized measure to evaluate a child's social and emotional development.

In the storytelling task, each child engaged in three rounds of storytelling with different partners and text-less storybooks. Each storybook had a sequence of coloured pictures with evolving scenes and characters that children could use to develop their own story. Each round differed in terms of the amount of instruction, number of scenes in the story and the type of partner. In a session, a dyad of children took turns to narrate a story to their partner, each turn resulting in a

storytelling episode. The dataset comprised of 29 sessions and 58 storytelling episodes. While each dyad session averaged 15 minutes and comprised of four stages ("Instruction", "Story Construction", "Story Sharing", and "Recall & Questionnaire"), the actual storytelling episode only lasted a little less than one and a half minutes on an average ($\mu = 77$ seconds).

Three cameras captured each storytelling episode, one facing the storyteller, the other facing the listener and the last one capturing a "bird's-eye view" of the dyad. The audio was also recorded using a high-quality microphone and synchronized in time with all the cameras. Multiple annotators coded each episode for a variety of verbal and non-verbal behaviours using ELAN [Brugman and Russel, 2004] (an annotation software). The following behaviours were annotated: gaze, posture, nod, eyebrow movement, mouth, utterances, voicing, on/off task, and attentive state. These behaviours were chosen either because they were found in prior work or were frequently observed in the storytelling sessions. Furthermore, the dataset also captured the participants' socio-demographic parameters such as the age, gender, household income, mother's qualification and results from the ASQ questionnaire etc. to analyze their influence on the acquisition of speaker cues and listener responses.

B. Feature Engineering

We extracted several useful features across the visual and vocal modalities from the high-quality audio and video recordings provided in the P2PSTORY dataset. We used OpenFace [Baltrušaitis et al., 2016] at a frequency of 30 Hz to render a set of highly detailed facial features shown in Table I. In addition to 18 Facial Action Units (FAUs), we extracted the velocity and acceleration of eye and head movements to better capture children's reactions in terms of their gross body movements. Moreover, we also calculated children's blink rate and pupil dilation using eye landmarks. The interested reader may refer our previous work [Jindal et al., 2020] for further details on these features.

Several studies [Jiang et al., 2011] have established that in addition to visual cues, audio features are also useful in increasing emotion recognition accuracy. Therefore, we used OpenSmile with a sampling frequency of 30 Hz, to extract some prosodic features listed in Table I. Specifically, we included the pitch (F0), first and second-order derivatives of Root Mean square energy of the fundamental frequency, and Mel-Frequency Cepstral Coefficients (MFCC), along with their derivatives. All derivatives were computed using Numpy's* gradient function which uses central differences in the interior and first differences at the boundaries with the default distance set to 1.

C. Listener Disengagement Prediction

The ability to detect when the listener is losing attention can be beneficial for a speaker. It can not only help in analyzing when and why the latter felt disengaged, but also allow the speaker to regain the listener's attention. The speaker may do so by commencing an engaging gesture [Sidner et al., 2005] or changing the subject of conversation. With this aim, we model the task of detecting listener disengagement as a time series classification problem.

Features	Description
FAUs	Indicate the presence of 18 Facial Action Units
gaze_vel	Velocity of eye gaze
head_vel_T	Translational velocity of head
head_vel_R	Rotational velocity of head
head_acc_T	Translational acceleration of head
head_acc_R	Rotational acceleration of head
blink_rate	First order differential of Eye Aspect Ration
pupil_dilation	Size of pupil (averaged for both the eyes)
F0	The fundamental frequency computed from the Cepstrum
mfcc′, mfcc″	First and Second order derivatives of Mel-Frequency Cepstral Coefficients 1-12
pcm_RMSenergy′, pcm_RMSenergy″	First and Second order derivatives of Root-mean-square signal frame energy

TABLE I
VISUAL AND VOCAL PROSODIC FEATURES USED IN THE STUDY. WHILE VISUAL FEATURES WERE EXTRACTED FOR BOTH THE SPEAKER AND LISTENER, PROSODIC FEATURES WERE ONLY EXTRACTED FOR THE SPEAKER. FOR A FEATURE X, (X)′ AND (X)″ ARE ITS FIRST AND SECOND ORDER DERIVATIVES.

Inspired by the recent study by Park *et al.* [Park et al., 2017b], we split all the storytelling episodes into disjoint 3 second windows. A 3 second window with 90 time steps or time-ordered observations†, is not only short enough to be considered stationary (and therefore easier to analyze) but also captures enough information to enable us to model its temporal dynamics. We used vocal prosodic features from the speaker, and visual & behavioural features from both the participants to model listener disengagement. We model LDP as a classification problem, by training our models to predict the listener's engagement state at the last (90^{th}) time step using the features of the first 89 time steps of the window.

Before formulating the LDP problem any further, we first had to resolve a mismatch between our feature sets, some of which were sampled at a different frequency. As discussed in Section III-B, the features extracted using OpenSmile and OpenFace were sampled at a frequency of 30 Hz. In contrast, annotated behavioural features from the P2PSTORY dataset were only available at a lower sampling frequency of 5 Hz. We resolved this mismatch by aggregating each annotated feature across all its observations sampled during the time window. Specifically, we introduced dummy variables for each value of a behavioural feature. Each dummy variable was associated with the ratio of time steps where a particular dummy variable was True to the total time steps in a window (89). For example, at any time step, a listener could either have *raised*, *furrow* or *neutral* eyebrows. Therefore, we added three dummy variables replacing the *eyebrows* feature: *eyebrows_raised*, *eyebrows_furrow* and *eyebrows_neutral*. Therefore, we had two categories of features for each 3 second window: visual & vocal multivariate time series features, and a set of aggregated annotated features.

*https://numpy.org/

†Features were sampled at a frequency of 30 Hz (Section III-B).

We could also have down-sampled the OpenFace and OpenSmile features, but that would have meant a lot of lost data. On the other hand, we could also have over-sampled behavioural features. However, we found that these features hardly showed any variation in the chosen 3 second interval, and hence, the dynamics of behavioural features would have had very little information. Therefore, instead of over-sampling, we transformed these behavioural annotations into window-level features by aggregation.

Finally, we provide a mathematical formulation for our LDP model. Let $\mathtt{W}_i$ be the $\mathtt{i}^{\mathtt{th}}$ 3 second window. Also let $\mathtt{T}_j$ be the visual & vocal multivariate time series features for the $\mathtt{j}^{\mathtt{th}}$ time step of $\mathtt{W}_i$. Therefore, the dynamic features for the first 89 steps in the window $\mathtt{W}_i$ can be represented as $[\mathtt{T}_1, \mathtt{T}_2, ..., \mathtt{T}_{89}]_i$. Furthermore, let $\mathcal{A}_i$ be the aggregated annotated features for $\mathtt{W}_i$. Then, for the speaker to predict the engagement state of the listener ($\mathcal{E}_i$) at the $90^{\mathtt{th}}$ time step of window $\mathtt{W}_i$, we need to model the following mapping $\mathcal{F}_{\mathcal{LDP}}$:

$$\mathcal{F}_{\mathcal{LDP}}([\mathtt{T}_1, \mathtt{T}_2, ..., \mathtt{T}_{89}]_i, \mathcal{A}_i) \mapsto \mathcal{E}_i \tag{1}$$

For our experiments, we focused on two widely used algorithms for time series classification: Random Forests and ResNet. We used different sets of time series characteristics (`mean`, `basic`, `tsfresh` etc.) concatenated with the aggregated annotated features as input to our Random Forest models. ResNet is a state-of-the-art deep learning model for time series classification [Fawaz et al., 2019]. We used ResNet to learn the dynamics of the vocal and visual time series features. We also made use of the aggregated annotated features, by combining them with the latent representation of the temporal features learned by ResNet, and then passing them all to the final softmax layer for prediction.

D. Backchanneling Extent Prediction

Backchannel responses act as feedback to speakers, which enables them to analyze the extent to which the listeners understand them [Dennis and Kinney, 1998]. These responses are pivotal to communicating effectively, and therefore an active artificial listener must be able to backchannel appropriately. While predicting the exact timing of backchanneling signals is a challenging task, many studies in the past have proposed models to predict backchanneling opportunities. In fact, recently Park *et al.* [Park et al., 2017b] proposed a backchanneling opportunity prediction model based on child-child interactions. In this work, however, we investigate speaker cues which elicit *high* and *low* levels of backchanneling responses from listeners, instead of predicting their timing.

Like LDP, we model the task of predicting backchanneling extent as a time series classification problem using 3 second windows. However, we only make use of speaker cues (visual and prosodic features) to predict the extent of backchanneling. Since most of the annotated prosodic features (pauses between phrases, speaker's energy, etc.) available in the P2PSTORY dataset had limited variation during episodes, we did not include these features in our analysis, unlike prior work [Park et al., 2017b]. Based on the speaker cues for a particular window $\mathtt{W}_i$, the BEP model learns to predict the extent of listener's backchanneling response in the next window *i.e.,* $\mathtt{W}_{i+1}$. Taking inspiration from Dennis *et al.* [Dennis and Kinney, 1998], we considered both verbal (listener utterances such as *"hmm"*, *"ooh"*, etc.) and non-verbal (smile, nod, onset of partner gaze, lean towards, raising brows) gestures as backchannels for our BEP model. To label a window as having '*high*' or '*low*' backchanneling, we first find the proportion p of time steps wherein the listener generates one or more of the aforementioned backchannel responses. Next, we introduce a backchanneling threshold parameter τ to convert p to binary labels, *high backchanneling* or *low backchanneling* as follows:

$$label_{BC} = \begin{cases} \textit{high backchanneling} & p > \tau \\ \textit{low backchanneling} & p \leq \tau \end{cases} \tag{2}$$

The threshold parameter τ regulates the minimum number of time steps with backchanneling responses (($\tau \times 90$ time steps)) required for the window to qualify as being indicative of *high backchanneling*. For our experiments, we use two values of τ: 0.25 and 0.50. A higher τ includes responses that are relatively more prolonged, and on the other hand, a lower τ includes both subtle and long-lasting backchannels.

Now, we mathematically formulate the BEP problem. Let $\mathtt{W}_i$ be the $\mathtt{i}^{\mathtt{th}}$ 3 second window, and $[\mathtt{T}_1, \mathtt{T}_2, ..., \mathtt{T}_{90}]_i$ be the visual & vocal multi variate time series features for $\mathtt{W}_i$. Further, let $\mathcal{B}_{i+1} \in \{\textit{high backchanneling}, \textit{low backchanneling}\}$ be the target variable indicating the extent of listener backchannels in the following window ($\mathtt{W}_{i+1}$). Then, to predict the extent of backchanneling, our must learn the following mapping $\mathcal{F}_{\mathcal{BEP}}$:

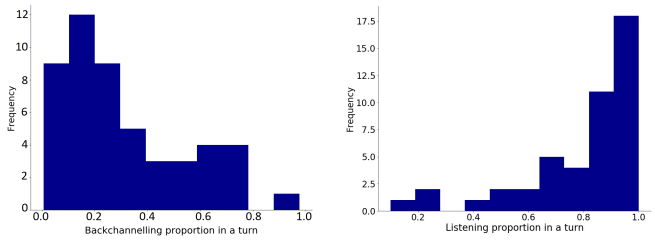
$$\mathcal{F}_{\mathcal{BEP}}([\mathtt{T}_1, \mathtt{T}_2, ..., \mathtt{T}_{90}]_i) \mapsto \mathcal{B}_{i+1} \tag{3}$$

Like LDP, we use both Random Forests and ResNet to predict the extent of backchannel responses.

E. Influence of Socio-demographic factors

Children gradually learn the art of weaving meaningful stories through "culturally-laden" interactions with their parents, peers and teachers. Studies across diverse cultural, socio-demographic and economic contexts have indicated that these factors may influence the development of narrative skills in young children [Curenton, 2010]. Like narrative skills, the acquisition of listener responses may also exhibit individual differences [Dittmann, 1972]. In our earlier work [Jindal et al., 2020], we had examined which socio-demographic and economic features influence the extent of backchanneling. The extent of backchanneling was expressed in terms of *backchanneling proportion* per storytelling episode *i.e.* the fraction of time (time steps) the listener exhibited any form of backchanneling in an episode. Figure 1 *(i)* illustrates the distribution of backchanneling proportion across all the storytelling episodes. For the sake of brevity, we shall not delve any further on the influence of socio-demographic factors on the extent of backchanneling and refer the interested reader to [Jindal et al., 2020] for more details.

In this paper, we build on our prior work [Jindal et al., 2020] and analyze whether socio-demographic and economic factors influence how much time children spend listening to



(i) (ii)

Fig. 1. Distribution of (i) *backchanneling* and (ii) *listener proportion* across all storytelling episodes.

their peers. We quantify the extent of listening in terms of the *listening proportion* or the fraction of time the listener actively listens to the storyteller. Figure 1 *(ii)* shows the distribution of listening proportion across all storytelling episodes.

In order to determine the socio-demographic features influencing the amount of time children spend listening, we conducted statistical tests to discover socio-demographic and economic features causing significant differences in the distribution of listening proportion across all the story-sharing episodes. In particular, for each socio-demographic or economic factor noted in Table II, we carried out a two-sample Kolmogorov-Smirnov (K-S) test under the null hypothesis that the distribution of listening proportion conditioned on its (the socio-demographic factor) values is similar. A two-sample K-S test examines whether the empirical cumulative distribution (ECDF) of a random variable significantly differs across the two samples. For example, consider the gender of the storyteller. The storyteller can either be a `male` or a `female`. Thus, the K-S test either rejects or accepts the null hypothesis that the distribution of listening proportion when the storyteller is male is not significantly different from its distribution when the storyteller is a female. As shown in Table II, the K-S test on the gender of the storyteller rejects the null hypothesis.

A notable exclusion from Table II is the age of the children. This is because all the children were approximately of the same age *i.e.* between five to six years old [Singh et al., 2018]. It must be noted that the K-S test only measures the difference between the empirical distributions of the two samples. Thus, socio-demographic factors having more than two values had to be 'treated'. We binned the values of such factors into '`High`' and '`Low`' buckets or chose the two most frequently occurring values for the K-S test. For example, the participants' ASQ scores were classified as '`High Score`' or '`Low Score`' based on multiple thresholds (Thresh $=$ 20 or 30), such that children having ASQ scores less than the threshold belonged to the '`Low Score`' category and vice versa. Likewise, the participants' total household income was also classified into '`High Income`' (Over $\$100,000$) and '`Low Income`' ($\$30,000$ to $\$75,000$) groups. Furthermore, participants' mother's highest education also had multiple values: `Graduate or professional training`, `College Graduate`, `High school graduate/ GED`, `Some college or vocational school` and `Some high school`, but since most mothers were college graduates, we used the two most frequently-occurring qualifications (`Graduate or professional training` or `College Graduate`) as our two samples. We also carried out randomization tests ($n = 1000$) by randomly permuting feature values and computing the K-S statistic for each randomly permuted feature set. We discuss our findings at a greater depth in Section IV-A.

F. Interpretability: Feature and Time Series Importance

Given the nature of our study, interpretability of our machine learning models is as important as their predictive accuracy. In this section, we provide a brief overview of some techniques which give us insights into features that most influence our models' predictions. While there are many ways to rank essential features, in this study, we primarily use Mean Decrease in Impurity (MDI) [Breiman, 2001] and Mean Decrease in Accuracy (MDA) [Breiman, 2002] also known as permutation feature importance.

MDI evaluates the importance of a feature by measuring the total decrease in impurity as a consequence of splitting on it, weighted by the proportion of samples reaching the node and averaged over all the trees. The importance of a feature is directly proportional to its MDI score. However, a significant drawback of MDI is that it is biased towards high-cardinality features *i.e.* categorical features having many values. To supplement our analysis and prevent misleading conclusions, we also calculated feature relevance using MDA or permutation feature importance. The elegant yet straightforward assumption behind the permutation feature importance algorithm is that randomly shuffling the observations of an irrelevant feature in the testing dataset should only have a negligible impact on the performance of a well-trained model. On the other hand, shuffling the observations of an important feature should confuse the model and result in a significant drop in performance (`accuracy`). Specifically, the MDA for a feature f is the total decrease in a model's `accuracy` as a result of randomly shuffling its observations, averaged over s independent runs of the algorithm. In our work, in addition to accuracy, we also estimated the importance of features using the model's `AUC` and `log-loss`[‡]. It must be noted that MDA is computed with the help of a trained model on the testing data.

The vanilla MDA algorithm [Breiman, 2002] expects testing data to be a two-dimensional matrix having dimensions $\mathcal{O} \times \mathcal{F}$, where $|\mathcal{O}|$ is the number of observations whereas $|\mathcal{F}|$ is the number of features. However, ResNet is trained and tested on a three-dimensional matrix, where each observation is a time series instead of a scalar. To discover important features for our ResNet models, we propose Algorithm 1, which adjusts the shuffling procedure of the vanilla MDA algorithm for three-dimensional datasets and Convolutional Neural Networks. Algorithm 1 takes a trained model $\mathbb{M}$, a three-dimensional feature matrix $\mathbb{X}$, observed values of the target $\mathbb{Y}$ and a scoring metric $\mathbb{S}$ as input, and returns the *importance score* of each feature. The feature matrix $\mathbb{X}$ is of the order $\mathcal{W} \times \mathcal{T} \times \mathcal{F}$ where $|\mathcal{W}|$ is

[‡]Accuracy, AUC and log-loss are referred to as scoring metrics. The log-loss metric was only used for our ResNet models.

Algorithm 1: Permutation Feature Importance

Input: Model ($\mathbb{M}$), Feature Matrix ($\mathbb{X}$), Targets ($\mathbb{Y}$), Scoring metric ($\mathbb{S}$)
Output: *Importance scores* = $\{(f, i)\ \forall f \in$ feature set, i is the corresponding feature importance$\}$

1 **begin**
2 $\quad \mathcal{S}_b \leftarrow$ *Scorer*($\mathbb{M}$, $\mathbb{X}$, $\mathbb{Y}$, $\mathbb{S}$) // `Baseline score`
3 $\quad$ Initialize *Importance scores* $\leftarrow \phi$
4 $\quad$ Initialize $\mathcal{S}_p \leftarrow 0$
5 $\quad$ **for** $f \leftarrow 1$ **to** $|\mathcal{F}|$ **do**
6 $\quad\quad \mathbb{X}_p \leftarrow$ *Shuffle*($\mathbb{X}$, f) // `shuffle feature` f
7 $\quad\quad \mathcal{S}_p \leftarrow$ *Scorer*($\mathbb{M}$, $\mathbb{X}_p$, $\mathbb{Y}$, $\mathbb{S}$)
8 $\quad\quad$ *score difference* $\leftarrow \mathcal{S}_b - \mathcal{S}_p$
9 $\quad\quad$ *Importance scores* $\leftarrow$ *Importance scores* $\cup$ (*score difference*, f)
10 $\quad$ **end**
11 $\quad$ **return** *Importance scores*
12 **end**

Algorithm 2: Time-Series Importance

Input: Model ($\mathbb{M}$), Feature Matrix ($\mathbb{X}$), Targets ($\mathbb{Y}$), Scoring metric ($\mathbb{S}$), Feature list (*feature_list*)
Output: Importance score (*Importance score*)

begin
$\quad \mathcal{S}_b \leftarrow$ *Scorer*($\mathbb{M}, \mathbb{X}, \mathbb{Y}, \mathbb{S}$)
$\quad$ Initialize $\mathcal{S}_p \leftarrow 0$
$\quad$ Initialize $\mathbb{X}_p \leftarrow \mathbb{X}$
$\quad$ **for** $f \in$ *feature_list* **do**
$\quad\quad \mathbb{X}_p \leftarrow$ *Shuffle*($\mathbb{X}_p, f$)
$\quad$ **end**
$\quad \mathbb{M}' \leftarrow$ *Train*($\mathbb{X}_p, \mathbb{Y}$)
$\quad \mathcal{S}_p \leftarrow$ *Scorer*($\mathbb{M}', \mathbb{X}_p, \mathbb{Y}, \mathbb{S}$)
$\quad$ *Importance score* $\leftarrow \mathcal{S}_b - \mathcal{S}_p$
$\quad$ **return** *Importance score*
end

the number of windows (observations), $|\mathcal{T}|$ is number of time steps in a window and $|\mathcal{F}|$ is the total number of features. The scoring metric $\mathbb{S}$ can be any performance metric such as `AUC`, `accuracy` etc. Algorithm 1 first evaluates the performance of the trained model $\mathbb{M}$ using the scoring metric $\mathbb{S}$ (baseline score, $\mathcal{S}_b$) (Line 2). Next, it permutes the feature matrix $\mathbb{X}$ (Line 6) using the `shuffle` procedure and re-evaluates the model on the same scoring metric resulting in the permuted score $\mathcal{S}_p$ (Line 7). The *importance score* of a feature is then the difference between the baseline and permuted score (Lines 8 & 9).

The `shuffle` procedure randomly permutes the time series observations of a given feature based on a randomly generated permutation map. A permutation map $\mathbf{M}$ is simply a $|\mathcal{W}|$-dimensional vector which defines a mapping $\mathbf{M} : \{(1,\ 2,\ ...\ |\mathcal{W}|)\} \rightarrow \mathbb{P}$, where $\mathbb{P}$ denotes the permutation set of $(1,\ 2,\ ...\ |\mathcal{W}|)$, such that if $\mathbf{M}[i] = j,\ (i \neq j)$ then the algorithm installs the time series observation from the i^{th} window of $\mathbb{X}$ to j^{th} window of $\mathbb{X}_p$ for a given feature f. We must note that the time series observations of only one feature are shuffled at a time while the temporal-order of each time series remains the same in both $\mathbb{X}$ and $\mathbb{X}_p$.

We also used Partial Dependency Plots (PDPs) to investigate how some important features influence our target variables. PDPs illustrate the functional relationship between a subset of features and the target variable when the influence of all the remaining features are averaged-out. They are useful in determining whether the relationship between a feature and the target is linear, monotonic or complex [Friedman, 2001].

Finally, to demonstrate whether the dynamics of time series features bear any importance to the predicting capability of our ResNet models, we devised an Algorithm 2 on similar lines as Algorithm 1. The fundamental assumption behind Algorithm 2 is that if the dynamics of time series features significantly influence the predictive performance of ResNet, then destroying the temporal information (by randomly shuffling $\mathbb{X}$ along the time dimension, $\mathcal{T}$) must result in a sharp drop in performance. The elegance of Algorithm 2 however, lies in the invariant that most characteristics of the shuffled time series are unaffected by the permutation. This is because time series characteristics (Section IV) like `mean` and `standard deviation` are independent of the temporal ordering of values in a time series. This invariant entails that ResNet models 'trained' on the permuted data learn from information no more than any Random Forest model.

Algorithm 2 differs from Algorithm 1 in the following ways. Whereas in Algorithm 1, a trained model is evaluated on $\mathbb{X}_p$ in the testing phase, in Algorithm 2, we train a new ResNet model and evaluate it on every permuted feature matrix. This way, we force the ResNet model to depend on information other than the dynamics of time series features. Secondly, we only permute the top $\eta = 10$ most important features found using Algorithm 1 to allow the model to have some temporal information. Finally, the `shuffle` procedures of Algorithms 2 and 1 are slightly different. In Algorithm 2, the top η features are permuted along the time dimension ($\mathcal{T}$) to destroy temporal information unlike Algorithm 1 where features are permuted along the $\mathcal{W}$ dimension and the temporal information of each time series observation remains intact.

IV. Results

We begin our analysis by investigating the influence of socio-demographic and economic features on active listening and backchanneling in Section IV-A. This is followed by an extensive examination of the listener disengagement (LDP) and backchanneling extent (BEP) prediction models in Sections IV-B and IV-C, respectively. Finally, we demonstrate the importance of time series features using Algorithm 2 in Section IV-D.

Sections IV-B and IV-C are structured as follows. We begin by assessing the performance of our models trained on different sets of features inspired by prior work. These experiments not only give us an insight into features integral for the LDP and BEP tasks but also allows us to view our work and feature set, in the context of listener disengagement and backchanneling literature. Next, we experiment with different characteristics of time series as features for Random Forests. In order to use a set of time series $\mathcal{D} = \{T_i\}_{i=1}^{n}$ as features for supervised learning algorithms such as Random Forests, one must first map $\mathcal{D}$ into a feature vector $\mathcal{X} = \{x_1, x_2, ...x_m\}$, where each feature x_j is a scalar value and a characteristic of

Features	Backchanneling		Listening	
	Storyteller	Listener	Storyteller	Listener
Gender	**0.40** *	0.20	0.17	0.22
Participants have same gender	0.13	0.13	**0.32** *	**0.32** *
Mothers' Highest education	0.21	**0.57** *	**0.29** *	0.18
ASQ Score (Thresh = 20)	–	**0.32** *	0.20	0.15
ASQ Score (Thresh = 30)	**0.27** *	**0.28** *	0.27	0.17
Uses words to describe feelings	**0.34** *	**0.32** *	**0.50** *	**0.53** *
Friendly with strangers	0.15	**0.28** *	0.17	0.14
Talks with adults (s)he knows well	0.22	0.20	0.18	0.19
Looks at parent when talking	0.27	**0.59** *	0.25	0.17
Has siblings	0.24	0.23	**0.34** *	0.28
Total household income	0.18	**0.27** *	**0.46** *	0.21

TABLE II
K-S STATISTICS: * INDICATE SIGNIFICANT DIFFERENCES AT 5% SIGNIFICANCE LEVELS.

Metric	Backchanneling		Listening	
	Low	High	Low	High
P	0.66	0.73	0.60	0.55
R	0.75	0.72	0.60	0.58
F1	0.66	0.69	0.57	0.53
AUC	0.84		0.74	

TABLE III
EXTENT OF BACKCHANNELING AND LISTENING USING RANDOM FORESTS WITH SOCIO-DEMOGRAPHIC FEATURES.

Importance socio-demographic features	
Backchanneling	**Listening**
Mother's highest education (L)	Gender (S)
Gender (S)	Mother's highest education (S)
ASQ Score (L)	Household income (S)
Gender (L)	Gender (L)
Household income (L)	Mother's highest education (L)

TABLE IV
MOST IMPORTANT FEATURES THAT PREDICT BACKCHANNELING AND LISTENING ((L): LISTENER, (S): SPEAKER).

some time series $T_i \in \mathcal{D}$. Time series can be effectively and efficiently characterised based on the distribution of their values, correlation properties, stationarity, entropy etc. [Fulcher, 2018, Christ et al., 2018]. In our prior work [Jindal et al., 2020], we had used the arithmetic means of all time series features at the window-level to predict listener disengagement and the extent of backchanneling using Random forests. However, in the present work, we experimented with various different characteristics of time series. In particular, we used two different sets of time series characteristics (or aggregates) for our prediction tasks: `Basic` and `Tsfresh`. The `Basic` set included the arithmetic mean (`M`), standard deviation (`SD`), minimum (`Min`), maximum (`Max`), median (`Md`), first (Q_1) and the third quartile (Q_3) of each time series feature at the window-level.

The `Tsfresh` set on the other hand, included a 'relevant subset' of 63 different time series characteristics such as absolute entropy, kurtosis, skewness etc.[§] and a total of 794 characterizations[¶] (or aggregates) for each time series feature at the window-level. Each aggregate is a time series characteristic with a specific configuration of its parameters. For instance, `approximate_entropy` is a time series characteristic, while `approximate_entropy(m, r)` is an aggregate for an integer `m` (length of compared run of data) and a positive float `r` (filtering level). Since too many irrelevant features may impair the quality of models, `tsfresh` also selects relevant features from the exhaustive set of aggregates it creates based on statistical hypothesis tests [Christ et al., 2018]. The final set of 'relevant' `tsfresh` features included 496 features for the LDP task and only 6 features for the BEP task in total.

Finally, we conclude each section by examining and comparing important features for both our Random Forest and ResNet models. In our earlier work [Jindal et al., 2020], we had only reported important features for the Random Forest models. However, in this paper, we also analyze features which contribute most to the predictive capability of our ResNet models using Algorithm 1.

[§]The interested reader may refer to https://tsfresh.readthedocs.io/en/latest/text/list_of_features.html for a complete list of time series characterization techniques.

[¶]`tsfresh` computes 794 features by default.

In our previous work [Jindal et al., 2020] we had examined the performance of LDP and BEP models across three experimental settings inspired by the hierarchical nature of our data. We had categorised our experiments as "Leave-One Subject Out" or LOSO, "Leave-One Episode Out" or LOEO, and "Random stratified". In the LOSO experiment for instance, a model (LDP or BEP) is trained on data from multiple subjects while it is tested on the left-out subject in each fold. We had observed an expected degradation of our models' performance from random stratified to the LOSO setting due to a reduction in the amount of information sharing between the training and testing sets. Since the LOSO experiment predicts on data from an unseen child, it is the most practically useful experimental setting. Performance in the LOSO experiments represent a lower bound and have been widely used in different studies due to their insights into the practical utility of models [Goswami et al., 2020, Agarwal et al., 2020]. In the following sections, we only report results under the LOSO experimental setting unless stated otherwise. We had previously discussed the impact of the experimental setting in great detail, and we urge the interested reader to refer to [Jindal et al., 2020] for more information.

A. Socio-Demographic Analysis

Table II presents the results of the Kolmogorov-Smirnov (K-S) and randomization tests on all socio-demographic factors. Those factors which led to a significant difference in backchanneling and listening proportion are highlighted and indicated with a *. It can be seen that factors such as how often listeners look at their parents when talking ($\mathcal{D} = 0.591$) and their mother's education ($\mathcal{D} = 0.569$) gave rise to the highest difference in the distributions of backchanneling proportion among all the examined socio-demographic factors. On the other hand, factors such as whether children use words to describe their feelings (Storyteller: $\mathcal{D} = 0.5$, Listener: $\mathcal{D} = 0.527$) and the storyteller's household income ($\mathcal{D} = 0.458$) influence the distribution of listening proportion the most.

To examine the direction of the difference *i.e.* to discover which factors cause an increase in the extent of time children spend listening and backchanneling, we plotted the Empirical Cumulative Distributions (ECDFs) of all socio-demographic factors which yielded significant differences. Figure 2 illustrate

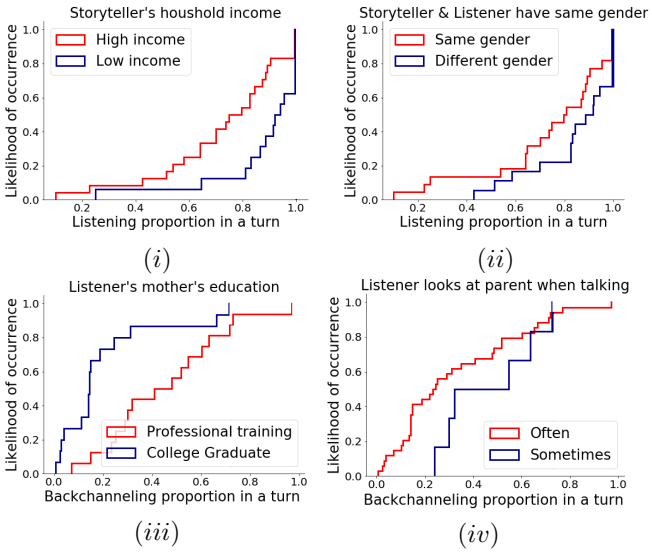


Fig. 2. Empirical CDFs of *listening proportion* corresponding to the (i) storyteller's total household income and (ii) whether both the storyteller and listener have the same gender. ECDFs of *backchanneling proportion* corresponding to the (iii) listeners' mothers' education and (iv) whether they look at their parents when talking to them.

ECDFs of some important socio-demographic factors which influence listening and backchanneling, respectively. Given an empirical cumulative distribution and a backchanneling proportion ($\mathcal{P}_{bc}$), listeners are much more likely to backchannel (frequently) if they `often` look at their parents while talking and their mothers have undergone `graduate or professional training`. Similarly, given a listening proportion ($\mathcal{P}_l$), listeners actively listen more often if they are of the same gender as their partners or belong to a high-income household.

Some other socio-demographic factors which influence the distribution of backchanneling proportion are the storyteller's gender and the listener's household income. The ASQ scores of the storyteller and the listener also yield significant differences. Similarly, responses to questions in the ages and stages questionnaire (ASQ), such as how often does the child use words to describe feelings and whether the child is friendly with strangers also impact the backchanneling proportion.

Correspondingly, aspects of the storytellers' socio-demography such as their household income, mother's education, whether they have siblings and how often they describe their feelings affect listening proportion. It is interesting to note that whether the storyteller and listener are of the same gender influences listening, but it has no bearing on the extent of backchanneling. In contrast, while ASQ scores and responses influence the backchanneling proportion, they result in insignificant differences in the listening proportion. Another exciting yet inexplicable finding is that listeners' socio-demography affects backchanneling more often in comparison to the storyteller's factors which influence the listening proportion more frequently.

Since we found that socio-demographic features indeed influence backchanneling and listening proportion, we trained Random Forest classifiers on these features to predict the extent of backchanneling and listening (‘`High`’ or ‘`Low`’) in a storytelling episode. These experiments allowed us to which discover socio-demographic and economic features best predict the amount of time children spent backchanneling and listening in an episode. To train the Random Forest models, we used all socio-demographic features mentioned in Table II. We labelled episodes with a BC proportion greater than median BC proportion ($\tau_{bc} = 0.23$) as having ‘`High backchanneling`’. Rest of the episodes were labelled as ‘`Low backchanneling`’. Similarly for the listening experiment, we labelled episodes having listening proportion greater than its median ($\tau_l = 0.86$) as ‘`High listening`’, and the rest as ‘`Low listening`’.

The results of these experiments are summarized in Tables III and IV. Table III summarizes the results of these experiments averaged over 5 folds of random stratified cross-validation. These results suggest that the Random Forests models can predict the extent of backchanneling and listening reasonably well, despite having access to a limited set of social and demographic features only.

Furthermore, Table IV lists important features which contribute most to the predictive capabilities of the Random Forest models, found using permutation feature importance (or Mean Decrease in Accuracy, Section III-F). It can be seen that the genders of both the partners in a dyad and the listener's mother's education are the most critical socio-demographic features which predict both, the extent of backchanneling and listening. This may imply that children behave differently depending on the gender of their partners. Moreover, the impact of mother's education on the child is natural, given that mothers have a profound influence on a child's overall development [Tamis-LeMonda and Rodriguez, 2008]. Studies have shown that educated mothers not only read to their children more often but often demonstrate sophisticated language skills which improve the quality of their child's verbal interactions [Scarborough and Dobrich, 1994, Raikes et al., 2006, Rowe et al., 2005]. Another important feature that influences backchanneling and listening is the household income of a child. This finding too follows from prior work which has shown that household income (and poverty) is strongly associated with less stimulating home environments which in turn adversely affects the cognitive and verbal abilities of children [Tamis-LeMonda and Rodriguez, 2008, Garrett et al., 1994].

B. Predicting Listener Disengagement

Table V summarizes results obtained for the Listener Disengagement Prediction task by our Random Forest and ResNet models under different experimental settings. Since we are more interested in accurately predicting disengagement, we only report performance metrics for the not-listening (negative) class. Below we discuss some key findings.

First, we discuss our observations from the feature ablation experiments. We found that adding prosodic features to the annotated ones (gaze, posture, nods etc.) significantly increases the `precision` of Random Forests [`Model 2`] while marginally decreasing its `F1` and `recall`. For the

M.No.	Reference	Features				Aggregates	P	R	F1	AUC
		Annotated	Prosodic	Visual	Demographic		Random Forests			
1	Park *et al.*, 2017 [Park et al., 2017a]	✓				Mean	0.80	0.57	0.65	0.89
2	Lee *et al.*, 2017 [Lee et al., 2017]	✓	✓			Mean	0.96	0.51	0.64	0.92
3		✓	✓	✓		Mean	**0.63**	**0.78**	**0.67**	**0.85**
4		✓	✓	✓		Mean & Stdev	0.96	0.35	0.49	0.93
5		✓	✓	✓		Basic	0.66	0.23	0.33	0.93
6		✓	✓	✓		Tsfresh	0.56	0.15	0.23	0.89
7		✓	✓	✓	✓	Mean	0.92	0.46	0.60	0.94
							ResNet			
8	Park *et al.*, 2017 [Park et al., 2017a]	✓				–	0.75	0.58	0.64	0.77
9	Lee *et al.*, 2017 [Lee et al., 2017]	✓	✓			–	0.35	0.92	0.35	0.64
10		✓	✓	✓		–	**0.67**	**0.79**	**0.70**	**0.89**

TABLE V
LISTENER DISENGAGEMENT PREDICTION: WE REPORT PRECISION (P), RECALL (R), F1-SCORE (F1) AND AUC FOR THE "NOT-LISTENING" CLASS.

ResNet however, the combination results in sharp drop in the F1 and precision, and a significant increase in recall. Therefore, while adding prosodic features did not result in a significant boost performance, their predictive utility cannot be ruled out. In fact, prosodic features such as F0 (Pitch), mfcc[1], mfcc[7] and mfcc[9] are amongst the most important features for ResNet [Figure 3 *(iii)*]. Moreover, including visual features along with the annotated and prosodic features, results in the best model for both Random Forests and ResNet [Model 2 & 9]. This stems from the fact that cognitive states such as engagement often manifest in facial expressions [Goswami et al., 2020]. Finally, the addition of socio-demographic features led to mixed results. While our Random Forest model reported an increase in precision and AUC, there was also a sharp drop in recall [Model 3 vs 7].

Next, we examine the impact of different sets of time series characteristics (or aggregates) on the performance of Random Forests [Models 3 to 6]. Given the same set of features (Annotated, Prosodic & Visual), we found that the arithmetic mean of time series outperforms other sets of aggregates by a substantial margin. Nevertheless, mean and standard deviation resulted in the most precise model (precision $= 0.96$). The reason behind the unsatisfactory performance of the basic and tsfresh feature sets could be that adding numerous highly correlated or irrelevant features result in poor quality models. Besides, tsfresh independently tests each time series characterization for its relevance [Christ et al., 2018]. However, some features may be excellent predictors of the target under investigation in combination with other features. Therefore, while univariate significance tests may be useful in limiting irrelevant features, they may also get rid of individually useless and collective useful features.

In conclusion, while Random Forests trained on different sets of features and time series characteristics gives promising results, ResNet with all the features (Model 10) yields the best listener disengagement prediction model (highest F1 and AUC). This is primarily because ResNet is able to capture the intricacies and dynamics of listener behaviour by learning from time series features.

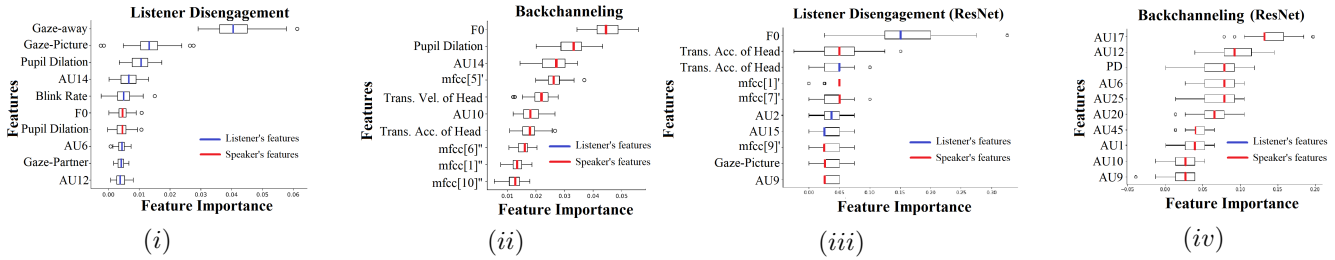
Figure 3 (i) and (iii) show the most important features for the best performing Fandom Forest [Model 3] and ResNet [Model 10]. We used the vanilla MDA algorithm [Breiman, 2001] for Random Forests, and Algorithm 1 for ResNet to compute feature importance across 50 independent runs. The annotated features, Gaze-Away and Gaze-Picture were the most important features for our Random Forest model. This is because listeners gazing-away may be perceived as being distracted [Lee et al., 2017]. The 3-D PDP [Figure 4 (ii)] also supports this conclusion. The PDP plot also reveals that children looking towards the picture are regarded as being attentive. We also found that pupil dilation and blink rate of listeners are important indicators of disengagement. To the best of our knowledge, these features have not been used in predicting disengagement before. But not all annotated features were useful. Features such as laugh, eyebrows, posture and nod were in fact the least important.

For our ResNet model, prosodic features of the speaker (F0 (Pitch), mfcc[1]' and mfcc[7]') turned out to be the most important. This is intuitive because hearing a monotonous speech may be boring. On the other hand, storytellers who modulate their voice may generate excitement and engagement. FAUs (AU02 (*outer brow raiser*) and AU15 (*lip corner depressor*)), and the acceleration of head were some other features which our ResNet model relied on to make predictions. Again, these features have never been used in prior work investigating disengagement and analysing social interactions.

C. Predicting the Extent of Backchanneling

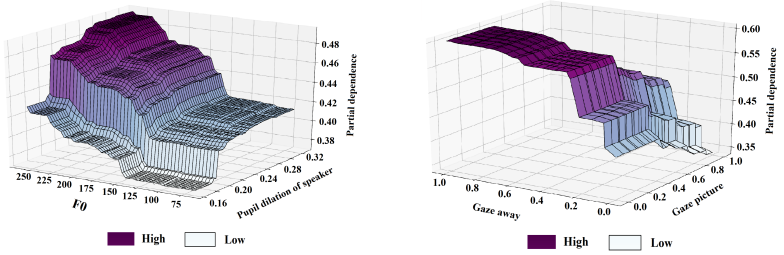
Table VI summarizes the results of our BEP experiments. We record the models' performance in terms of various performance metrics obtained for the *high backchanneling* (positive) class, averaged across all LOSO test splits. In the next few paragraphs, we discuss our key observations from these results.

First, we compare the performance of our Random Forest models for different values of the backchanneling threshold parameter *i.e.* for $\tau = 0.25$ and 0.50 [Models 1 to 8]. We observed an interesting trend in their precision and recall values. Specifically, most models (6 out of 8) trained for $\tau = 0.25$, had lower precision and higher recall in comparison to corresponding models trained for $\tau = 0.50$ *i.e.* $P_{\tau=0.25} < P_{\tau=0.50}$, and $R_{\tau=0.25} > R_{\tau=0.50}$. Thus, when τ drops from 0.50 to 0.25, there is an increase in the number of true positives, but at the cost of an increase in false positives. We can explain this trend based on the average backchanneling proportion of each subject (Figure 5) and our models' expectations. From Figure 5 it can be clearly seen that



(i) (ii) (iii) (iv)

Fig. 3. Permutation Feature Importance (MDA): Top 10 most important features for (i, ii) Random Forests and (iii, iv) ResNet.



(i) (ii)

Fig. 4. Partial Dependence Plots: (i) Combined effect of speaker's F0 and Pupil Dilation on backchanneling signals (ii) Impact of the proportion of time when the listener is gazing towards the picture (Gaze-Picture) and away (Gaze-Away) on listener's disengagement.

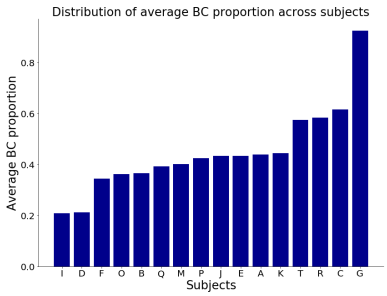


Fig. 5. Average *backchanneling proportion* for each subject across different episodes.

nearly 75% of our subjects spent less than 40% of an episode backchanneling on an average. In fact, only one subject (Subject *G*) had an average backchanneling proportion well in excess of 0.50. On the other hand, all children spent more than 25% of an episode backchanneling, except subjects *I* and *D*. Therefore, when a model is trained with the backchanneling threshold set to 0.25, it expects children to backchannel a lot, and occasionally predicts "*low backchanneling*" as "*high backchanneling*". Nevertheless, these models always predict "*high backchanneling*" windows accurately and hence have high `recall`. On the other hand, when $\tau = 0.50$, the model expects children to rarely backchannel to a great extent. Thus, models trained with $\tau = 0.50$ are more likely to predict "*high backchanneling*" as "*low backchanneling*", especially when the BC proportion is closer to 0.50. However, these models predict "*high backchanneling*" windows with a reasonably high BC proportion with high `precision`.

However, we expect our models to have a perfect balance between `precision` and `recall` regardless of the backchanneling threshold. In this regard, ResNet performed equally well on both the thresholds. Nonetheless, the model performed slightly better in terms of `precision` and `recall` at $\tau = 0.25$. The superior performance of ResNet may be attributed to the fact that it uses more information in the form of time series dynamics to predict the extent of backchanneling.

Next, we focus on our feature ablation experiments wherein we compared the performance of our models using different sets of features inspired by prior work. We first examine the performance of Random Forests models [`Models` 1 to 8]. For both the thresholds ($\tau = 0.25$ and 0.50), all the feature ablations showed similar variations. We observed that the performance of our model improved when trained on all prosodic features, in addition, to pitch [`Model` 1 vs 2]. Other prosodic features like `mfcc[5]'`, `mfcc[6]''`, `mfcc[1]''` and `mfcc[10]''` are in fact, some of the most important features for our Random Forests (Figure 3 (ii)). Adding annotated gaze features to vocal prosody also led to a slight improvement in performance [`Model` 3]. Finally, the best Random Forest model was the one trained on all visual and prosodic features [`Model` 5 for $\tau = 0.25$ and `Model` 4 for $\tau = 0.50$]. The addition of socio-demographic features to visual and prosodic features results in a boost in performance, especially in the case of $\tau = 0.25$. However, `Model` 8 is only of theoretical interest. It has limited practical utility since socio-demographic features may not always be available and their use may raise privacy concerns.

Again for ResNet, results for both values of τ showed the same trend. An interesting observation was that pitch alone was a reasonable predictor of the backchanneling extent [`Model` 9]. Using all other prosodic as well as annotated gaze features results in a slight albeit insignificant improvements [`Model` 10 & 11]. This is because prosodic and annotated gaze features are not as crucial for ResNet as they are for Random Forests, as shown in Figure 3 (iv). Like Random Forest, ResNet trained on all the visual and prosodic features performed the best overall [`Model` 12]. In fact, we found that visual features and especially FAUs were some of the most important for ResNet.

In addition to different feature sets, we also experimented with multiple sets of time series characteristics for Random Forests. The results are consistent with our findings from corresponding LDP experiments discussed in the last section. Given the same set of features (Visual & Prosodic), when τ is set to 0.25, a combination of mean and standard deviation yields the best performance [`Model` 5]. On the other hand,

M.No.	Reference	Features				Aggregates	$\tau = 0.25$				$\tau = 0.5$			
		Annotated	Prosodic	Visual	Demographic		P	R	F1	AUC	P	R	F1	AUC
							Random Forests							
1	Ward, 1996 [Ward, 1996]		✓(Pitch)			Mean	0.61	0.71	0.65	0.66	0.62	0.45	0.32	0.66
2	Jiang, 2011 [Jiang et al., 2011]		✓			Mean	0.77	0.84	0.80	0.86	0.92	0.51	0.40	0.83
3	Lee *et al.*, 2017 [Lee et al., 2017]; Morency *et al.*, 2010 [Morency et al., 2010]	✓(Gaze)	✓			Mean	0.75	0.83	0.81	0.85	0.92	0.44	0.50	0.84
4			✓	✓		Mean	0.90	0.57	0.68	0.87	**0.66**	**0.98**	**0.78**	**0.86**
5			✓	✓		Mean & Stdev	**0.71**	**0.97**	**0.81**	**0.88**	0.90	0.55	0.58	0.86
6			✓	✓		Basic	0.60	0.25	0.30	0.88	0.97	0.58	0.63	0.88
7			✓	✓		Tsfresh	0.66	0.83	0.73	0.73	0.64	0.65	0.55	0.74
8			✓	✓	✓	Mean	0.83	0.88	0.85	0.92	0.85	0.51	0.62	0.88
							ResNet							
9	Ward, 1996 [Ward, 1996]		✓(Pitch)			–	0.52	0.96	0.67	0.55	0.47	0.70	0.49	0.55
10	Jiang, 2011 [Jiang et al., 2011]		✓			–	0.57	0.70	0.68	0.62	0.60	0.69	0.51	0.58
11	Lee *et al.*, 2017 [Lee et al., 2017]; Morency *et al.*, 2010 [Morency et al., 2010]	✓(Gaze)	✓			–	0.55	0.67	0.70	0.63	0.45	0.66	0.55	0.52
12			✓	✓		–	**0.72**	**0.91**	**0.83**	**0.90**	**0.70**	**0.89**	**0.80**	**0.78**

TABLE VI
BACKCHANELING EXTENT PREDICTION: WE REPORT THE PRECISION (P), RECALL (R), F1-SCORE (F1) AND AUC FOR "HIGH BACKCHANNELING" CLASS.

for $\tau = 0.50$ arithmetic mean outperforms all other sets of aggregates [Model 4].

Next, we compare the performance of the Random Forests and ResNet trained on the same features. We observed that ResNet performed better than Random Forests in most settings, except for models trained on prosodic features [Model 2], and a combination of gaze and prosodic features [Model 3] for $\tau = 0.25$. This demonstrates that modelling the dynamics of time series features indeed helps in predicting the extent of backchanneling.

Finally, Figure 3 (ii) and (iv) present the distribution of *importance scores* of the 10 most important features for our Random Forests and ResNet model, respectively, found using Permutation Feature Importance (MDA). For Random Forests, the pitch (F0) of the speaker is the most important factor influencing listener backchannels. This also follows from prior work in predicting listener backchannels [Ward and Tsukahara, 2000, Noguchi and Den, 1998]. Besides, we also observe that the speaker's *pupil dilation*, AU10 (*upper lip raiser*), AU14 (*dimpler*), and *translational head acceleration* are strong predictors of backchanneling responses. To the best of our knowledge, none of these features have been used to predict backchanneling before. To further analyze the combined influence of F0 and *pupil dilation* (best predictors of backchanneling) on the extent of backchanneling, we plotted the 3-D PDP shown in Figure 4 (i). It is evident that both these features have a positive impact on seeking backchannel responses. It is interesting to note that around the range of $0.20 - 0.22$, the speaker's *pupil dilation* causes a sharp increase in the listener's backchanneling response. In the case of ResNet, the speaker's FAUs and *pupil dilation* are the most important features. Unlike Random Forests, prosodic features such as mfcc′ are not strong predictors of backchanneling for ResNet.

D. Importance of time series features

We conclude this section by presenting the results of a short yet important study that validates our hypothesis that the dynamics of time series features are rich sources of information to predict disengagement and backchanneling. Specifically, we used Algorithm 2 to find the importance of time series features on the predictive utility of our LDP and BEP models. We ran the algorithm for 20 independent iterations for both the tasks, and in each iteration, the ResNet model was trained and evaluated using LOSO cross-validation on the permuted dataset (Section III-F). Overall, we observed that there was a significant decrease in the performance of both the LDP and BEP models when trained and tested on the permuted data, as compared to the baseline results [Models 10 & 12] discussed in the previous sections. The LDP model reported an average drop of 29.8% in F1 and 28.6% in AUC. On the other hand, the average drop in F1 and AUC for the BEP model was 36.4% and 25.7%, respectively. This establishes the fact that modeling the temporal characteristics of the features significantly improved the performance of the ResNet model for both the tasks.

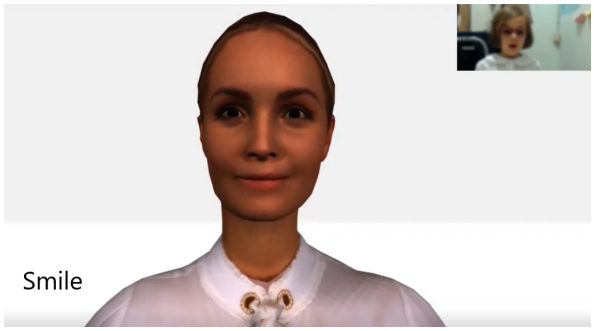


Fig. 6. *Alice, a social listener.* We deployed our BEP model on a 3-D avatar called *Alice*. Alice uses the model to identify backchanneling opportunities.

V. CONCLUSION

In this work, we proposed two models to help peer learning robots come across as active listeners and engaging storytellers to foster early language development in children. Specifically, we formulated Listener Disengagement (LDP) and a Backchanneling Extent Prediction (BEP) models. Extensive experiments with different sets of features and time series characteristics yielded exciting results. To lend interpretability to our models, we used Permutation Feature Importance in addition to Partial Dependency Plots to find important features for our models. We found that visual and vocal features such as *blink rate*, *pupil dilation*, *velocity & acceleration of head*, *FAUs*, and MFCC coefficients were extremely useful in

predicting disengagement and backchanneling extent. To the best of our knowledge, these features have not been used by studies in the past. We also established the utility of time series features in predicting disengagement and the extent of backchanneling. Further, we also examined the influence of socio-demographic features on the amount of time children spent backchanneling and listening to their peers.

To evaluate our models as potential real-time systems, we deployed the best performing LDP and BEP models to predict disengagement and backchanneling during some episodes in the P2PSTORY dataset. We deployed our backchanneling extent prediction model as a lightweight asynchronous python-flask application on a 3-D avatar[‖], called *Alice*. Alice uses our BEP model to identify backchanneling opportunities and chooses a backchannel response (`smile`, `nod`, `head shake`, `eyebrow raise`) at random[**].

A limitation of our work is that we did not evaluate our models on a real social robot, and believe that further experiments are essential to corroborate their effectiveness. However, since children consider robots as social beings [Kahn Jr et al., 2012] and readily learn new information from them [Movellan et al., 2009], we believe that our findings from analysing child-child dyadic interactions will provide useful lessons in improving child-robot interactions.

Acknowledgment

The authors would like to thank Prof. Rajni Jindal for her insightful feedback.

References

Mansi Agarwal, Maitree Leekha, Ramit Sawhney, Rajiv Ratn Shah, Rajesh Kumar Yadav, and Dinesh Kumar Vishwakarma. Memis: Multimodal emergency management information system. In *European Conference on Information Retrieval*, pages 479–494. Springer, 2020.

G Aist and J Mostow. When speech input is not an afterthought: A reading tutor that listens. In *Workshop on Perceptual User Interfaces*, 1997.

G Aist, J Mostow, B Tobin, P Burkhead, A Corbett, A Cuneo, B Junker, and MB Sklar. Computer-assisted oral reading helps third graders learn vocabulary better than a classroom controlabout as well as one-on-one human-assisted oral reading, 2001.

Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. Openface: an open source facial behavior analysis toolkit. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–10. IEEE, 2016.

Shalini Bhatia, Roland Goecke, Zakia Hammal, and Jeffrey F Cohn. Automated measurement of head movement synchrony during dyadic depression severity interviews. In *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, pages 1–8. IEEE, 2019.

Dan Bohus and Eric Horvitz. Models for multiparty engagement in open-world dialog. In *Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 225–234. Association for Computational Linguistics, 2009.

Cynthia Breazeal, Kerstin Dautenhahn, and Takayuki Kanda. Social robotics. In *Springer handbook of robotics*, pages 1935–1972. Springer, 2016.

[‖]We obtained a 3-D model of a middle-aged female from https://renderpeople.com/free-3d-people/ for our demo.

[**]Predicting the type of backchannel response is beyond the scope of our work.

Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

Leo Breiman. Manual on setting up, using, and understanding random forests v3.1, 2002. URL http://oz.berkeley.edu/users/breiman/Using_random_forests_V3.

Hennie Brugman and Albert Russel. Annotating multi-media/multi-modal resources with ELAN. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal, May 2004. European Language Resources Association (ELRA). URL http://www.lrec-conf.org/proceedings/lrec2004/pdf/480.pdf.

Maximilian Christ, Nils Braun, Julius Neuffer, and Andreas W Kempa-Liehr. Time series feature extraction on basis of scalable hypothesis tests (tsfresh–a python package). *Neurocomputing*, 307: 72–77, 2018.

Herbert H Clark, Susan E Brennan, et al. Grounding in communication. *Perspectives on socially shared cognition*, 13(1991):127–149, 1991.

Stephanie M. Curenton. Narratives as learning tools to promote school readiness. *Early Education and Development*, 21(3):287–292, 2010. doi: 10.1080/10409289.2010.485532. URL https://doi.org/10.1080/10409289.2010.485532.

Iwan de Kok and Dirk Heylen. The multilis corpus–dealing with individual differences in nonverbal listening behavior. In *Toward Autonomous, Adaptive, and Context-Aware Multimodal Interfaces. Theoretical and Practical Issues*, pages 362–375. Springer, 2011.

Alan R Dennis and Susan T Kinney. Testing media richness theory in the new media: The effects of cues, feedback, and task equivocality. *Information systems research*, 9(3):256–274, 1998.

Allen T Dittmann. Developmental factors in conversational behavior. *Journal of Communication*, 22(4):404–423, 1972.

Sidney D'Mello, Rick Dale, and Art Graesser. Disequilibrium in the mind, disharmony in the body. *Cognition & emotion*, 26(2): 362–374, 2012.

Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 33(4):917–963, 2019.

Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

Ben D Fulcher. Feature-based time-series analysis. In *Feature engineering for machine learning and data analytics*, pages 87–116. CRC Press, 2018.

Patricia Garrett, Nicholas Ng'andu, and John Ferron. Poverty experiences of young children and the quality of their home environments. *Child development*, 65(2):331–345, 1994.

Mononito Goswami, Lujie Chen, and Artur Dubrawski. Discriminating cognitive disequilibrium and flow in problem solving: A semi-supervised approach using involuntary dynamic behavioral signals. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34. AAAI, 2020.

Zeeshan Ali Haq and Ziaul Hasan. Eye-blink rate detection for fatigue determination. In *2016 1st India International Conference on Information Processing (IICIP)*, pages 1–5. IEEE, 2016.

Dongmei Jiang, Yulu Cui, Xiaojing Zhang, Ping Fan, Isabel Ganzalez, and Hichem Sahli. Audio visual emotion recognition based on triple-stream dynamic bayesian network models. In Sidney D'Mello, Arthur Graesser, Björn Schuller, and Jean-Claude Martin, editors, *Affective Computing and Intelligent Interaction*, pages 609–618, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg. ISBN 978-3-642-24600-5.

Rajni Jindal, Maitree Leekha, Minkush Manuja, and Mononito Goswami. What makes a better companion? towards social & engaging peer learning. In *Proceedings of the Twenty-forth European Conference on Artificial Intelligence*. IOS Press, 2020.

Peter H Kahn Jr, Takayuki Kanda, Hiroshi Ishiguro, Nathan G Freier, Rachel L Severson, Brian T Gill, Jolina H Ruckert, and Solace Shen. robovie, you'll have to go into the closet now: Children's social and moral relationships with a humanoid robot. *Developmental psychology*, 48(2):303, 2012.

Mariska E Kret. The role of pupil size in communication. is there room for learning? *Cognition and Emotion*, 32(5):1139–1145, 2018.

Jin Joo Lee, Cynthia Breazeal, and David DeSteno. Role of speaker cues in attention inference. *Front. Robotics and AI*, 4:47, 2017. ISSN 2296-9144. doi: 10.3389/frobt.2017.00047. URL https://www.frontiersin.org/article/10.3389/frobt.2017.00047.

Zofia Malisz, Marcin Włodarczak, Hendrik Buschmeier, Joanna Skubisz, Stefan Kopp, and Petra Wagner. The alico corpus: analysing the active listener. *Language resources and evaluation*, 50(2):411–442, 2016.

Andrew A. McReynolds, Sheba P. Naderzad, Mononito Goswami, and Jack Mostow. Toward learning at scale in developing countries: Lessons from the global learning xprize field study. In *Provisionally accepted to the Seventh Annual ACM Conference on Learning at Scale*. ACM, 2020.

Marek P Michalowski, Selma Sabanovic, and Reid Simmons. A spatial model of engagement for a social robot. In *9th IEEE International Workshop on Advanced Motion Control, 2006.*, pages 762–767. IEEE, 2006.

Louis-Philippe Morency, Iwan de Kok, and Jonathan Gratch. A probabilistic multimodal approach for predicting listener backchannels. *Autonomous Agents and Multi-Agent Systems*, 20(1):70–84, 2010.

Jack Mostow, Greg Aist, Paul Burkhead, Albert Corbett, Andrew Cuneo, Susan Eitelman, Cathy Huang, Brian Junker, Mary Beth Sklar, and Brian Tobin. Evaluation of an automated reading tutor that listens: Comparison to human tutoring and classroom instruction. *Journal of Educational Computing Research*, 29(1):61–117, 2003.

Javier Movellan, Micah Eckhardt, Marjo Virnes, and Angelica Rodriguez. Sociable robot improves toddler vocabulary skills. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 307–308. ACM, 2009.

Hiroaki Noguchi and Yasuharu Den. Prosody-based detection of the context of backchannel responses. In *Fifth International Conference on Spoken Language Processing*, 1998.

Behnaz Nojavanasghari, Tadas Baltrušaitis, Charles E Hughes, and Louis-Philippe Morency. Emoreact: a multimodal approach and dataset for recognizing emotional responses in children. In *Proceedings of the 18th acm international conference on multimodal interaction*, pages 137–144, 2016.

Hae Won Park, Mirko Gelsomini, Jin Joo Lee, and Cynthia Breazeal. Telling stories to robots: The effect of backchanneling on a child's storytelling. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, pages 100–108. IEEE, 2017a.

Hae Won Park, Mirko Gelsomini, Jin Joo Lee, and Cynthia Breazeal. Telling stories to robots: The effect of backchanneling on a child's storytelling. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, pages 100–108. IEEE, 2017b.

Carole Peterson, Beulah Jesso, and Allyssa McCabe. Encouraging narratives in preschoolers: An intervention study. *Journal of child language*, 26(1):49–67, 1999.

Ronald Poppe, Khiet P Truong, Dennis Reidsma, and Dirk Heylen. Backchannel strategies for artificial listeners. In *International Conference on Intelligent Virtual Agents*, pages 146–158. Springer, 2010.

Helen Raikes, Barbara Alexander Pan, Gayle Luze, Catherine S Tamis-LeMonda, Jeanne Brooks-Gunn, Jill Constantine, Louisa Banks Tarullo, H Abigail Raikes, and Eileen T Rodriguez. Mother–child bookreading in low-income families: Correlates and outcomes during the first three years of life. *Child development*, 77(4):924–953, 2006.

James Rehg, Gregory Abowd, Agata Rozga, Mario Romero, Mark Clements, Stan Sclaroff, Irfan Essa, O Ousley, Yin Li, Chanho Kim, et al. Decoding children's social behavior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3414–3421, 2013.

Meredith L Rowe, Barbara Alexander Pan, and Catherine Ayoub. Predictors of variation in maternal talk to children: A longitudinal study of low-income families. *Parenting: Science and Practice*, 5(3):259–283, 2005.

Evangelos Sariyanidi, Hatice Gunes, and Andrea Cavallaro. Automatic analysis of facial affect: A survey of registration, representation, and recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(6):1113–1133, 2014.

Hollis S Scarborough and Wanda Dobrich. On the efficacy of reading to preschoolers. *Developmental review*, 14(3):245–302, 1994.

Candace L Sidner, Christopher Lee, Cory Kidd, Neal Lesh, and Charles Rich. Explorations in engagement for humans and robots. *arXiv preprint cs/0507056*, 2005.

Nikhita Singh, Jin Joo Lee, Ishaan Grover, and Cynthia Breazeal. P2pstory: Dataset of children as storytellers and listeners in peer-to-peer interactions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 434:1–434:11, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-5620-6. doi: 10.1145/3173574.3174008. URL http://doi.acm.org/10.1145/3173574.3174008.

Catherine S Tamis-LeMonda and Eileen T Rodriguez. Parents role in fostering young childrens learning and language development. *Encyclopedia on early childhood development*, 1:1–11, 2008.

Keith Topping and Stewart Ehly. *Peer-assisted learning*. Routledge, 1998.

Peter Tymms, Christine Merrell, Allen Thurston, John Andor, Keith Topping, and David Miller. Improving attainment across a whole district: school reform through peer tutoring in a randomized controlled trial. *School Effectiveness and School Improvement*, 22(3):265–289, 2011.

Nigel Ward. Using prosodic clues to decide when to produce back-channel utterances. In *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96*, volume 3, pages 1728–1731. IEEE, 1996.

Nigel Ward and Wataru Tsukahara. Prosodic features which cue back-channel responses in english and japanese. *Journal of pragmatics*, 32(8):1177–1207, 2000.

Victor H Yngve. On getting a word in edgewise. In *Chicago Linguistics Society, 6th Meeting, 1970*, pages 567–578, 1970.

Ageing Research Reviews 61 (2020) 101078

Contents lists available at ScienceDirect

Ageing Research Reviews

journal homepage: www.elsevier.com/locate/arr

Review

Ubiquitin biology in neurodegenerative disorders: From impairment to therapeutic strategies

Dhiraj Kumar, Rashmi K. Ambasta, Pravir Kumar*

Molecular Neuroscience and Functional Genomics Laboratory, Delhi Technological University (Formerly Delhi College of Engineering), Delhi 110042, India

ARTICLE INFO

Keywords:
Ubiquitin E3 ligase
Ubiquitin proteasome system
Ubiquitination
Neurodegenerative diseases
Signaling
Therapeutics

ABSTRACT

The abnormal accumulation of neurotoxic proteins is the typical hallmark of various age-related neurodegenerative disorders (NDDs), including Alzheimer's disease, Parkinson's disease, Huntington's disease, Amyotrophic lateral sclerosis and Multiple sclerosis. The anomalous proteins, such as Aβ, Tau in Alzheimer's disease and α-synuclein in Parkinson's disease, perturb the neuronal physiology and cellular homeostasis in the brain thereby affecting the millions of human lives across the globe. Here, ubiquitin proteasome system (UPS) plays a decisive role in clearing the toxic metabolites in cells, where any aberrancy is widely reported to exaggerate the neurodegenerative pathologies. In spite of well-advancement in the ubiquitination research, their molecular markers and mechanisms for target-specific protein ubiquitination and clearance remained elusive. Therefore, this review substantiates the role of UPS in the brain signaling and neuronal physiology with their mechanistic role in the NDD's specific pathogenic protein clearance. Moreover, current and future promising therapies are discussed to target UPS-mediated neurodegeneration for better public health.

1. Introduction

According to the united nation's 2019 statistical reports of world population, a steady rise from one in eleven people (9%) to one in six people (16 %) with age of over 65 years by 2050. Further, it is projected to become triple, from 143 million in 2019 to 426 million in 2050 with 80 years aged individuals (United Nations, 2019). While the aging figures imply an improved life expectancy, but neurodegenerative pathologies remained an unresolved mystery for the aged people. Here, age and sex are the two important variables affecting human health and diseases, one being the main risk factor, while other impacting the disease incidences, severity, and the treatment responses. In fact, post mortem and living brain studies have revealed neuro-epigenetic signatures of age and sex in our brain (Gilbert et al., 2019). Moreover, timely research has demonstrated numerous neurodegenerative mechanisms in the brain, including mitochondrial dysfunctions, ROS production, protein oxidation, lipid peroxidation, ER stress, nitrosative stress, protein aggregation (amyloid, tau, synuclein, Lewy bodies, and

Abbreviations: AD, Alzheimer's disease; ALS, amyotrophic lateral sclerosis; ASK1, apoptosis signal-regulating kinase 1; AMPA, α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid; AMPAR, α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor; Aβ, amyloid beta; AKT, protein kinase B; ARE, adenylate-uridylate-rich element; ATP, adenosine triphosphate; CaMKII, calcium-calmodulin-dependent protein kinase II; cAMP, cyclic adenosine monophosphate; CD28/36, cluster of differentiation 28/36; CLIPTAC, click-formed PROTAC; cPLA2, cytosolic phospholipases A2; CREB, cAMP response element-binding protein; CREM, cAMP-responsive element modulator; DNA, deoxyribonucleic acid; DUBs, deubiquitinating enzymes; ERK1/2, extracellular signal-regulated protein kinases 1 and 2; E1s, ubiquitin-activating enzymes; E2s, ubiquitin-conjugating enzymes; E3s, ubiquitin-ligating enzyme; ER, endoplasmic reticulum; ERAD, endoplasmic reticulum associated degradation; FOXO, forkhead family of transcription factors; GABA, gamma-aminobutyric acid; GluA1, glutamate ionic receptor A1; GluR, glutamate receptor; HD, Huntington's disease; HSR, heat stress response; IFN-γ, interferon gamma; JNK, c-jun N-terminal kinase; KEAP1, Kelch-like ECH-associated protein 1; LBD, Lewy body disease; MAPK, mitogen-activated protein kinase; MDM2, mouse double minute 2 homolog; mHtt, mutant huntingtin; MHC, major histocompatibility complex; MS, multiple sclerosis; NDDs, neurodegenerative disorders; NEDD8, neural precursor cell expressed, developmentally down-regulated 8; NF-kB, nuclear factor kB; NMDA, N-methyl-D-aspartate; NMDAR, N-methyl-D-aspartate receptor; Nrf2, nuclear factor erythroid 2–related factor 2; NUB1, negative regulator of ubiquitin like proteins 1; PA28, proteasome activator; PD, Parkinson's disease; PI3K, phosphoinositide 3-kinase; PKA/C, protein kinase A/C; PROTAC, proteolysis targeting chimera; PRU, pleckstrin-like receptor for ubiquitin domain; ROS, reactive oxygen species; SOD1, superoxide dismutase 1; SUMO1, small ubiquitin-like modifier 1; TCR, T-cell receptor; TDP-43, TAR DNA-binding protein 43; TUBE, tandem ubiquitin binding entity; UBA, ubiquitin associated domain; UBB^{+1}, mutant ubiquitin; USP9X, ubiquitin specific peptidase 9 X; Ubl, ubiquitin like protein; UPS, ubiquitin proteasome system

* Corresponding author at: Molecular Neuroscience and Functional Genomics Laboratory, Department of Biotechnology, Delhi Technological University (Formerly Delhi College of Engineering), Room # FW4TF3, Mechanical Engineering Building, Shahabad Daulatpur, Bawana Road, Delhi 110042, India.
E-mail address: pravirkumar@dtu.ac.in (P. Kumar).

https://doi.org/10.1016/j.arr.2020.101078
Received 6 November 2019; Received in revised form 24 March 2020; Accepted 22 April 2020
Available online 12 May 2020
1568-1637/ © 2020 Elsevier B.V. All rights reserved.

inclusion bodies), DNA damage, microglial activation, neuro inflammation, and proteasomal dysfunction. Ubiquitin proteasome system (UPS) dysfunction is central to all the neurodegeneration mechanisms, since it regulates the proteostasis in nearly all these processes (Douglas and Dillin, 2010). Interestingly, disturbed proteostasis as a consequence of aging have also reported to cause circadian and sleep disturbances in the patients and are believed to drive disease pathogenesis early in the neurodegeneration (Musiek and Holtzman, 2016). Altogether, they affect specific neurons in distinct parts of the brain, i.e. selective neuronal and regional vulnerability in different neurodegenerative disorders (NDDs), including Alzheimer's disease (AD), Parkinson's disease (PD), Huntington's disease (HD), Amyotrophic lateral sclerosis (ALS), and Multiple sclerosis (MS). In addition, their progression in the brain occurs in a stereotypical and predictable fashion, whose mechanism is still unknown (Fu et al., 2018). Additionally, these pathologies are governed by a systemic response that includes glial cells, vascular cells and peripheral cells, apart from neurons as a sophisticated network in the brain. Therefore, integrative biology of systemic proteostasis equilibrium in the human brain needs to be addressed for unraveling the potential therapeutic modalities to target the neurodegenerative pathologies.

Like every cell, neuron has an efficient system to clear toxins and to maintain cellular-integrity with help of a well-coordinated ubiquitin coding system. Here, ubiquitin coding is a type of post-transnational modification of a protein that regulates its stability, longevity, functionality and localization in a cell (Hershko, 2005). The functional UPS machinery is comprised of multiple components, including ubiquitin protein, ubiquitin activating enzyme (E1), ubiquitin conjugating enzyme (E2), ubiquitin ligating enzyme (E3), deubiquitinating enzymes (DUBs), and the proteasome. Initially, ubiquitin is activated by E1 enzyme via ATP hydrolysis, i.e. adenylation of the ubiquitin C-terminal carboxyl group and its transfer to a cysteinyl residue on E1 enzyme (Rose, 2005). It is followed by the conjugation of ubiquitin, i.e. transfer of activated ubiquitin from E1 to a cysteinyl residue of E2 enzyme. The final step of ubiquitin transfer to the substrate protein's lysine residue with isopeptide bond takes place either directly from E2 or after thioester formation of ubiquitin with the E3 ligase enzyme. These three main steps of ubiquitin-activation, conjugation, and ligation are known as Ubiquitination. The subsequent ubiquitin attachment at the different lysine site on substrate protein (multi-monoubiquitination) or on the previously attached ubiquitin (polyubiquitination) determines the cellular fate depending upon the state and conformations of polyubiquitin chains, including methionine (M1), and lysine [K6, K11, K27, K29, K33, K48, and K63; (Callis, 2014)]. These ubiquitination signals regulate diverse biological functions in neurons, such as DNA damage responses, endocytosis, neuronal-development, neural network morphogenesis, synapse remodeling, ribosome biogenesis, cell adaptation and signal transduction pathways (Kwon and Ciechanover, 2017). Interestingly, ubiquitination is a cross-road between cell survival and death due to its potent ability to clear malfunctioned or cytotoxic proteins that enable cell to overcome proteotoxic stress. Besides, cell adopts any of these mechanisms, including proteasomes and autophagy to target the ubiquitinated protein for its selective degradation. In fact, investigations on the model substrates revealed Lys48 polyubiquitination as a predominant signal for proteasomal degradation while others known to direct protein-protein interactions to form multi-protein complex and modulate the activities of their substrates. Since, Lys-48 linked poly ubiquitin chain form a closed conformation where hydrophobic patches of ubiquitin molecules sequestered at the ubiquitin-ubiquitin interfaces, thereby disallowing them to interact with other target receptors. While, Lys-63 linked polyubiquitin chain, forms an extended structure exposing their hydrophobic patches thereby allowing their interaction with the receptor proteins. In this way, Lys-48 and Lys-63 polyubiquitin chains are the predominant markers for proteasomal and autophagosomal protein clearance from the cell (Fig. 1), and affirm ubiquitination as a shared signal for protein quality control in neurons either through their targeting for diverse cellular functions or targeting them for degradation via proteasomal or autophagosomal/lysosomal pathway (Ji and Kwon, 2017).

2. How ubiquitination is crucial for neuronal signaling and brain homeostasis?

Over hundred billion interconnected wire-frame networks of neurons transmit well-coordinated electrochemical signals to regulate complex physiological functions in our brain. In general, these functions include Pre- and Post-synaptic regulation, neurotransmitter's vesicle transport and release, synaptic density and plasticity, and neuronal receptor modulations (DiAntonio et al., 2001; Lin and Man, 2013). Moreover, ubiquitination determines learning and memory functions via regulating the abundance of selected proteins involved in the synaptic plasticity (Caroni et al., 2012). Interestingly, ubiquitination is also associated with drug addiction mediated behavioral changes and neuronal adaptations via targeting drug induced proteins for proteasomal degradation (Massaly et al., 2015). Furthermore, it also controls the transcription process in neurons by regulating the degradation of transcriptional repressors like IκB and CREM (cAMP-responsive element modulator) (Liu and Chen, 2011). Alike neurons, ubiquitination also acts in glial cells like astrocytes and microglia to maintain protein homeostasis, cytokine balance and healthy environment in the brain. Therefore, it is necessary to discuss the ubiquitin system's biology in the brain for proper understanding of complex biophysiological functions to regulate brain homeostasis.

2.1. UPS mediated synaptic plasticity: A gateway for learning and memory functions in the brain

The substantial research in several model systems has forced us to believe that synapses are the principal site for information storage in the brain. However, memory consolidation relies on the strength of the synapses in different parts of the brain, including their initial storage in the hippocampus followed by their transfer to the cortex region (Abraham et al., 2019). Interestingly, UPS plays a vital role in controlling these synaptic connections during memory and learning processes of synaptic plasticity via regulating protein synthesis and degradation (Bingol and Sheng, 2011). Both short-term and long-term synaptic plasticity's regulation included the turnover of diverse protein kinases- cAMP-dependent protein kinase A (PKA), calcium-calmodulin-dependent protein kinase II (CaMKII), protein kinase C (PKC), and mitogen-activated protein kinase (MAPK). For instance, NMDA receptor's stimulation triggered CaMKII autophosphorylation and consequent redistribution of proteasomes from dendritic shaft to the spine and its proteasomal activity in hippocampal neurons (Bingol et al., 2010). Similarly, serotonin stimulated sensory neurons exhibited proteasomal degradation of inhibitory R subunit of PKA and subsequent PKA and PKC mediated degradation of CREB repressor- CREB1b to regulate long-term facilitation in *Aplysia* model (Upadhya et al., 2004). Likewise, a recent cross-talk between GluA1 phosphorylation and ubiquitination demonstrated that GluA1 ubiquitination causes enhanced protein kinase A (PKA) phosphorylation at Ser-845 (Guntupalli et al., 2017; Widagdo et al., 2017). The former study indicates the potential of ubiquitination in PKA mediated phosphorylation of AMPA receptor subunits (GluR1 and GluR4) to trigger their synaptic incorporations in neuronal membrane during long-term potentiation (Esteban et al., 2003; Zheng and Keifer, 2009). On contrary, MAPK family proteins such as ERK1/2 is responsible for gene transcription, protein translation and post-translational modification during information processing and memory formation while others, including ERK5, p38 or JNKs have a distinct role in learning and memory (Medina and Viola, 2018). Besides, proteasomal degradation of NF-kB regulators- IjB and IkB kinases regulated gene transcription and its associated formation and stability of fear memory in the amygdala (Jarome et al., 2011). Moreover, UPS

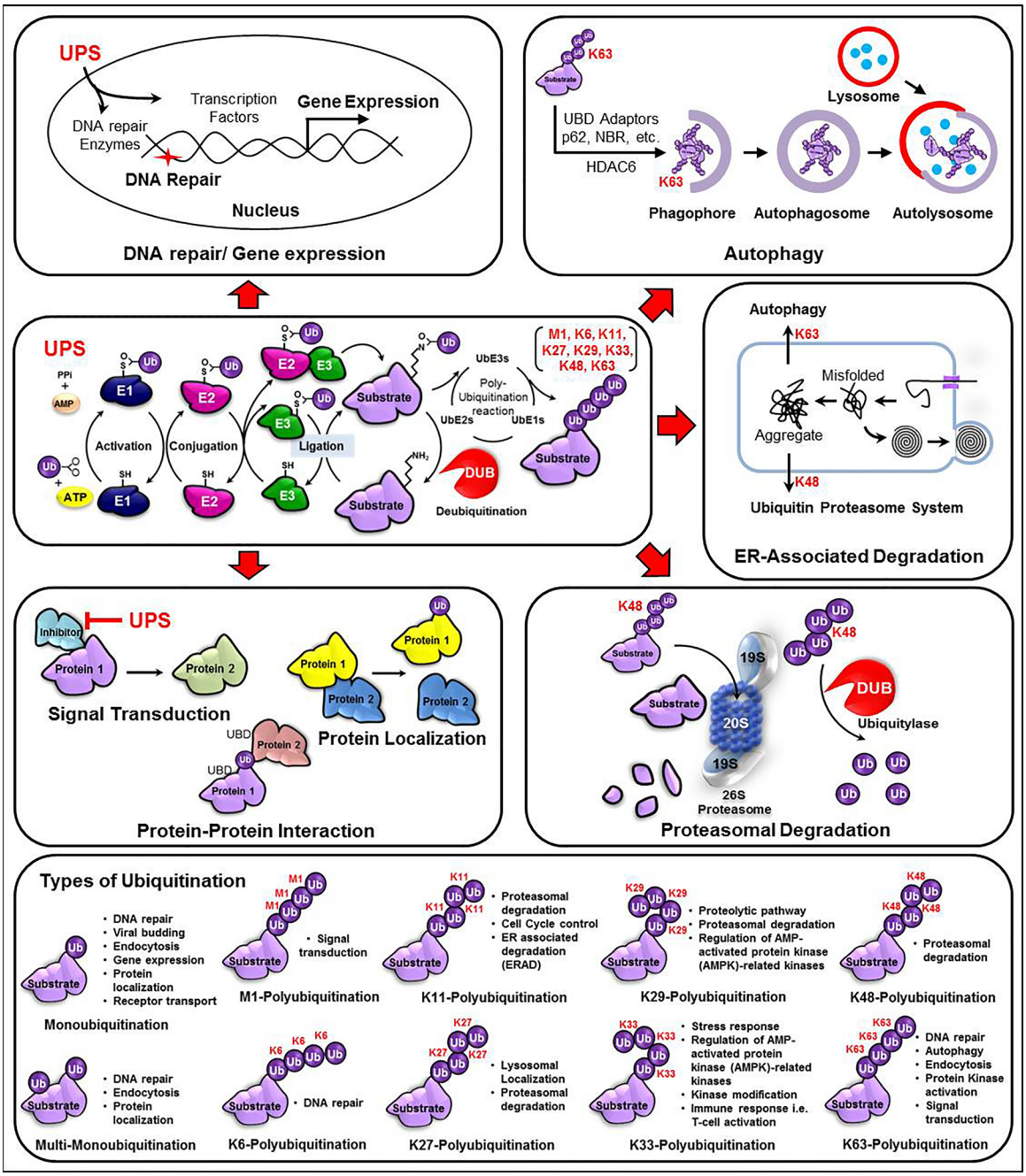


Fig. 1. The ubiquitination process and their cellular responses based on the ubiquitin chain architecture: The ubiquitination involves the complex interplay of ubiquitin activating enzymes (E1s), ubiquitin conjugating enzymes (E2s), ubiquitin ligases (E3s) and deubiquitinases (DUBs) to form a variety of ubiquitin tags, including mono-ubiquitin, multi-mono-ubiquitin and poly-ubiquitin chains (M1, K6, K11, K27, K29, K33, K48, K63) that perform a variety of functions in a cell. UPS- Ubiquitin proteasome system; Ub- Ubiquitin, UBD- Ubiquitin binding domain; p62- Nucleoprotein 62, NBR- Neighbor of BRCA1 gene; HDAC6- Histone deacetylase 6.

also regulates both pre- and post-synaptic protein densities and receptor shuttling to govern synaptic transmission activity during memory and learning process. For instance, NMDAR activity dependent proteasomal degradation of MOV10 resulted in translational activation of alpha-CaMKII, Limk1, and the depalmitoylating enzyme lysophospholipase1 (Lypla1) during synaptic plasticity (Banerjee et al., 2009). Other synaptic proteins like Dunc-13 (pre-synaptic vesicle priming protein), Syntaxin 1 (synaptic vesicle exocytosis protein), and RIM1α (pre-synaptic scaffolding protein) are proteasomally degraded by different E3 ligases namely Fbxo45, STARING (Syntaxin 1-interacting RING finger protein) and SCRAPPER respectively (Chin et al., 2002; Hsieh et al., 2018; Yao et al., 2007). Similarly, UPS also modulates structural proteins, neurotransmitter receptors and regulatory proteins in postsynaptic terminal. For instance, proteasome inhibitor MG132 prevented internalization of membrane receptors, including ionotropic GLR-1 glutamate receptors and NMDA-induced AMPA receptors (Widagdo et al., 2015). In addition, postsynaptic scaffold protein PSD-95 stabilized the surface expression of AMPA receptors, while its proteasomal degradation by Mdm2 (E3 ligase) triggered AMPA receptor internalization (Colledge et al., 2003). Similarly, NMDA receptors also retro translocated and degraded by UPS with the help of SCF-type ligase and contribute towards short-term synaptic plasticity regulation (Kato et al., 2005). Likewise, the glycine and GABA receptors also internalized and degraded upon ubiquitination in the neurons (Lin and Man, 2013). Furthermore, other post-synaptic scaffolding proteins such as GRIP1, GKAP, Shank, SPAR, and Arc have a role in synaptic strength

Table 1
Key role of E3 ligases in synaptic plasticity mediated memory and learning processes within the brain.

S.No.	E3 ligases	Role in Memory and Learning	Target Protein	Experimental Model	References
1	APC^{Cdh1}	Interacts with EphA4 receptor to mediate ephrin-induced proteasomal degradation of AMPARs	AMPAR	EphA4 knockout mice	(Fu et al., 2011)
2	CBX4	Contextual fear conditioning memory Formation	HDAC7	C57BL6J mice	(Jing et al., 2017)
3	CRL4	Target voltage- and calcium-activated BK channel for ER retention	Cereblon, BK channel	CRBN mutant mice	(Song et al., 2018)
4	Fbx2	Controls synaptic NR1and contribute to homeostatic regulation of NMDA receptor	NR1	Hippocampal neurons	(Kato et al., 2005)
5	MDM2	Glutamate receptor trafficking	PSD-95	Rat forebrain crude synaptosomes	(Colledge et al., 2003)
6	MGRN1	Potentiates the transcriptional cellular response to proteotoxic stress	–	C57BL/6J mice and Brain lysate	(Benvegnù et al., 2017a)
7	Nedd4	Role in trafficking of GluA1-containing AMPAR	GluA1, AMPAR	Nedd4(+/-) mice	(Camera et al., 2016)
8	Parkin	Role in short term memory through regulation of hippocampal synaptic plasticity	Synphilin, α-synuclein, synaptotagmin XI, Pael receptor, tubulin, p38/JTV-1, mitofusin	Parkin-deficient mice	(Rial et al., 2014)
9	RNF167	Regulates GluA2 associated synaptic AMPAR currents	GluA2, AMPAR	Sprague–Dawley rats	(Lussier et al., 2012)
10	$SCF^{\beta\text{-}TRCP}$	Regulate postsynaptic density of spine-associated Rap GTPase activating protein (SPAR) in association with Polo-like Kinase 2	SPAR	E19 Long Evans rat hippocampi	(Ang et al., 2008)
11	SCRAPPER	Role in RIM1 (Rab3-interacting molecule-1) mediated synaptic vesicle release and neurotransmission	RIM1	SCR (+/−) mice	(Yao et al., 2011)
12	TRIM3A	Role in γ-actin turnover at hippocampal synapses regulating spine densities, long-term potentiation, and short-term contextual fear memory consolidation	γ-actin	Trim3 −/− mice	(Schreiber et al., 2015)
13	TRIM67	Role in appropriate brain development and behavior	Netrin receptor DCC	Trim67-/- mice	(Boyer et al., 2018)
14	TRIM9	Role in neuronal morphogenesis and spatial learning and memory	SNAP25, Netrin-1	Trim9-/- mice	(Winkle et al., 2016)
15	UBE3A	Regulate postsynaptic SK2 levels for NMDA receptor activation. Also regulate glutamate receptor trafficking and CaMKII abundance	SK2, Arc	UBE3A-deficient mice	(Mabb and Ehlers, 2010; Sun et al., 2015)

regulation by their own proteasomal degradation to control the surface expression of diverse receptors, including AMPA (GluR2) and NMDA by different E3 ligases- TRIM3, SCFβ-TRCP, and Triad3A (Ferrara et al., 2019; Lin and Man, 2013; Mabb et al., 2014). There is growing list of E3 ligases regulating the synaptic proteins and receptors to modulate learning and memory process in the brain (Table 1). In summary, kinase action contributes towards memory formation by two ways, including the regulation of ion channel density/conductivity, i.e. regulation of membrane receptor trafficking and regulation of gene transcription/ local translation that governs synaptic transmission and synaptogenesis respectively.

2.2. Redox regulated UPS activity governs lipid metabolism, inflammatory responses and cellular immunity in the brain

The brain is main metabolizers of oxygen, but with relatively delicate defensive antioxidant mechanisms. Here, mitochondrial machinery metabolizes oxygen to produce ATP that yields several reactive oxygen species (ROS) including oxygen ions, free radicals, and peroxides in neurons. In fact, low level of ROS ensures the activation of various signaling pathways through transcriptional activation, while its abundance causes oxidative or nitro-oxidative stress in neurons. In fact, various studies have demonstrated the role of oxidative stress in UPS impairment thereby aggravating the pathogenesis of NDDs. For instance, High levels of many polyubiquitinated proteins belonging to protein synthesis, quality control system, energy metabolism, and cytoskeleton network are reported in the human AD brain with respect to the age-matched controls. Moreover, the identification of quality control system proteins, including PPIA, HSP-90β, and UCH-L1 in the pool signifies oxidative damage linked UPS impairment and proteotoxicity in AD (Tramutola et al., 2018). Similarly, Nitrosative stress is found to S-nitrosylate E3-ligase Parkin and consequent downregulation of E3 ligase-UPS pathway providing a link between free radical toxicity and protein accumulation in PD (Chung et al., 2004). Likewise, another study has revealed defective form of UCH-L1, a major target of oxidative damage, supporting a direct link between oxidative stress and UPS dysfunction in the pathogenesis of AD and PD (Choi et al., 2004). Further, proteomic analysis has identified other unconventional ubiquitin chain targets like Ubc6 that are critical for proteasomal function (Xu et al., 2009). In addition, they trigger severe inflammatory responses against the degenerated neurons and muscles, oxidized proteins, and lipid peroxidation, or glycated products in the brain (Popa-Wagner et al., 2013). For instance, ROS mediated phosphorylation of NF-κB inhibitor by IκB kinase (IKK) may regulate proteasomal degradation of IκB and consequent release of NF-κB (Alkalay et al., 1995; Kanarek and Ben-Neriah, 2012; Rajakumar et al., 2013). Thus, rescued NF-κB migrates into the nucleus and promotes the transcription of pro-inflammatory mediators along with c-Fos/Activator protein 1 (Collins et al., 2016). Another regulator of redox signaling; NRF2 remains bound to its adaptor KEAP1 under basal conditions and are maintained at a low level in the cytoplasm through ubiquitination. Though in the diseased state it migrates into the nucleus and binds to the ARE sequence and consequently, trigger transcriptions of multiple genes involved in the expression of antioxidants, efflux transporters, glutathione, ATP synthesis, and several proteasomal subunits, etc. (Sivandzade et al., 2019; Changjiang et al., 2018). Furthermore, oxidative stress initiates the release of pleiotropic cytokines like IFN-γ that induces the proteasomal activator PA28 (11S multimeric complex) to trigger over expression of immunoproteasomes and consequent MHC class- I mediated antigenic peptide's presentation in astrocytes and microglial cells (Jansen et al., 2014; Launay et al., 2013). Besides, oxidative stress as well linked with the PI3K/AKT pathway mediated over expression of the FOXO transcription factors that guide neuronal survival and stress responses in nervous system (Lefaki et al., 2017; Uranga et al., 2013). In addition, UPS also protects against lipid peroxidation and consequent toxin production, which may otherwise cause neural apoptosis

(Nagamine et al., 2010). For instance, a UPS component, Parkin monoubiquitinates CD36 and stabilize it at plasma membrane to facilitate the uptake of fatty acids stored in lipid droplets or oxidized by mitochondria. Moreover, under cellular stress, CD36 triggers Ca2+- and ERK1/2-mediated activation of cPLA2 to release arachidonic acid from ER and nuclear membrane phospholipids that aids in synaptic transmission and its conversion into bioactive eicosanoids by microglia to have pleiotropic effects on inflammation (Abumrad and Moore, 2011). UPS also regulates both T-cell receptor (TCR) and co-stimulatory CD28 signaling via Cbl family ubiquitin ligases and USP9X deubiquitinase mediated regulation of pro-inflammatory cytokines and their receptor's expression to direct effective immunity and self-tolerance (Naik et al., 2014). These scientific advancements have improved our understanding about the ubiquitin system mediated redox biology, lipid metabolism, inflammation and immune system regulation in the brain cells by checking the levels of ROS, oxidized proteins, transcriptional factors, and pro-inflammatory agents within limits to maintain cellular homeostasis.

This way, ubiquitination system regulates protein turnover in neurons and glial cells to facilitate multifaceted processes of neuronal signaling, synaptic regulation and neuronal adaptations to maintain brain homeostasis (Fig. 2).

3. Intervening roles of UPS in the neurodegenerative pathology

3.1. UPS in drug addiction mediated behavioral changes and neuronal adaptations

Drug addiction is a compulsive pattern of drug seeking and enduring propensity to relapse it, in effect, of the neuro-adaptations in specific brain regions analogous to long-term memory consolidation. Latest research has defined drug addiction mechanism to be associated with alterations in synaptic transmission in mesocorticolimbic and corticostriatal pathways and the transcriptional changes in neurons for their durable changes in plasticity and behavior. In fact, UPS is evident to control the trafficking or internalization of various receptors, including Glutamate (mGluRs, AMPARs, NMDARs) and non-glutamate receptors (GABA, nAChRs) for maintaining proper synaptic transmissions. For instance, cocaine addiction mediated synaptic plasticity changes have been observed in the nucleus accumbens (NAcc) (Massaly et al., 2013). Similarly, proteasomal degradation of N-ethylmaleimide-sensitive fusion (NSF) protein (a post synaptic density protein), modulate AMPA glutamate receptor's expression and involved in cocaine addiction mediated memory reconsolidation (Ren et al., 2013). In addition, repeated exposure to both amphetamine and cocaine altered glutamatergic neurotransmission in NAcc and potentiated AMPAR synaptic transmission, while synaptic de-potentiation was only associated with cocaine-induced protein synthesis and metabotropic glutamate receptor-5 (mGluR5) activation in NAcc medium spiny neurons (Jedynak et al., 2016). Likewise, proteasomal degradation of GKAP and Shank destabilized PSD-95 and consequently, down regulated NR2B subunit of glutamate N-methyl-Daspartate receptors (NMDARs) to trigger behavioral sensitization in amphetamine addictions in NAcc (D'Souza, 2015; Hung et al., 2010). Another instance where protein ubiquitination is involved in the neuroplastic changes in nAChR's associated synaptic transmission underlying alcohol and nicotine addictions (Feduccia et al., 2012). Furthermore, morphine, cocaine, and nicotine evoked synaptic plasticity was reported with nitric oxide and BDNF induced ubiquitination and reduction in GABA and GABARs in ventral tegmental area and prefrontal cortex of the brain (Lüscher and Malenka, 2011). These studies have highlighted the importance of investigating diverse ubiquitination enzymes involved in the regulation of drug-induced neuroplastic changes to mimic behavioral pathology in human addicts.

3.2. Intricacies of UPS in the pathophysiology of neurodegenerative disorders

A prominent feature of neurodegenerative disorders includes the accumulation of misfolded proteins in response to overwhelming effect of degradative pathways in neurons causing cytotoxicity and neurodegeneration. For instance, the histopathology of AD brain is marked by the presence of neurofibrillary tangles and amyloid containing senile plaques throughout the brain. Moreover, they exhibit progressive atrophy of cortical and subcortical structures, affecting hippocampus and consequent cognitive impairment in the AD patients (Deture and Dickson, 2019). While PD exhibits α-synuclein containing cytoplasmic inclusions called Lewy bodies in the brain resulting in the selective degeneration of dopaminergic neurons in the *substantia nigra pars compacta* and associated motor symptoms in PD (Mhyre et al., 2012). Besides, HD is characterized by the deposition of poly glutamine expansion containing mutant huntingtin protein. That results in the degeneration and atrophy of the striatum, cerebral cortex and other subcortical regions of HD brain with the advancement of disease (Sari, 2011). Additionally, ALS also exhibit inclusion bodies containing SOD1, FUS, Optineurin, and TDP-43, etc., and trigger degeneration in the motor cortex, brain stem and central spinal cord (Ragagnin et al., 2019). On the other hand, MS is an inflammatory demyelinating disease, where auto-antigens and inflammatory agents trigger inflammation, neurodegeneration and tissue damage (Peterson and Fujinami, 2007). These anomalies trigger multiple pathological events, including mitochondrial dysfunction, ER stress, oxidative stress, and inflammation in vulnerable neurons. Importantly, misfolded protein degradation is associated with the principal action of ubiquitin-proteasome system involving the explicit set of ubiquitination enzymes (E1, E2, E3 and DUBs) especially ubiquitin E3 ligase for targeting a specific protein for degradation. Therefore, recent advancements in ubiquitination research have shifted its paradigm from being a prime suspect for neurodegenerative pathophysiology to an attractive therapeutic target for clearing disease-related proteins.

3.2.1. Alzheimer's disease: clearance of Aβ, AβPP, Tau and α-synuclein

Alzheimer's disease (AD) acknowledged as the world's leading cause of dementia, is characterized by extracellular amyloid-β (Aβ) and intracellular tau deposits as senile plaque and neurofibrillary tangle in the brain respectively. Accumulating evidence signifies the alteration in proteostasis network, i.e. impaired proteasome as a salient feature in AD, involving ER stress, oxidative stress, mitochondrial dysfunction, and inflammation mediated neurodegeneration (Fig. 3A) (Keller et al., 2000a). For instance, various studies have shown oligomeric Aβ deposition in the lumen of ER, in effect, of AβPP mutation along with elevated ER stress markers (eIF2α, GRP78) and ubiquitin E3 ligase-HRD1 (Gerakis and Hetz, 2018; Umeda et al., 2011). Moreover, ER stress triggered Aβ neurotoxicity by E2-25K/Hip-2 mediated activation and stabilization of caspase-12 (Fonseca et al., 2014). Likewise, Aβ also induced ER stress mediated JNK3 activation and consequent AβPP phosphorylation and endocytosis for its proteolytic processing to form amyloid beta (Yoon et al., 2012). Similarly, a reduction in MGRN1 E3 ligase levels in hippocampal neurons alter AβPP's trafficking along the secretory pathway and leads to an increased Aβ40 and Aβ42 release (Benvegnù et al., 2017b). Importantly, Keller et al., has demonstrated the role of multicatalytic proteasome (MCP) inhibition in cellular dysfunction and associated neurodegeneration in PC6 cells and Fisher 344 rats to be sensitive to the oxidative stress (Keller et al., 2000b, 2000c). Moreover, UCH-L1, a deubiquitylase enzyme is also reported to be oxidatively modified and dysfunctional in AD brain by the Butterfield group and others (Castegna et al., 2002; Choi et al., 2004; Sultana et al., 2006). The oxidative modifications of UCH-L1 at cysteine (Cys^{152}) and methionine ($Met^{1, 6, 12,124,179}$) residues are responsible for the tau-immunoreactive tangles and Aβ overproduction in AD and PD brains (Choi et al., 2004; Liu et al., 2015). In addition, Guglielmotto et al.,

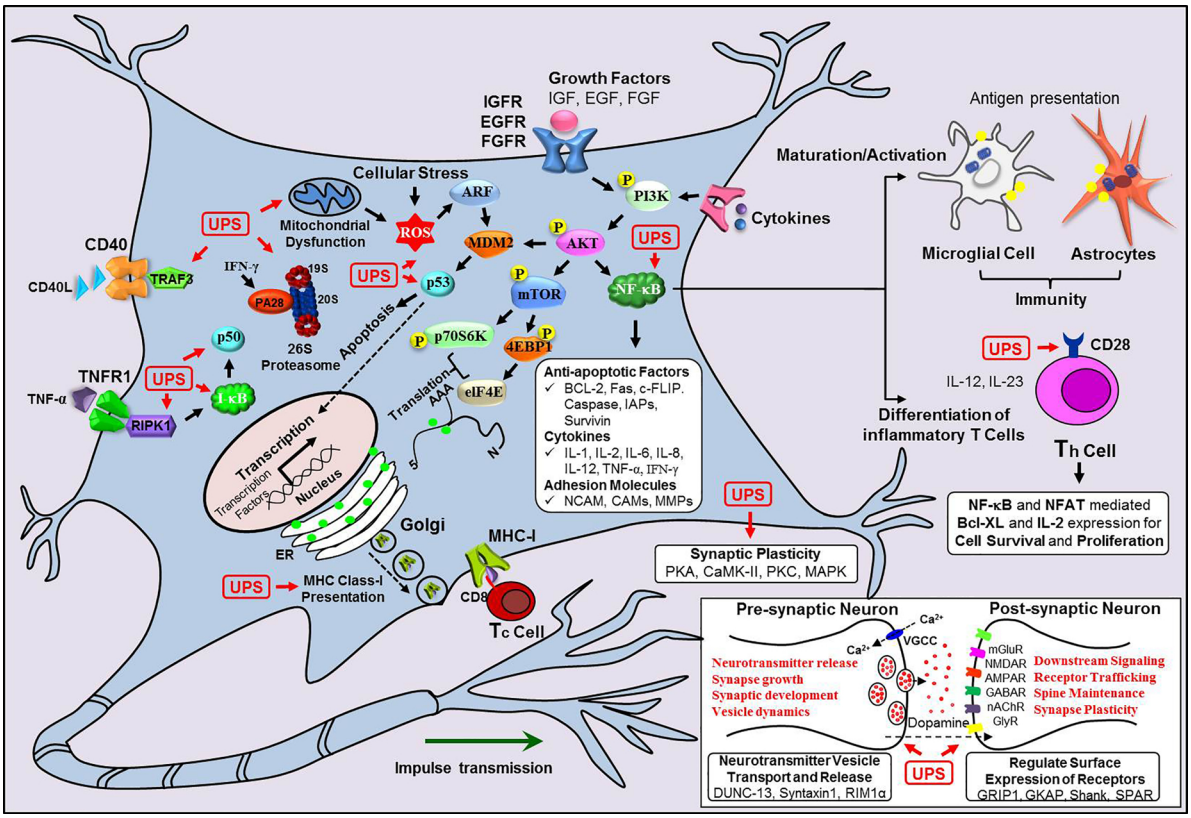


Fig. 2. Role of Ubiquitin Proteasome System in the Brain signaling, Synaptic regulation and Cellular homeostasis: UPS is involved in multifaceted functions, including MHC Class-I antigen presentation, Cellular immunity, T-cell receptor and CD28-signaling, transcription and translation process, and synaptic transmission within the brain. Moreover, it governs mitochondrial trafficking; ROS mediated signaling, and p53 linked apoptotic regulation; NF-kB mediated inflammatory responses (release of anti-apoptotic factors, cytokines and adhesion molecules), diverse receptor's expressions (CD40, TNFR1, Growth factor receptors) to maintain cellular-homeostasis. It also activates microglia and astrocytes to trigger immunoproteasome linked antigen presentation and subsequent T-cells response to generate various cytokines and associated immune defense under cellular stress. Additionally, UPS governs the synaptic transmission by regulating neurotransmitter release, Synapse growth, Synaptic development, Vesicle dynamics in the Pre-synaptic neurons while receptor trafficking, Downstream signaling spine maintenance, and Synaptic plasticity in Post-synaptic neurons. IGF-Insulin growth factor; EGF-Epidermal growth factor; FGF-Fibroblast growth factor; IGFR-Insulin growth factor receptor; EGFR-Epidermal growth factor receptor; FGFR-Fibroblast growth factor receptor; CD8-Cluster of differentiation 8; CD40-Cluster of differentiation 40; TNFα-Tumor necrosis factor alpha; TNFR1-Tumor necrosis factor receptor 1; MHC- Major histocompatibility factor; ER-Endoplasmic reticulum; TRAF3-TNF Receptor Associated Factor 3; RIPK1-Receptor-interacting serine/threonine-protein kinase 1; I-κB-Inhibitor of kappa B; NF-κB-Nuclear Factor kappa beta; IFNγ-Interferon gamma; PA28-Proteasome activator 28; ARF-ADP ribosylation factors; MDM2-Mouse double minute 2 homolog; PI3K-Phosphatidylinositol-4,5-bisphosphate 3-kinase; AKT-Protein kinase B; mTOR-mammalian target of rapamycin; eIF4E-Eukaryotic translation initiation factor 4E; p70S6K-Ribosomal protein S6 kinase beta-1; IAPs-Inhibitor of apoptosis; IL-Interleukins; NCAM-Neural cell adhesion molecule; CAMs-Cell adhesion molecules; MMPs-Matrix metalloproteinases; NFAT-Nuclear factor of activated T-cells; PKA/PKC/CaMK-II-Protein kinases; VGCC-Voltage gated calcium channel; DUNC-13- Drosophila homolog UNC-13; RIM1α-Regulating synaptic membrane exocytosis 1 alpha; GRIP1-Glutamate Receptor Interacting Protein 1; GKAP-Guanylate kinase-associated protein; SPAR-Spine-associated Rap-Gap; mGluR-Metabotropic glutamate receptor; NMDAR-N-methyl-D-aspartate receptor; AMPAR-α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor; GABAR-Gamma-Aminobutyric acid; nAChR-Nicotinic acetylcholine receptor; GlyR-Glycine receptor.

reported UCH-L1 inhibition-based induction of BACE1 up-regulation and consequent neuronal death in control and 5xFAD transgenic AD mouse signifying Aβ peptide accumulation (Guglielmotto et al., 2017). On contrary UCH-L1 overexpression have demonstrated BACE1 degradation and affected APP processing and Aβ production in both *in-vitro* and *in-vivo* studies thereby enhancing synaptic function and contextual memory (Gong et al., 2006; Zhang et al., 2014, 2012). However, Aβ inactivates ubiquitin proteasome system and subsequent impairment in multivesicular bodies (MVB) sorting pathway (Bustamante et al., 2018), nevertheless, UPS also play a decisive role in Aβ clearance. For instance, mutant ubiquitin and ubiquitin-like ubiquilin-1 proteins are linked with Aβ accumulation while up regulated proteasomal degradation of BACE1 and γ-secretase components leads to decrease in Aβ accumulation (Hong et al., 2014; Gong et al., 2010). Another E3 ligase-substrate adaptor protein- FBXO2 expedited the clearance of AβPP and regulate synaptic connections, which otherwise reduced in AD (Atkin et al., 2014). Similarly, over expression of other E3 ligases- Parkin and CHIP along with E2 conjugating enzyme- UbcH5B and heat shock proteins- Hsp70/90 curtailed AβPP expression, Aβ burden, hyper phosphorylated tau and inflammation in the AD mouse model (Gadhave et al., 2016; Sweeney et al., 2017). Moreover, multiple E3 ligase, including Parkin, SIAH1, MUL1, ARIH1, and SMURF1 is reported to regulate mitophagy by ubiquitinating outer mitochondrial membrane proteins (Jacoupy et al., 2019; Reddy and Oliver, 2019). Likewise, MDM2 and NUB1 facilitated tau degradation in the brain by disrupting GSK-3β's interaction with p53 and tau respectively (Proctor and Gray, 2010; Richet et al., 2012). Furthermore, inhibition of proteasomal ubiquitylase Usp14, UCHL1 and intrusion of proteasome activator Blm10 have also enhanced tau degradation studies (Boselli et al., 2017; Tramutola et al., 2016). On contrary very few studies have reported α-synuclein ubiquitination in AD (Mukaetova-Ladinska et al., 2000; Nemes et al., 2004) although numerous studies have shown its aggregation along with amyloid plaques and neurofibrillary tangles in AD (Twohig and Nielsen, 2019). Collectively, the above results depicted the role of UPS in both early and late-stage AD with characteristic synaptic dysfunction and neurodegeneration.

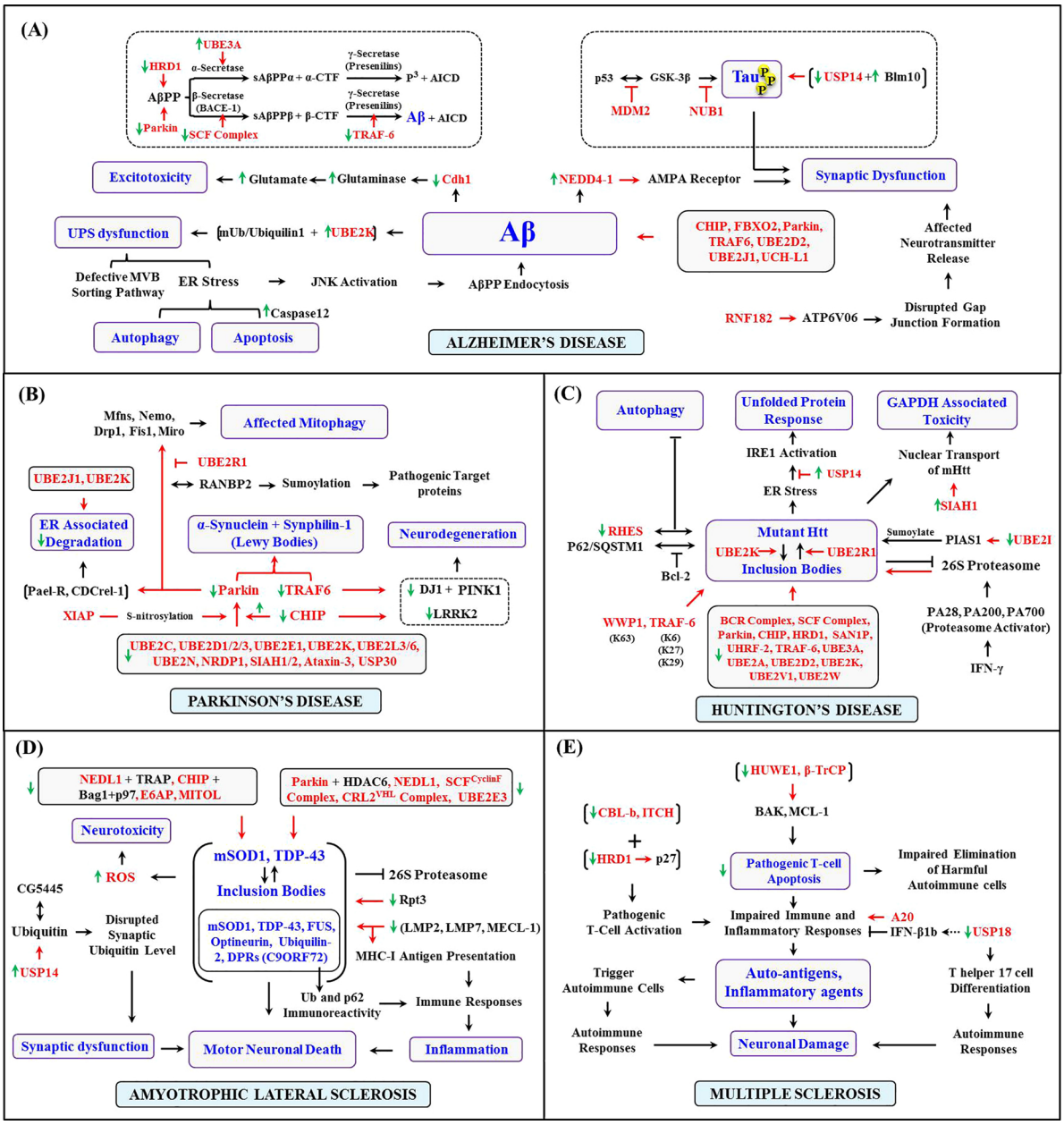


Fig. 3. Ubiquitination mediated neurodegeneration mechanism in neurodegenerative disorders: (A) Alzheimer's Disease- The AβPP is proteolytically processed to generate non-amyloid (p3) and amyloid (Aβ) peptide with help of α- and β-secretase respectively, followed by γ-secretase. Under pathogenic state, E3 ligases HRD1 and Parkin failed to regulate the level of AβPP and give rise to its amyloidogenic products. Similarly, down-regulated SCF complex, TRAF-6 and up regulated UBE3A triggered the selection of amyloidogenic pathways for Aβ synthesis. Moreover, disrupted Cdh1 activity, lead to glutaminase overload associated excited-toxicity in neurons. Furthermore, Neurotransmitter release and synaptic functions are compromised due to impaired AMPA signaling, and gap junction formation due to NEDD4-1 and RNF182 mediated proteasomal degradation of AMPA receptors and ATP6V06 (Endosomal proton pump member). Likewise, multi vesicular body-sorting pathway for lysosomal degradation is affected due to UBE2K mediated mutant ubiquitin or ubiquitin like ubiquilin-1 conjugation. In addition, ER-stress trigger JNK activation and induces endocytosis of AβPP to produce Aβ peptides. Consequently, it governs autophagy and caspase-12 mediated apoptosis as a facultative response against Aβ toxicity. Several other UPS components like CHIP, FBXO2, Parkin, TRAF6, UBE2D2, UBE2J1 and UCH-L1 have reported to clear Aβ that get affected in the diseased state. Like Aβ, Tau protein is marked by down regulated USP14 and elevated proteasome activator Blm10 for proteasomal degradation. Moreover, NUB1 prevents hyper-phosphorylation of Tau by Glycogen synthase kinase-3beta (GSK-3β) to alleviate its toxicity, while MDM2 mediates inhibition of interaction between GSK-3β and p53 that enables p53 to trigger transcriptional activation of protective genes to clear off hyper phosphorylated tau. These UPS components get disrupted in diseased state and caused hyper-phosphorylated tau mediated neurotoxicity due to the neuro-fibrillary tangles. (B) Parkinson's Disease- The Lewy body (α-synuclein and Synphilin-1) pathology is mainly governed by defective E3 ligase Parkin. It regulates mitochondrial membrane proteins (Mfns, Nemo, Drp1, Fis1, and Miro) and is affected in the diseased state that triggers impaired mitophagy. Likewise, defective parkin disrupted endoplasmic reticulum associated degradation (ERAD) by affected Pael-R and CDCrel-1 ubiquitination. It also triggers SUMOylation via interacting with RANBP2 for the clearance of pathogenic target proteins. Similarly, CHIP and TRAF6 are associated with the ubiquitination of LRRK2 and PINK1/DJ1 respectively that get affected in the disease state and cause neurodegeneration. (C) Huntington's Disease- The mutant huntingtin forms inclusion bodies- both soluble and insoluble forms. Here, Proteasome activators PA28, PA200 and PA700 are helpful in reversing the inhibition of proteasome by mHtt and inclusion bodies. However, it governs ER stress mediated unfolded protein response and induces GAPDH associated toxicity through its nuclear transport by SIAH1. Many UPS enzymes act differently i.e. positively and negatively to regulate the clearance of mHtt, including WWP1 and TRAF6 that direct K63 and K6, K27, K29 poly-ubiquitination respectively, and retard proteasomal degradation under diseased state. (D) Amyotrophic Lateral Sclerosis- Inclusion bodies in ALS is populated with mutant SOD1, TDP-43, FUS, Optineurin, Ubiquilin2 and dipeptide repeat proteins that affect the different cellular process in neurons. Impaired proteasome and MHC-I mediated immunity and inflammation triggers the

neuronal death in the diseased state. Moreover, CG5445 and USP14 scavenge the ubiquitin protein and thereby interfere with the synaptic functions in ALS. Furthermore, mitochondrial dysfunction mediated ROS production also triggers inflammation and associated neurotoxicity in ALS. (E) Multiple Sclerosis- Auto-antigens and inflammatory agents are the principal cause for the progression of MS. Where depleted expression of certain E3s, including HUWE1, β-TrCP, CBL-b, ITCH and HRD1 lead to compromised pathogenic T-cell apoptosis and its activation to aggravate the symptoms of MS. However, IFN-β1b treatment enabled the suppression of immune and inflammatory responses by targeting auto-antigens and inflammatory agents who prevented neuronal damage in Multiple sclerosis. Dashed arrow- Regulation; Red arrow- Ubiquitination and degradation; Black arrow- Activation; Up arrow (Green) - Elevated expression; Down arrow (Green) - Reduced expression.

3.2.2. Parkinson's disease: Clearance of α-synuclein in Lewy bodies

Parkinson's disease (PD) is one of the prevalent neurodegenerative disorders characterized by progressive loss of muscular coordination and control with primary α-synuclein deposits in Lewy bodies. The brains of PD patients revealed cellular stress, including oxidative stress, ER stress and mitochondrial dysfunction to be the integral part for impairment in proteostasis and dopaminergic neuronal loss (Fig. 3B). In fact, several lines of evidence have presented an intimate relation between UPS aberrations and PD pathogenesis. For instance, deletion of PSMC1/Rpt2 (ATPase Subunit) and variations in PSMC4/Rpt3 proteasomal subunits have exhibited ubiquitin and α-synuclein positive neurons, thereby suggesting the role of UPS in α-synuclein degradation (Vilchez et al., 2014). Similarly, aberrations in E3 ligase Parkin and ubiquitylase UCH-L1 have been widely reported to cause early-onset familial forms of PD with characteristic symptoms of rigidity, tremor, and bradykinesia (Mabb and Ehlers, 2010). Since, mutant Parkin failed to clear toxic accumulation of their substrates, including, Pael-R and CDCrel-1 and contribute towards ER stress associated neurodegeneration (Dawson and Dawson, 2010). Likewise, Parkin in association with peptidases like PINK1 and DJ-1 is famous for the regulation of mitochondrial dynamics, which otherwise lead to oxidative stress, lipid peroxidation and consequent toxicity in neurons (Cenini et al., 2019). Recent studies have revealed that PINK1 activates Parkin in response to mitochondrial depolarization by phosphorylating ubiquitin at Ser65 that recruits autophagy receptors to clear damaged mitochondria via mitophagy (Chin and Li, 2016). Altogether, these proteins and SIAH-1/2 are known to facilitate proteasomal degradation of parkin's substrate synphillin-1 and thus prevent inclusion formation in association with α-synuclein (Szargel et al., 2008). Another Parkin's substrate, PARIS (ZNF746) reported to suppress PGC-1α expression and consequent neurodegeneration in PD signifying another mechanism of Parkin associated neuro protection in PD (Shin et al., 2011). Interestingly, USP13 mediated Parkin's ubiquitination may facilitate its stability via interaction with Beclin-1 and contribute towards the clearance of α-Synuclein (Liu et al., 2019; Lonskaya et al., 2013). Furthermore, a ubiquitylase UCHL1 is also evident to function like E3 ligase that governed K63-linked ubiquitination of alpha-synuclein and promotes its clearance, while its mutation leads to α-Synuclein accumulation (Cartier et al., 2012). Several other ubiquitin-linked proteins such as HRD1, MDM2 and CHIP governed an important role in maintaining protein homeostasis to prevent neurotoxicity in PD. For instance, HRD1 is associated with degradation of ER linked proteins like Pael-R, while MDM2 regulates GSK3B levels and CHIP activates Parkin to mitigate ER stress, proteotoxicity, and mitochondrial stress in PD (Kaneko, 2016; Tetzlaff et al., 2008). On the other side, mutant α-synuclein also found to curtail proteolytic activity of UPS via direct interaction with its 20S core proteolytic unit and attain filament structure. Moreover, recent study has identified CDK5 mediated phosphorylation dependent proteasomal degradation of E3 ligase GP78 triggering neural death in PD (Wang et al., 2018b). Based on these evidences, there is a direct relation between abnormal synuclein aggregation and dopaminergic neuronal death in PD.

3.2.3. Huntington's disease: Clearance of mHtt and inclusion bodies

Huntington's disease (HD) is a poly glutamine expansion disorder of huntingtin protein causing progressive neurodegeneration. Its pathological hallmark includes the mutant huntingtin (mHtt) aggregates with positive inclusions of the ubiquitin proteasome system. In fact, different studies have reported both aggregated as well as soluble forms of polyubiquitinated mHtt in HD transgenic R6/2 mice model, HD patient material, and post-mortem brains (Chang et al., 2015; Juenemann et al., 2015). In spite of mHtt ubiquitination, proteasomes are incapable of cleaving poly glutamine tract in these proteins thereby leading to its aggregation in neurons (Fig. 3C). For instance, in a study mammalian 26S proteasome was found to cleave mHtt within flanking sequences or after first glutamine of a polyQ peptide, but unable to digest polyQ repeats that obstructed their passage through proteasomes and caused its impairment and aggregation in neurons (Juenemann et al., 2013). Moreover, FRET experiments have defined mHtt's interaction with LMP2 subunit of proteasome, thereby impairing their chymotrypsin and caspase like activities leading to compromised peptide cleavage in HD patients (Holmberg et al., 2004). It has previously observed that only oligomeric or soluble forms of mHtt triggered proteasomal impairment while inclusion bodies failed to do so (Díaz-Hernández et al., 2006). Scientists have reported improvement in mHtt clearance by altering proteasomal activities. For instance, proteasome activators such as PA28, PA200, PA700 and 20S catalytic immune-subunits- LMP2 (PSMB9), LMP7 (PSMB8), and Mecl-1 (PSMB10) facilitated polyQ cleavage and clearance due to IFN-γ stimulation (Heink et al., 2005; Schipper-Krom et al., 2012). Interestingly, optical pulse chase experiments have demonstrated selective compartmentalized degradation of mHtt i.e. faster removal in the cytoplasm of the cell body than in neuronal processes (Zhao et al., 2016). Apart from proteasome alteration, E3 ligases have demonstrated to clear mHtt through ubiquitination. For example, nuclear ubiquitin ligases SAN1P and UHRF-2 have facilitated the degradation of mHtt aggregates (Iwata et al., 2009). Likewise, HRD1, Parkin and CHIP also decreased inclusion bodies through mHtt ubiquitination (Harding and Tong, 2018). Similarly, Bcl-2 along with Parkin and VCP found to decrease the affinity of p62/SQSTM1 to polyubiquitinated proteins and consequent reduction in mHtt aggregates (Zhou et al., 2015). Another study revealed K48 mediated ubiquitination of Htt fragments in UBE3A dependent manner (Bhat et al., 2014). Interestingly, other atypical ubiquitination (K6, K27, K29) by TRAF6 and K63 by WWP1 of mHtt found to inhibit proteasome mediated mHtt degradation and subsequent aggregation in HD (Nath and Lieberman, 2017; Zucchelli et al., 2011). Where as an inflammatory kinase like IKK found to promote mHtt degradation through proteasomal and lysosomal pathways (Thompson et al., 2009). Likewise, some E2s conjugating enzymes, including, hE2-25k, CDC34, and UBE2W are reported to regulate the ubiquitination and solubility of mHtt aggregate in different HD models (Mitra and Finkbeiner, 2008; Wang et al., 2018a). Interestingly, recent fundings about ubiquitination of mHtt containing inclusion bodies suggested it a dynamic process involving the recruitment of (de)ubiquitination proteins continuously to trigger ubiquitin-specific clearance pathways in HD (Juenemann et al., 2018).

3.2.4. Amyotrophic lateral sclerosis: Clearance of SOD1, TDP-43 and inclusion bodies

ALS is a rare motor-neurodegenerative disorder affecting brain and spinal cord that triggers progressive paralysis with troubles in chewing, talking, and walking. Recent studies have reported UPS associated impaired turnover of several proteins, including SOD1, FUS, Optineurin, TDP-43, and translation product of intronic hexanucleotide

Table 2
Mechanistic role of ubiquitin.(ation enzymes for the clearance of toxic protein aggregates in neurodegenerative diseases (Fuchsberger et al., 2017; Joshi et al., 2016; Saito et al., 2014; Rodrigues et al., 2016; Rosen et al., 2010; Liu et al., 2008; Gong et al., 2010; Geetha et al., 2012; Singh et al., 2017; Zhao et al., 2016; Morgan et al., 2007; Hong et al., 2014; Burr et al., 2011; Guglielmotto et al., 2017; Ko et al., 2009; Allen et al., 2005; Zhong et al., 2005; Dawson and Dawson, 2010; Rott et al., 2017; Xu et al., 2015; Murata et al., 2013; Mulherkar et al., 2009; Tsang et al., 2009; Haddad et al., 2013; Fiesel et al., 2014; Geisler et al., 2014; Araki and Nagata, 2011; Fiesel et al., 2014; Durcan et al., 2012; Ristic et al., 2014; Gersch et al., 2017; Alexopoulou et al., 2016; Lu et al., 2013; Miller et al., 2005; Mealer et al., 2014; Bhutani et al., 2012; Bae et al., 2006; Zucchelli et al., 2011; Maheshwari et al., 2014; Howard et al., 2007; Ochaba et al., 2016; de Pril et al., 2007; Wang et al., 2018a; Hyrskyluoto et al., 2014; Uchida et al., 2016; Allen et al., 2005; Miyazaki et al., 2004; Hans et al., 2014; Vaden et al., 2015; Xu et al., 2016; Mandel et al., 2012; Sellebjerg et al., 2012; Honke et al., 2016; Malhotra et al., 2013).

NDDs	UPS Components	Molecular Functions	Mechanism of Action	References
Alzheimer's Disease (Aβ, Tau)	**APC/Cdh1**	Regulate glutamate level to control Aβ linked excitotoxicity in neurons	The ubiquitin E3 ligases such as SCF complex, TRAF6, and HRD1 regulate proteasomal degradation of β-secretase, Presenilin (γ-secretase) and AβPP respectively, thereby controlling the level of Aβ in the brain. Other E3s, including Parkin and CHIP directly clear Aβ in the brain. While downregulated UBE3A inhibit Aβ formation by preventing α-secretase degradation. On contrary Aβ also inhibit few E3s, including Cdh1 and NEDD4-1 thereby affecting membrane receptors, synaptic density and neurotransmission. In addition, multiple E2s, including UBE2A, UBE2D1, UBE2K, and deubiquitinases- UBE2J1 and UCH-L1 are associated with the conjugation and recycling of ubiquitin to regulate turnover of various synaptic proteins. However, in diseased state alterations in these E2s, E3s and DUBs affect gap junctions, membrane receptors, synaptic protein densities, and associated synaptic transmission and neurological behaviors in AD patients.	(Fuchsberger et al., 2017)
	CHIP	Ubiquitinates phosphorylated Tau and also promote Aβ degradation		(Joshi et al., 2016)
	HRD1	ER-associated degradation of misfolded MHC-I heavy chains, Pael-R and AβPP		(Saito et al., 2014)
	NEDD4-1	Aβ triggered ubiquitination/ degradation of AMPARs to regulate synaptic function, strength and dendritic spine density		(Rodrigues et al., 2016)
	Parkin	Reduces Aβ levels and improve long term potentiation and behavioral abnormalities		(Rosen et al., 2010)
	RNF182	Ubiquitinate ATP6V06 protein and disrupt gap junction formation and neurotransmitter release in the diseased state		(Liu et al., 2008)
	SCF Complex (SKP1, CUL1, Fbx2)	Degrade BACE1 enzyme that attenuate amyloidosis and improves synaptic function		(Gong et al., 2010)
	TRAF6	Abrogate Aβ-mediated inhibition of p75 neurotrophin receptor's ubiquitination and restore neuronal cell survival. Also degrade mutated Presenilin		(Geetha et al., 2012)
	UBE3A	Ubiquitination of α-secretase (ADAM10) and its compromised action reduces Aβ level through non-amyloidogenic pathway		(Singh et al., 2017)
	UBE2A	Coordinates amyloid peptide's degradation through proteolysis		(Zhao et al., 2016)
	UBE2D1	Genetically associated with the late onset Alzheimer's susceptible gene APOE		(Morgan et al., 2007)
	UBE2K (E2-25K/Hip2)	Upstream regulator of ASK1 and JNK that manage Aβ neurotoxicity. Also, colocalize with mutant ubiquitin (UBB^{+1}) to inhibit proteasome activity resulting in Presenilin accumulation		(Hong et al., 2014)
	UBE2J1,	Increase cellular levels of mono-ubiquitin and consequent protein turnover by UPS that improves contextual memory, and restores synaptic function in brain		(Burr et al., 2011; Guglielmotto et al., 2017)
	UCH-L1			
Parkinson's Disease (Lewy bodies; α-synuclein, Synphilin1)	**CHIP**	Enhance parkin's activity and ubiquitinates α-synuclein and LRRK2 for degradation	Multiple E3 ligases, including Parkin, TRAF6, SIAH1, UBE3A, CHIP and DJ1 directly or indirectly ubiquitinate and clear α-synuclein from the cell. Among them, Parkin is the key E3 ligase centrally regulating the clearance of alpha synuclein with help of other E3 ligases- NRDP1, XIAP, CHIP, and DJ1, where NRDP1 and XIAP1 stabilize Parkin, while CHIP enhances Parkin's activity. Moreover,	(Ko et al., 2009)
	Gigaxonin	Regulate degradation of microtubule associated protein 1B and intermediator filaments to maintain neuronal health		(Allen et al., 2005)
	NRDP1	Regulate Parkin's stability and activity via regulating its proteasomal degradation		(Zhong et al., 2005)
	Parkin	Ubiquitinate and clear α-synuclein aggregates and various mitochondrial membrane proteins to regulate mitophagy		(Dawson and Dawson, 2010)
	RANBP2	SUMOylate its pathogenic target and trigger protein degradation through Parkin interaction		(Rott et al., 2017)
	SIAH1	Ubiquitinate and degrade alpha-synuclein and synphilin-1 proteins		(Xu et al., 2015)
	TRAF6	Promotes ubiquitination of α-synuclein, DJ1		(Murata et al.,

(*continued on next page*)

Table 2 (*continued*)

		and PINK1	Parkin also regulates the ubiquitination of other E3 ligase- RANBP2 and its SUMOylated mitochondrial membrane proteins to regulate mitophagy. While Gigaxonin ubiquitinates MAP1B and regulate the stability of intermediators filaments. Altogether, these E3 ligases along with a diverse set of conjugating enzymes, including UBE2A, UBE2C, UBE2D, UBE2E, UBE2J, UBE2K, UBE2L, UBE2N, UBE2R and various deubiquitinases like Ataxin3, UCHL1, USP8, USP9X, USP30 regulate membrane trafficking, synaptic transmission and microtubular organization to control neuro-muscular functions in PD	2013)
	UBE3A	Colocalize with alpha synuclein to clear it in a proteasome-dependent manner		(Mulherkar et al., 2009)
	XIAP	Trigger S-nitrosylation of Parkin and thereby affect neuronal survival in PD		(Tsang et al., 2009)
	UBE2A	Associate with parkin and regulate the clearance of dysfunctional mitochondria in mice and human cells		(Haddad et al., 2013)
	UBE2C, UBE2D family members, UBE2-E1 /K /L3 /L6 /N	Contribute towards mitophagy by redundantly charging RING-HECT hybrid ligase Parkin with ubiquitin		(Fiesel et al., 2014; Geisler et al., 2014)
	UBE2K, UBE2J1	Ubiquitinate and selectively degradation misfolded membrane proteins from the ER		(Araki and Nagata, 2011)
	UBE2R1	Negative regulator of Parkin that restrict parkin translocation to mitochondria		(Fiesel et al., 2014)
	Ataxin-3	Coupled with parkin to govern ubiquitination and decomposition of polyubiquitin chain		(Durcan et al., 2012)
	UCH-L1, USP9X	Mutated form lead to impairment in abnormal protein clearance		(Ristic et al., 2014)
	USP30	Associated with PINK1 and Parkin to clear damaged mitochondria in PD		(Gersch et al., 2017)
	USP8	Associated with the clearance of α-synuclein and modify its toxicity in PD		(Alexopoulou et al., 2016)
Huntington's Disease (PolyQ, mHtt)	**BCR Complex (BTB-CUL3 - RBX1)**	Associated with the proteasomal degradation of mHtt in association with NUB1-NEDD8-CUL3 pathway	Diverse E3s act differently to regulate the cytotoxicity of mHtt aggregates. For instance, BCR complex, SCF complex, TRAF6 and CHIP regulate mHtt levels via ubiquitination and proteasomal degradation. Besides, SIAH1 regulate nuclear transport associated toxicity of mHtt. In addition, UBE3A regulate dopamine regulated phosphoprotein (DARPP-32) and directs neurotransmission. Similarly, RHES binds Beclin1 to regulate autophagy process and overcome the cytotoxicity caused by the protein aggregates. Collectively these E3s with the assistance of multiple E2s and DUBs- UBE2A, UBE2D2, UBE2K, UBE2V1, UBE2W, UBE2I, and USP14 effectively remove the cellular aggregates of mHtt to relieve the cytotoxicity in neurons.	(Lu et al., 2013)
	CHIP	suppression of aggregation and toxicity of PolyQ huntingtin		(Miller et al., 2005)
	RHES	Binds autophagy regulator Beclin-1 to activate autophagy, while its co-expression with mHtt blocks Rhes-induced autophagy		(Mealer et al., 2014)
	SCF Complex (Skp1-Cul1-F-box)	Deficient Cul1 and Skp1 are associated with increased mHtt levels in brain		(Bhutani et al., 2012)
	SIAH1	Enhance nuclear translocation of mHtt and cytotoxicity in association with GAPDH		(Bae et al., 2006)
	TRAF6	Associated with ubiquitination and degradation of mHtt and PolyQ proteins		(Zucchelli et al., 2011)
	UBE3A	Its deficiency trigger protein overload and decrease in dopamine regulated phosphoprotein DARPP-32, and play crucial role in HD progression		(Maheshwari et al., 2014)
	UBE2A, UBE2D2, UBE2V1	Ubiquitinates poly-glutamine aggregates and controls the size and number of protein aggregates		(Howard et al., 2007)
	UBE2I	Conjugate SUMO protein to E3 SUMO ligase- PIAS1, which regulates SUMOylation of huntingtin protein		(Ochaba et al., 2016)
	UBE2K (E2-25/Hip2)	Interacts directly with huntingtin and may mediate ubiquitination of neuronal intranuclear inclusions in HD		(de Pril et al., 2007)
	UBE2W	Markedly decreases mHTT aggregate formation and increases the level of soluble monomers		(B. Wang et al., 2018)
	USP14	Effectively reduce cellular aggregates and prevent cellular degeneration by counteracting ER stress-mediated IRE1 activation		(Hyrskyluoto et al., 2014)

(*continued on next page*)

Table 2 (*continued*)

Amyotrophic Lateral Sclerosis (TDP43, mSOD1, TRAP, Dvl1 and other insoluble protein aggregates)	**$CRL2^{VHL}$ Complex (Cul2-Rbx1-EloBC-VHL)**	Role in the ubiquitination and proteasomal degradation of misfolded TDP-43	The multiple protein aggregates, including TDP-43, mSOD1, TRAP, Dvl1 and other insoluble proteins are prominent in ALS. Multiple E3 ligases such as $CRL2^{VHL}$ complex, NEDL1 and Gigaxonin are reported to trigger ubiquitination and proteasomal degradation of TDP-43, mSOD1 and MAP1B proteins. Altogether, these E3s along with E2s- UBE2E3/ UBCM2 and deubiquitinase USP14 regulate ER-associated degradation, intermediary filaments, neuromuscular junctions and synaptic transmission to regulate motor functions in ALS.	(Uchida et al., 2016)
	Gigaxonin	Regulate intermediary filament by ubiquitination and degradation of microtubule associated protein-1B		(Allen et al., 2005)
	NEDL1	Associated with the ubiquitination and degradation of mutant SOD1, translocon-associated protein-delta (TRAP), and Dishevelled-1 (Dvl1)		(Miyazaki et al., 2004)
	UBE2E3/ UBCM2	Enhance the ubiquitination of TDP-43 and controls its solubility and associated neurodegeneration		(Hans et al., 2014)
	USP14	Maintain synaptic ubiquitin level for proper synaptic functions and development of neuromuscular junctions		(Vaden et al., 2015)
Multiple Sclerosis (Soluble oligomers, auto-antigens and auto-inflammatory agents)	**HRD1**	Facilitates T-cell proliferation by the destruction of cyclin-dependent kinase inhibitor $p27^{kip1}$ and thus aids in T-cell mediated immunity	Multiple E3s regulate cellular functions such as HRD1 ubiquitinate p27 (a CDK inhibitor) and elevate T-cell expression. Similarly, HUWE1 and FBXW1A regulate the expression of anti- and pro-apoptotic proteins MCL1 and BAK respectively that govern T-cell survival. Additionally, ITCH and CBLB also regulate T cell antigen surface expression. Collectively these UPS enzymes regulate T-cell expression, neuro-inflammation, and autoimmune responses to protect against myelin damage in MS.	(Xu et al., 2016)
	HUWE1	Regulate the levels of anti-apoptotic protein MCL-1 and pro-apoptotic BAK, which get disrupted and thus enhance the survival of pathogenic T-cells		(Mandel et al., 2012)
	FBXW1A (β-TrCP)			
	ITCH	Maintain cell surface expression of cytotoxic T lymphocyte antigen 4 (CTLA-4) on $CD4^{+}CD25^{high}$ T cells thereby regulating T cell function in MS		(Sellebjerg et al., 2012)
	CBLB			
	USP18	Acts as a negative regulator of type-I interferon signaling and thus a therapeutic response to IFN-β in MS pathology. Also necessary for T helper-17 cell differentiation and autoimmune responses		(Honke et al., 2016; Malhotra et al., 2013)

Ubiquitin E3 Ligase- Red color; **Ubiquitin E2 Conjugating Enzyme- Green color**; **Deubiquitinase Enzyme- Blue color**

Ubiquitin E3 Ligase- Red color; Ubiquitin E2 Conjugating Enzyme- Green color; Deubiquitinase Enzyme- Blue color.

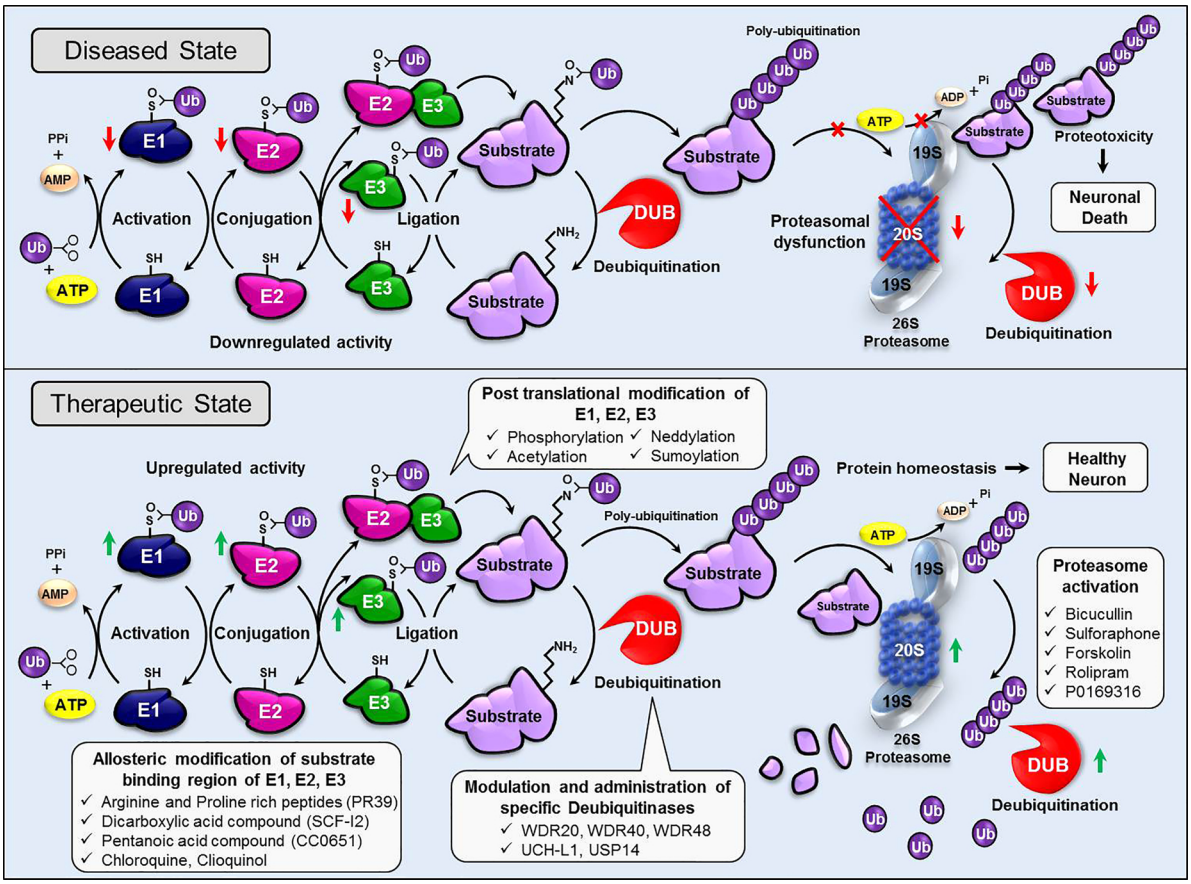


Fig. 4. Therapeutic approaches to alleviate neurodegenerative pathologies: Under diseased conditions the activity of ubiquitinating enzymes (E1s, E2s, and E3s), deubiquitinase enzymes (DUBs), proteasome assembly and functions are compromised leading to proteotoxic stress in the neurons causing its death. The disease pathologies might be reversed upon adapting the following remedial strategies- i) Allosteric modification of the substrate binding region, which could increase or decrease the activity of E1, E2 and E3 enzymes in order to regulate the clearance of toxic, misfolded or unwanted proteins. ii) Post translational modification of E1, E2, and E3 enzymes lead to the protection of these enzymes from their self-ubiquitination and degradation thereby enabling them to be active during diseased conditions for elevated clearance of harmful proteins. iii) Modulation and administration of specific deubiquitinases could aid in overcoming the burden of native or mutant ubiquitin charged enzymes and substrates, facilitating the UPS machinery for the efficient clearance; iv) Proteasomal activation aids in the mitigation of proteotoxic burden of neurons enabling them to maintain their homeostatic state. E1-Ubiquitin E1 activating enzyme; E2-Ubiquitin E2 conjugating enzyme; E3-Ubiquitin E3 ligase; DUB-Deubiquitinase enzyme; Ub-Ubiquitin protein; ATP-Adenosine triphosphate; AMP-Adenosine monophosphate; PPi-Pyrophosphate.

repeats of C9ORF72 gene as the most susceptible cause for ALS (Fig. 3D) (Burk and Pasterkamp, 2019). For instance, conditional knockout of proteasomal subunit PSMC4/Rpt3 disclosed aggregates of TDP-43 and FUS proteins with affected locomotor function and motor neuronal loss, which were absent in autophagy related protein Atg7 knockout study, signifying the involvement of proteasome in ALS development (Tashiro et al., 2012). Moreover, inducible subunits of proteasome including, LMP2, LMP7 and MECL-1 get overexpressed early during disease pathogenesis and generates peptides crucial for major histocompatibility class-I presentation and subsequent immunity in the SOD1 mice model of ALS (Johnston-Carey et al., 2016). Likewise, alteration in E3 ligases- TRAF6 and Mahogunin ring finger 1 (MGRN1) conferred mutant SOD1 linked cytotoxic protein aggregates in the ALS mouse model (Chhangani et al., 2016; Semmler et al., 2019). Conversely, the role of UPS has also been demonstrated in the clearance of ALS pathological proteins. For example, the clearance of mutated or excessive SOD1 in ER is triggered by an E3 ligase NEDL1 with help of translocon protein TRAP, while in the cytosol by E3 ligases- E6-AP and Dorfin (Mishra et al., 2013). Another E3 ligase CHIP along with ubiquitin binding proteins- Bag1 and VCP (p97), facilitated mutant-SOD1 degradation by translocating it from 19S ATPase subunit to core 20S proteasome particle (Choi and Lee, 2010). Moreover, mitochondrial E3 ligase MITOL is found to ubiquitinate mSOD1 and mitofusin2 to regulate mitochondrial dynamics and its associated ROS production (Sugiura et al., 2013). Similarly, TDP-43 is ubiquitinated by Parkin and facilitated its translocation from the nucleus to cytosol and its accumulation in a multi-protein complex with HDAC6 (Hebron et al., 2013). Moreover, a novel E3 ligase ZNF179 is evident to ubiquitinate TDP-43 and ameliorates TDP-43 associated neuropathy in ALS (Lee et al., 2018). Furthermore, a CCNF missense mutation has identified to alter E3 ligase complex ($SCF^{Cyclin-F}$) thereby causing accumulation of abnormally ubiquitinated TDP-43 proteins (Williams et al., 2016). Likewise, HRD1 reported to regulate optineurin degradation and microtubule-dependent aggresome formation (Mao et al., 2017). Additionally, C9ORF72 gene associated mutations are responsible for dipeptide repeats in them leading to their aggregation and immunoreactivity against ubiquitin and p62 in ALS (Freibaum and Taylor, 2017; Schludi et al., 2015). Recently, a p53 family protein (p63) has been reported to augment cellular levels of E3 ligase TRIM63 (MuRF1) that regulates the degradation of muscular proteins during muscular-atrophy in ALS (von Grabowiecki et al., 2016). This way, UPS regulates protein turnover dynamics in ALS to combat neurodegenerative pathologies.

3.2.5. Multiple sclerosis: Clearance of auto-antigens and inflammatory agents

Multiple sclerosis (MS) is a chronic inflammatory progressive neurodegenerative disorder of CNS, triggering demyelination and muscular

Table 3
Potential compounds modulating UPS activity for therapeutic implications in neurodegenerative disorders.

S.No.	Compounds (PubChem CID)	Impact on UPS	Target UPS Component	Molecular Action	Target Pathology	BBB permeability	References
1	Benzamil (108107108,107)	UPS activator	_	Rescue acid sensing ion channel (ASIC) dependent acid toxicity and enhance UPS activity	HD, MS	Yes	(Durham-Lee et al., 2011; Friese et al., 2007; Wong et al., 2008)
2	Betulinic acid (64,971)	Proteasome activator	20S β5 proteasome subunit	Activates chymotrypsin-like activity of proteasome and enhance GABAergic activity, Prevent neurobehavioral activity in Alzheimer's disease	AD, Autism	Yes	(Das et al., 2016; Huang et al., 2007; Navabi et al., 2018)
3	Bicuculline (10,237)	Proteasome activator and UPS activator	Rpt6 (19S Proteasome regulatory subunit) and other UPS component	Stimulate calcium/calmodulin-dependent protein kinase II (CaMKII) mediated proteasomal activity stimulate AMPARs ubiquitination to modulate synaptic connections in neurons	AD, PD	No	(Djakovic et al., 2009; Widagdo et al., 2017)
4	Clioquinol (2788)	Proteasome inhibitor	20S Proteasome	Inhibit 20S proteasome inhibition by clioquinol-copper complex and affect its chymotrypsin-like activity. Disrupt interaction between metals and Aβ peptide in the brain and down regulates mutant Htt level in HD models	AD, HD	Yes	(Regland et al., 2001; Wang et al., 2009)
5	Curcumin (969,516)	_	APC, HSP70, HSP90	Inhibits APC leading to cell cycle arrest and apoptosis in mylenomas. Increase Hsp70/90 activity to prevent amyloid formation and subsequent neuronal death	AD, Tauopathies, HD	Yes	(Lee and Langhans, 2012; Maiti et al., 2014; Tsai et al., 2011)
6	Delta-9-Tetrahydrocannabinol (Δ9-THC) (5,321,846)	DUB Stimulator	USP3	Its antioxidant effects and increased level of USP3 exert direct neuroprotective role in cellular models of Parkinson's disease	PD	Yes	(Carroll et al., 2012; Schou et al., 1977)
7	Doxycycline (54,671,203)	E3 inhibitor	HUWE1	Reshapes α-synuclein oligomers into off-pathway, high-molecular-weight species that do not evolve into fibrils preventing its toxic effects	AD, PD	Yes	(Balducci and Forloni, 2019; González-Lizárraga et al., 2017)
8	Forskolin (47,936)	Proteasome inhibitor	26S proteasome	Enhance proteasomal degradation of protein aggregates (mutant FUS, SOD1, TDP43, Tau) in cAMP/PKA pathway dependent manner	PD, ALS	Yes	(Guo et al., 2016; Lokireddy et al., 2015)
9	IU1 (675,434)	Proteasome activator	USP14	Boost proteasome activity by ameliorating USP14 mediated suppression of protein degradation including tau	AD	Yes	(Kiprowska et al., 2017; Lee et al., 2010; Min et al., 2017)
10	Mitoxantrone (4212)	DUB inhibitor	USP11	Target ubiquitin-specific peptidase 11 and acts as immunosuppressive drug protective in multiple sclerosis	MS	No, but may cross the disrupted BBB in MS	(Cocco and Marrosu, 2014; Li et al., 2012b)
11	NC-043 (CAS#1053172-87-4)	DUB inhibitor	USP30	Antagonizes Parkin-dependent mitophagy in cells	PD	Unknown	(Yue et al., 2014)
12	Nicotine (89,594)	E2, Proteasome, DUB stimulator	UBE2J2, Proteasome subunits, USP16, USP34	Stimulates nicotinic acetylcholine receptors with alterations in UbE2s, proteasomal activity and DUBs-USP16, USP34 reduce disease progression, improved symptoms and/or decreased L-dopa-induced dyskinesias	PD	Yes	(Hawkins et al., 2004; Quik et al., 2015)
13	PD169316 (4712)	Proteasome activator	26S Proteasome	Trigger proteasome for PROTAC and ubiquitin dependent α-synuclein degradation by p38 MAPK inhibition	PD	Unknown	(Leestemaker et al., 2017)
14	PDTC (Pyrrolidine dithiocarbamate) (65,351)	Proteasome inhibitor	Immunoproteasome	Inhibit immunoproteasome and play protective role by acting as NF-κB inhibitor, antioxidant, Akt and Nrf2 pathway inducer that improve spacial learning in AD and motor functions in PD	AD, PD	Yes	(Abd-El-Fattah et al., 2014; Chabicovsky et al., 2010; Kan et al., 2016; Malm et al., 2007)
15	Pimozide (16,362)	DUB inhibitor	USP1	Role in treatment of psychotic disorders and stabilize neuro-muscular transmission and survival in the disease	ALS, Schizophrenia	Yes	(Lee et al., 2016; Patten et al., 2017)
16	Rasagiline derivative M-30 (6,918,822) and Iron chelator VK-28	Proteasome stimulator	–	Improved lactacystin-induced DA neuron loss, proteasomal inhibition, iron accumulation, and microglial activation in Substantia nigra	AD	Yes	(Weinreb et al., 2009; Zhu et al., 2007)
17	Resveratrol (445,154)	E3 and DUB stimulator	MID1, HRD1, USP10	Induces PP2A activity by targeting MID1 to reduce tau phosphorylation. Stimulate Aβ degradation and enhance USP10-mediated ubiquitylation of p53 to induce apoptosis	AD	Yes	(Andrade et al., 2018; Schweiger et al., 2017; Yan et al., 2016)

(*continued on next page*)

Table 3 (*continued*)

S.No.	Compounds (PubChem CID)	Impact on UPS	Target UPS Component	Molecular Action	Target Pathology	BBB permeability	References
18	Rolipram (5092)	Proteasome stimulator	26S Proteasome	Enhance proteasomal degradation of protein aggregates in cAMP/PKA pathway dependent manner	AD	Yes	(Liu et al., 2012; Lokireddy et al., 2015; Myeku et al., 2016)
19	Sildenafil (135,398,744)	Proteasome stimulator	26S proteasome	Act as phosphodiesterases 5 inhibitor and stimulate proteasome to reduce protein aggregate	AD and Proteotoxic diseases	Yes	(Gómez-Vallejo et al., 2016; Myeku and Duff, 2018)
20	Sulforaphane (5350)	Proteasome stimulator	26S Proteasome	Induce Nrf2 mediated proteasomal activity and mHtt degradation and inhibit rotenone-induced locomotor deficiency and dopaminergic neuronal loss. Also, up-regulate p75 neurotrophin receptors to protect against Aβ burden	AD, PD, HD	Yes	(Jazwa et al., 2011; Liu et al., 2014; Zhang et al., 2017; Zhou et al., 2016)
21	Trehalose (7427)	E3 stimulator	CHIP	Bind to expanded polyglutamine and stabilize partially unfolded polyglutamine-containing protein and increase CHIP expression to protect against polyglutamine mediated toxicity	HD	Yes	(Allavena et al., 2019; Casarejos et al., 2014; Fernandez-Estevez et al., 2014)

AD- Alzheimer's disease; PD- Parkinson's disease; HD- Huntington's disease; ALS- Amyotrophic lateral sclerosis; MS- Multiple sclerosis; BBB- Blood Brain Barrier.

atrophy with prominent visual impairments. It primarily affects the nerves in the brain and spinal cord resulting in a range of physical, mental and psychiatric problems. Earlier diagnosis reported human leukocyte antigen as first genetic factor while later several other causes have been identified such as chronic oxidative injury, mitochondrial damage, and limited axonal transport with affected protein homeostasis (Mahad et al., 2015). However, numerous studies have reported impaired activity and composition of UPS, but exact mechanism of UPS associated MS is still remained elusive. For instance, abnormally ubiquitinated axons are reported in normally myelinated white matter in the MS brain (Giordana et al., 2002). Moreover, UPS has also found to trigger inflammatory processes in MS that was abrogated upon interferon-beta-1b treatment (Minagar et al., 2012). Another study evidenced the ubiquitination mediated pathological immunity triggered by Th17 cells via targeting its activator RORγt (retinoic acid-related orphan receptor-γt) and co-activator SRC1 (steroid receptor co-activator 1) for proteasomal degradation (He et al., 2016). Another UPS component USP18, a deubiquitinating enzyme was deficient and unresponsive to therapeutic IFNβ treatment of diseased conditions (Malhotra et al., 2013). Similarly, another ubiquitylase A20 has been implicated in the regulation of inflammation and cell death in multiple sclerosis (Wertz et al., 2015). Interestingly, a recent study has demonstrated ubiquitination independent 26S proteasomal degradation of an auto antigen, myelin basic protein (MBP) in MS (Belogurov et al., 2014). Supporting this notion, a pleiotropic cytokine osteopontin found to be processed by extracellular proteasomes to form chemotactic fragments and subsequent inflammatory responses in MS (Dianzani et al., 2017). Furthermore, impaired elimination of harmful autoimmune cells in MS is attributed to affected UPS components. For instance, pro and anti-apoptotic protein BAK and MCL1 get repressed due to altered expressions of its associated E3 ligases- HUWE1 and β-TrCP in MS patients (Mandel et al., 2012). Likewise, other E3 ligases CBL-b and ITCH associated with T-cell activation were considerably diminished in MS brains (Sellebjerg et al., 2012). Additionally, UPS has a much more complicated role in MS pathogenesis governed by combinatorial effect of both inflammatory and neurodegenerative pathways (Fig. 3E). Therefore, further identification of UPS components is required to elucidate their molecular pathogens to devise novel therapies. Moreover, the specific action of UPS enzymes regulating clearance of toxic metabolites in AD, PD, HD, ALS, and MS has been summarized in Table 2.

4. Therapeutic implications of ubiquitination system for combating neurodegenerative pathologies

Recent advancements in the pharmacological interventions pertaining ubiquitin proteasome system has uncovered prospective treatment avenues for neurodegenerative diseases. The dynamic regulation of UPS components is crucial for achieving toxic metabolite clearance in neurons. For instance, mostly alteration in the activity of many E3 ligases- Parkin, CHIP, TRAF6, HRD1; E2 conjugating enzymes- UBE2D, UBE2J, UBE2K; and DUBs- USP14, USP18, and UCH-L1 have reported to trigger diverse neurological pathologies, including synaptic dysfunction, motor impairment, mitochondrial dysfunction, ER and oxidative stress in NDDs (Zheng et al., 2016). However, there is growing evidence that identified many active forms of (de)ubiquitinases- RNF182, NEDD4-1, UBE3A, SIAH1, UBE2R1 involved in the progression of neurodegenerative pathologies (Bae et al., 2006; Fiesel et al., 2014; Liu et al., 2008; Rodrigues et al., 2016; Singh et al., 2017). Therefore, the following therapeutic strategies can be adopted to alleviate the nuisance caused by UPS dysregulation (Fig. 4).

4.1. Current therapeutic strategies to target UPS mediated pathologies

One of the strategies could be *(i) the allosteric modification of the substrate binding region of ubiquitination enzymes* with the help of small

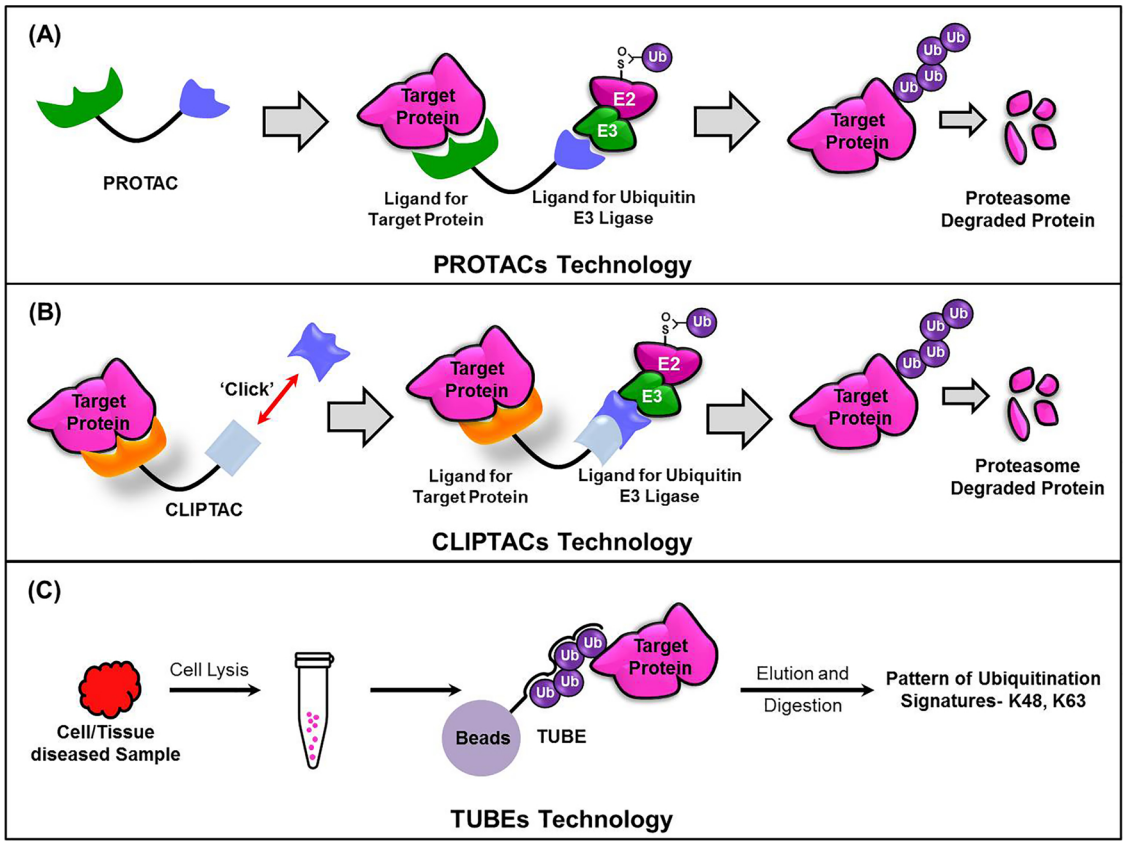


Fig. 5. Recent technologies for targeted protein degradation and ubiquitination research: (A) PROTACs Technology- This technique employs proteolysis targeting chimeras containing ligands for both target protein and E3 ligase to recruit E3 ligase to the target protein for its proteasomal degradation. (B) CLIPTACs Technology- This technique employs two-cell permeable ligands, which undergo a cycloaddition reaction to form a "In-cell click-formed proteolysis targeting chimeras" with ligands for both target protein and E3 ligase and proteasomally degrades the target protein. (C) TUBEs Technology- This technique employs the high-affinity poly-ubiquitin binding matrix, i.e. Tandem Ubiquitin Binding Entities to identify unique patterns of ubiquitination using K48 and K63 selective TUBEs.

molecules, which would enable us to increase or decrease the specificity towards their substrate to regulate protein degradation (Paiva et al., 2018). Such allosteric drugs include arginine and proline rich peptides, chloroquine and its derivatives, clioquinol, dicarboxylic acid compound (SCF-I2), and pentanoic acid compound (CC0651) that showed promising therapy against various diseases and can be studied for their efficacies in NDDs (Ceccarelli et al., 2011; Gizyńska et al., 2019; Khan et al., 2017; Kim et al., 2016; Nalbandian et al., 2015; Orlicky et al., 2010). Similarly, another way could be *(ii) the administration of the deubiquitinases and DUB regulator,* which can enhance the deubiquitination of mutant poly ubiquitin (UBB^{+1}) or normal polyUb chains to release proteasome burden and regulate synaptic functions. For instance, a DUB called UCHL1 is well identified for their roles in synaptic plasticity and prevention against neurodegeneration (Cartier et al., 2009). Another DUB, USP14 has implicated in the regulation of hippocampal short-term synaptic-plasticity (Walters et al., 2014). Recently, a study demonstrated the importance of WD repeat proteins WDR20, WDR40, and WDR48 in regulating the expression and catalytic activity of DUBs for controlling the synaptic plasticity and thus neuronal health (Hodul et al., 2017). Apart from the allosteric drugs, other ways could be *(iii) the post-transnational modifications of the UPS components,* including phosphorylation, acetylation, SUMOylation, and neddylation to modulate their activities. To illustrate, phosphorylation of Parkin (an E3 ligase) by PINK1 boosted its ligase activity, and increases unfolded protein response in stressed neurons (Kane et al., 2014). Likewise, Neddylation, i.e. Nedd8 (an Ubl) mediated modification of Cullin-1 enhanced the activity of a ubiquitin E3 ligase SCF^{skp2} towards the proteasomal degradation of p27 (Duda et al., 2008; Bornstein et al., 2006). Similarly, acetylation, SUMOylation and phosphorylation of MDM-2 augmented p53 ubiquitination (Miyauchi et al., 2002; Nihira et al., 2017). Moreover, *(iv) Selective target based specific ubiquitin-like (UBL) modifications* can aid in the therapeutics of neurodegenerative disorders; Since, Ubl modifications such as ISGlation, FATtylation of pathogenic proteins has shown interesting results in toxic protein clearance. For instance, FAT10 (an Ubl) is identified to bind and modulate the solubility of poly glutamine proteins revealing it as a therapeutic agent for Huntington's disease (Nagashima et al., 2011). Other Ubls like ATG8 and ISG15 have reported to facilitate the autophagic clearance of toxic proteins in association with p62/SQSTM1 binding and IFN-γ induction respectively (Nakashima et al., 2015; Pankiv et al., 2007; Lu et al., 2014, 2017). Likewise, another potential approach could be *(v) the activation of proteasome with suitable drugs,* which may be achieved through any of the following ways. (a) Up-regulation of 19S and 20S complex assembly, (b) eliciting the catalytic activity of 20S core proteasomal unit, and (c) triggering unfolding of aggregated proteins by modulating chaperonic activity of either chaperonins or ATPase subunit of 19S complex (Leestemaker et al., 2017). For example, a study has demonstrated to restrain Aβ levels upon resveratrol treatment by stimulating its proteasome-mediated degradation (Li et al., 2012a). Another study reported a small-molecule bicuculline stimulating calcium/calmodulin-dependent protein kinase II (CaMKII) mediated proteasomal activity to modulate synaptic connections in neurons (Djakovic et al., 2009). Likewise, another molecule IU (1-[1-(4-fluorophenyl)-2, 5-dimethylpyrrol-3-yl]-2-pyrrolidin-1-ylethanone) boosted proteasome activity by ameliorating USP14 mediated suppression of ubiquitinated conjugate's degradation (Lee et al., 2010). In addition, a natural occurring compound Sulforaphane is reported to induce proteasomal activity via triggering a transcriptional factor Nrf2

to govern mHtt degradation (Liu et al., 2014). Likewise, Forskolin, Rolipram and other cyclic AMP (cAMP) inducing compounds have shown enhanced proteasomal degradation of protein aggregates in cAMP/PKA pathway dependent manner (Lokireddy et al., 2015; Myeku et al., 2016). Recently, another small-molecule PD169316 has been found to trigger proteasomal activation and thus PROTAC and ubiquitin dependent α-synuclein degradation via hampering p38 MAPK pathway (Leestemaker et al., 2017). There is growing evidence of potential drugs and leading compound discoveries that have shown promising results in targeting the progression of various neurodegenerative diseases, including AD, PD, HD, ALS, and MS (Table 3). In spite of the advancements in drug discovery, the designing of substrate-specific drugs against diverse (de)ubiquitinases remained a challenging task due to structural similarities in their domains. In order to solve this problem a recent technology of Proteolysis-Targeting Chimera (PROTAC™) has evolved, which has been taken up by big pharmaceutical companies to target different cancers and neurodegenerative disorders.

4.2. Clinical applications of chimera construct to target neurodegenerative diseases

The selective protein degradation achieved by proteolysis targeting chimeras (PROTACs) has revolutionized the therapeutic avenues towards neurodegenerative diseases like AD and PD. The current PROTAC technology employed the use of hetero bifunctional molecules, including a ligand for an intracellular target protein and a recruiting group for a ubiquitin E3 ligase linked by a linker (Lai and Crews, 2017). Target protein degradation is initiated when PROTACs stimulate the target protein and ubiquitin E3 ligase to form a ternary complex. Then, these chimeras facilitate the ubiquitination of target proteins by specific E3 ligases, when ubiquitination machinery comes in close proximity followed by its recognition and degradation by the 26*S* proteasome in a eukaryotic cell (Fig. 5A). The specialties of these PROTACs are their ability to cross blood-brain barrier and their routes of administration, including oral, intravenous, subcutaneous, and intrathecal. Moreover, Arvinas Biopharmaceutical Company has explored the potential of PROTACs technology for targeting neurodegenerative diseases like Alzheimer's and Parkinson's disease. For instance, they demonstrated the 50 % reduction of tau proteins in the hippocampus upon direct injection of tau-directed PROTAC protein degraders. Further, they reported tau- and α-synuclein targeted PROTAC protein degraders to cross the blood-brain barriers *in-vivo,* signifying their therapeutic potential against AD and PD (http://arvinas.com/therapeutic-programs/next-generation-targets/). Another group has developed PROTACs that target polypeptide types and degrade the Tau protein. Moreover, they have demonstrated TH006 (a PROTAC) to regulate the Tau level in the brain of an AD mouse model (Chu et al., 2016). Interestingly, Jiang group has also developed a peptide PROTAC by recruiting Keap1-Cul3 ubiquitin E3 ligase and knockdown Tau by the UPS (Lu et al., 2018). Silva et al., has also demonstrated QC-01-175 effected degradation of aberrant Tau utilizing $CRL4^{CRBN}$ E3-ligase in frontotemporal dementia patient-derived neuronal cell models (Silva et al., 2019). Different pharmaceutical companies, including Genentech, Pfizer and distinct academicians have shown interest in this area and made collaborations with Arvinas. Other industries like Astex pharmaceuticals are working on the CLIPTACs protein degradation technology, i.e. in-cell click-formed proteolysis targeting chimeras (Fig. 5B). This technique employed two-cell permeable ligands that undergo a cycloaddition reaction to form a CLIPTAC, i.e. click-formed PROTAC inside the cell and thus facilitate ubiquitination and subsequent proteasomal degradation of target proteins, For successful application of CLIPTAC, we should ensure the administration of two distinct chemical entities separated in time to prevent PROTAC formation outside the cell. For instance, a CLIPTAC is developed with help of an eight-poly-D-arginine tag to improve the cell permeability of artificial ligand AP21998 and short peptide ALAPYIP to recruit mutant (F36V) FKBP12 and the E3 ubiquitin ligase CRL2 VHL complex for its proteasomal degradation (Schneekloth et al., 2004). Further, a tetrazine tagged thalidomide derivative has been designed that rapidly reacts with a trans-cyclo-octene tagged ligand of the target protein in cells to form a cereblon E3 ligase recruiting PROTAC molecule to target BRD4 and ERK1/2 for proteasomal degradation (Lebraud et al., 2016). Recently, Wang et al., has designed a CRBN-recruiting Brd4-PROTAC through multiple optimization cycles and achieved >90 % tumor regression in the corresponding xenograft model (Zhou et al., 2018). Another structure-guided design of [1,4] oxazepines (QCA570) as potent bromo domain and extra-terminal (BET) degrader is developed to restrict cell growth in acute leukemia cell lines (Qin et al., 2018). Likewise, another company, Captor Therapeutics is working to develop a molecular repair platform, i.e. low molecular compounds capable of reconstituting interactions between ubiquitin E3 ligases and their target proteins to employ UPS for target protein degradation. Similarly, other groups at LifeSensors are working to identify the bio markers for PD and other NDDs based on their TUBEs technology, i.e. pulling of significant amounts of ubiquitinated proteins from patient samples (Fig. 5C). The TUBE (Tandem ubiquitin binding entity) technology is based on the principle of identifying tetra-ubiquitin with a markedly higher affinity than single ubiquitin-associated (UBA) domains, thus enabling efficient purification of poly-ubiquitinated proteins from cell extracts in native conditions (Hjerpe et al., 2009). Therefore, this technology would enable us to identify the precise bio-markers for the complex diseases. For instance, Serna et al., has employed TUBEs technology-based microarray analysis to identify the cellular processes regulated by ubiquitination and its associated pathologies with aberrant ubiquitination levels. To achieve this, they used Adriamycin-stimulated MCF7 cells to differentiate protein ubiquitylation levels between cells that are sensitive or resistant to Adriamycin treatment (Serna et al., 2016). Interestingly, Yoshida et al. also employed this technique along with the anti-Lys-ε-Gly-Gly antibodies to identify specific ubiquitin ligase–substrate pairs. They designed trypsin-resistant (TR)-TUBE for expression in cells and co-expressed TR-TUBE with a ubiquitin ligase that stabilized the ubiquitinated substrates by masking the ubiquitin chains and thereby enabled them to identify ubiquitin ligase-substrate pairs (Yoshida et al., 2015). In this manner, the pharmaceutical companies are exploring the starring role of proteolysis in neurodegeneration, which would enable us to develop the potential future therapies avoiding non-specific side effects with enhanced drug efficacies against neurodegenerative disorders.

5. Conclusions

Although ubiquitination system has been extensively studied for the protein clearance in neurodegenerative proteinopathies over last few decades, but ubiquitination markers, sites and mechanisms remained largely unknown. The dynamic regulation of polyubiquitination system is quite intriguing at different physiological and/or pathologic conditions. Herein, we surveyed the role of ubiquitination in learning/memory processing, signaling and brain homeostasis. Moreover, it studied UPS dysregulation mechanisms in the pathophysiology of several major neurodegenerative diseases and their therapeutic regulation strategies. However, more research is required to formulate substrate-specific drugs using proteolysis-targeting chimera products as future therapy for successful treatment of neurodegenerative pathologies.

Declaration of competing interests

There is no conflict or competing interest declared by the authors.

Acknowledgments

We would like to thanks the senior management of Delhi Technological University for constant support and encouragement.

References

Abd-El-Fattah, M.A., Abdelakader, N.F., Zaki, H.F., 2014. Pyrrolidine dithiocarbamate protects against scopolamine-induced cognitive impairment in rats. Eur. J. Pharmacol. 723, 330–338. https://doi.org/10.1016/j.ejphar.2013.11.008.

Abraham, W.C., Jones, O.D., Glanzman, D.L., 2019. Is plasticity of synapses the mechanism of long-term memory storage? NPJ Sci. Learn. 4, 9. https://doi.org/10.1038/s41539-019-0048-y.

Abumrad, N.A., Moore, D.J., 2011. Parkin reinvents itself to regulate fatty acid metabolism by tagging CD36. J. Clin. Invest. 121, 3389–3392. https://doi.org/10.1172/JCI59219.

Alexopoulou, Z., Lang, J., Perrett, R.M., Elschami, M., Hurry, M.E.D., Kim, H.T., Mazaraki, D., Szabo, A., Kessler, B.M., Goldberg, A.L., Ansorge, O., Fulga, T.A., Tofaris, G.K., 2016. Deubiquitinase Usp8 regulates α-synuclein clearance and modifies its toxicity in Lewy body disease. Proc. Natl. Acad. Sci. U. S. A. 113, E4688–E4697. https://doi.org/10.1073/pnas.1523597113.

Alkalay, I., Yaron, A., Hatzubai, A., Orian, A., Ciechanover, A., Ben-Neriah, Y., 1995. Stimulation-dependent IκBα phosphorylation marks the NF-κb inhibitor for degradation via the ubiquitin-proteasome pathway. Proc. Natl. Acad. Sci. U. S. A. 92, 10599–10603. https://doi.org/10.1073/pnas.92.23.10599.

Allavena, G., Del Bello, B., Tini, P., Volpi, N., Valacchi, G., Miracco, C., Pirtoli, L., Maellaro, E., 2019. Trehalose inhibits cell proliferation and amplifies long-term temozolomide- and radiation-induced cytotoxicity in melanoma cells: a role for autophagy and premature senescence. J. Cell. Physiol. 234, 11708–11721. https://doi.org/10.1002/jcp.27838.

Allen, E., Ding, J., Wang, W., Pramanik, S., Chou, J., Yau, V., Yang, Y., 2005. Gigaxonin-controlled degradation of MAP1B light chain is critical to neuronal survival. Nature 438, 224–228. https://doi.org/10.1038/nature04256.

Andrade, S., Ramalho, M.J., Pereira, M.D.C., Loureiro, J.A., 2018. Resveratrol brain delivery for neurological disorders prevention and treatment. Front. Pharmacol. 9, 1261. https://doi.org/10.3389/fphar.2018.01261.

Ang, X.L., Seeburg, D.P., Sheng, M., Harper, J.W., 2008. Regulation of postsynaptic RapGAP SPAR by polo-like kinase 2 and the SCFβ-TRCP ubiquitin ligase in hippocampal neurons. J. Biol. Chem. 283, 29424–29432. https://doi.org/10.1074/jbc.M802475200.

Araki, K., Nagata, K., 2011. Protein folding and quality control in the ER. Cold Spring Harb. Perspect. Biol. 3, a007526. https://doi.org/10.1101/cshperspect.a007526.

Atkin, G., Hunt, J., Minakawa, E., Sharkey, L., Tipper, N., Tennant, W., Paulson, H.L., 2014. F-box only protein 2 (Fbxo2) regulates amyloid precursor protein levels and processing. J. Biol. Chem. 289, 7038–7048. https://doi.org/10.1074/jbc.M113.515056.

Bae, B.Il, Hara, M.R., Cascio, M.B., Wellington, C.L., Hayden, M.R., Ross, C.A., Ha, H.C., Li, X.J., Snyder, S.H., Sawa, A., 2006. Mutant Huntingtin: nuclear translocation and cytotoxicity mediated by GAPDH. Proc. Natl. Acad. Sci. U. S. A. 103, 3405–3409. https://doi.org/10.1073/pnas.0511316103.

Balducci, C., Forloni, G., 2019. Doxycycline for Alzheimer's disease: fighting β-amyloid oligomers and neuroinflammation. Front. Pharmacol. 10, 738. https://doi.org/10.3389/fphar.2019.00738.

Banerjee, S., Neveu, P., Kosik, K.S., 2009. A coordinated local translational control point at the synapse involving relief from silencing and MOV10 degradation. Neuron 64, 871–884. https://doi.org/10.1016/j.neuron.2009.11.023.

Belogurov, A., Kudriaeva, A., Kuzina, E., Smirnov, I., Bobik, T., Ponomarenko, N., Kravtsova-Ivantsiv, Y., Ciechanover, A., Gabibov, A., 2014. Multiple sclerosis autoantigen myelin basic protein escapes control by ubiquitination during proteasomal degradation. J. Biol. Chem. 289, 17758–17766. https://doi.org/10.1074/jbc.M113.544247.

Benvegnù, S., Mateo, M.I., Palomer, E., Jurado-Arjona, J., Dotti, C.G., 2017a. Aging triggers cytoplasmic depletion and nuclear translocation of the E3 ligase mahogunin: a function for ubiquitin in neuronal survival. Mol. Cell 66, 358–372.e7. https://doi.org/10.1016/j.molcel.2017.04.005.

Benvegnù, S., Wahle, T., Dotti, C.G., 2017b. E3 ligase mahogunin (MGRN1) influences amyloid precursor protein maturation and secretion. Oncotarget 8, 89439–89450. https://doi.org/10.18632/oncotarget.20143.

Bhat, K.P., Yan, S., Wang, C.E., Li, S., Li, X.J., 2014. Differential ubiquitination and degradation of huntingtin fragments modulated by ubiquitin-protein ligase E3A. Proc. Natl. Acad. Sci. U. S. A. 111, 5706–5711. https://doi.org/10.1073/pnas.1402215111.

Bhutani, S., Das, A., Maheshwari, M., Lakhotia, S.C., Jana, N.R., 2012. Dysregulation of core components of SCF complex in poly-glutamine disorders. Cell Death Dis. 3, e428. https://doi.org/10.1038/cddis.2012.166.

Bingol, B., Sheng, M., 2011. Deconstruction for reconstruction: the role of proteolysis in neural plasticity and disease. Neuron 69, 22–32. https://doi.org/10.1016/j.neuron.2010.11.006.

Bingol, B., Wang, C.F., Arnott, D., Cheng, D., Peng, J., Sheng, M., 2010. Autophosphorylated CaMKIIα acts as a scaffold to recruit proteasomes to dendritic spines. Cell 140, 567–578. https://doi.org/10.1016/j.cell.2010.01.024.

Bornstein, G., Ganoth, D., Hershko, A., 2006. Regulation of neddylation and deneddylation of cullin1 in SCF Skp2 ubiquitin ligase by F-box protein and substrate. Proc. Natl. Acad. Sci. U. S. A. 103, 11515–11520. https://doi.org/10.1073/pnas.0603921103.

Boselli, M., Lee, B.H., Robert, J., Prado, M.A., Min, S.W., Cheng, C., Catarina Silva, M., Seong, C., Elsasser, S., Hatle, K.M., Gahman, T.C., Gygi, S.P., Haggarty, S.J., Gan, L., King, R.W., Finley, D., 2017. An inhibitor of the proteasomal deubiquitinating enzyme USP14 induces tau elimination in cultured neurons. J. Biol. Chem. 292, 19209–19225. https://doi.org/10.1074/jbc.M117.815126.

Boyer, N.P., Monkiewicz, C., Menon, S., Moy, S.S., Gupton, S.L., 2018. Mammalian TRIM67 functions in brain development and behavior. eNeuro 5https://doi.org/10.1523/ENEURO.0186-18.2018. ENEURO.0186-18.2018.

Burk, K., Pasterkamp, R.J., 2019. Disrupted neuronal trafficking in amyotrophic lateral sclerosis. Acta Neuropathol. 137, 859–877. https://doi.org/10.1007/s00401-019-01964-7.

Burr, M.L., Cano, F., Svobodova, S., Boyle, L.H., Boname, J.M., Lehner, P.J., 2011. HRD1 and UBE2J1 target misfolded MHC class I heavy chains for endoplasmic reticulum-associated degradation. Proc. Natl. Acad. Sci. U. S. A. 8, 2034–2039. https://doi.org/10.1073/pnas.1016229108.

Bustamante, H.A., González, A.E., Cerda-Troncoso, C., Shaughnessy, R., Otth, C., Soza, A., Burgos, P.V., 2018. Interplay between the autophagy-lysosomal pathway and the ubiquitin-proteasome system: a target for therapeutic development in Alzheimer's disease. Front. Cell. Neurosci. 12, 126. https://doi.org/10.3389/fncel.2018.00126.

Callis, J., 2014. The ubiquitination machinery of the ubiquitin system. Arab. B 12, e0174. https://doi.org/10.1199/tab.0174.

Camera, D., Coleman, H.A., Parkington, H.C., Jenkins, T.A., Pow, D.V., Boase, N., Kumar, S., Poronnik, P., 2016. Learning, memory and long-term potentiation are altered in Nedd4 heterozygous mice. Behav. Brain Res. 303, 176–181. https://doi.org/10.1016/j.bbr.2016.01.054.

Caroni, P., Donato, F., Muller, D., 2012. Structural plasticity upon learning: regulation and functions. Nat. Rev. Neurosci. 13, 478–490. https://doi.org/10.1038/nrn3258.

Carroll, C.B., Zeissler, M.L., Hanemann, C.O., Zajicek, J.P., 2012. Δ9-tetrahydrocannabinol (Δ9-THC) exerts a direct neuroprotective effect in a human cell culture model of Parkinson's disease. Neuropathol. Appl. Neurobiol. 38, 535–547. https://doi.org/10.1111/j.1365-2990.2011.01248.x.

Cartier, A.E., Djakovic, S.N., Salehi, A., Wilson, S.M., Masliah, E., Patrick, G.N., 2009. Regulation of synaptic structure by ubiquitin C-terminal hydrolase L1. J. Neurosci. 29, 7857–7868. https://doi.org/10.1523/JNEUROSCI.1817-09.2009.

Cartier, A.E., Ubhi, K., Spencer, B., Vazquez-Roque, R.A., Kosberg, K.A., Fourgeaud, L., Kanayson, P., Patrick, C., Rockenstein, E., Patrick, G.N., Masliah, E., 2012. Differential effects of UCHL1 modulation on alpha-synuclein in PD-like models of alpha-synucleinopathy. PLoS One 7, e34713. https://doi.org/10.1371/journal.pone.0034713.

Casarejos, M.J., Perucho, J., López-Sendón, J.L., García De Yébenes, J., Bettencourt, C., Gómez, A., Ruiz, C., Heutink, P., Rizzu, P., Mena, M.A., 2014. Trehalose improves human fibroblast deficits in a new CHIP-mutation related ataxia. PLoS One 9, e106931. https://doi.org/10.1371/journal.pone.0106931.

Castegna, A., Aksenov, M., Aksenova, M., Thongboonkerd, V., Klein, J.B., Pierce, W.M., Booze, R., Markesbery, W.R., Butterfield, D.A., 2002. Proteomic identification of oxidatively modified proteins in Alzheimer's disease brain. Part I: Creatine kinase BB, glutamine synthase, and ubiquitin carboxy-terminal hydrolase L-1. Free Radic. Biol. Med. 33, 562–571. https://doi.org/10.1016/S0891-5849(02)00914-0.

Ceccarelli, D.F., Tang, X., Pelletier, B., Orlicky, S., Xie, W., Plantevin, V., Neculai, D., Chou, Y.C., Ogunjimi, A., Al-Hakim, A., Varelas, X., Koszela, J., Wasney, G.A., Vedadi, M., Dhe-Paganon, S., Cox, S., Xu, S., Lopez-Girona, A., Mercurio, F., Wrana, J., Durocher, D., Meloche, S., Webb, D.R., Tyers, M., Sicheri, F., 2011. An allosteric inhibitor of the human Cdc34 ubiquitin-conjugating enzyme. Cell 145, 1075–1087. https://doi.org/10.1016/j.cell.2011.05.039.

Cenini, G., Lloret, A., Cascella, R., 2019. Oxidative stress in neurodegenerative diseases: from a mitochondrial point of view. Oxid. Med. Cell. Longev. 2019, 2105607. https://doi.org/10.1155/2019/2105607.

Chabicovsky, M., Prieschl-Grassauer, E., Seipelt, J., Muster, T., Szolar, O.H.J., Hebar, A., Doblhoff-Dier, O., 2010. Pre-clinical safety evaluation of pyrrolidine dithiocarbamate. Basic Clin. Pharmacol. Toxicol. 107, 758–767. https://doi.org/10.1111/j.1742-7843.2010.00573.x.

Chang, R., Liu, X., Li, S., Li, X.J., 2015. Transgenic animal models for study of the pathogenesis of Huntington's disease and therapy. Drug Des. Devel. Ther. 9, 2179–2188. https://doi.org/10.2147/DDDT.S58470.

Changjiang, L., Cheng, L., Haitao, W., Peijie, H., Zhang, Y., Yang, Y., Chen, J., Chen, M., 2018. Activation of the KEAP1-NRF2-ARE signaling pathway reduces oxidative stress in Hep2 cells. Mol. Med. Rep. 18, 2541–2550. https://doi.org/10.3892/mmr.2018.9288.

Chhangani, D., Endo, F., Amanullah, A., Upadhyay, A., Watanabe, S., Mishra, R., Yamanaka, K., Mishra, A., 2016. Mahogunin ring finger 1 confers cytoprotection against mutant SOD1 aggresomes and is defective in an ALS mouse model. Neurobiol. Dis. 86, 16–28. https://doi.org/10.1016/j.nbd.2015.11.017.

Chin, L.S., Li, L., 2016. Ubiquitin phosphorylation in Parkinson's disease: implications for pathogenesis and treatment. Transl. Neurodegener. 5, 1. https://doi.org/10.1186/s40035-015-0049-6.

Chin, L.S., Vavalle, J.P., Li, A., 2002. Staring, a novel E3 ubiquitin-protein ligase that targets syntaxin 1 for degradation. J. Biol. Chem. 277, 35071–35079. https://doi.org/10.1074/jbc.M203300200.

Choi, J.S., Lee, D.H., 2010. CHIP promotes the degradation of mutant SOD1 by reducing its interaction with VCP and S6/S6' subunits of 26S proteasome. Anim. Cells Syst. (Seoul) 14, 1–10. https://doi.org/10.1080/19768351003765145.

Choi, J., Levey, A.I., Weintraub, S.T., Rees, H.D., Gearing, M., Chin, L.S., Li, L., 2004. Oxidative modifications and down-regulation of ubiquitin carboxyl-terminal hydrolase L1 associated with idiopathic Parkinson's and Alzheimer's diseases. J. Biol. Chem. 279, 13256–13264. https://doi.org/10.1074/jbc.M314124200.

Chu, T.T., Gao, N., Li, Q.Q., Chen, P.G., Yang, X.F., Chen, Y.X., Zhao, Y.F., Li, Y.M., 2016. Specific knockdown of endogenous tau protein by peptide-directed ubiquitin-proteasome degradation. Cell Chem. Biol. 23, 453–461. https://doi.org/10.1016/j.chembiol.2016.02.016.

Chung, K.K.K., Thomas, B., Li, X., Pletnikova, O., Troncoso, J.C., Marsh, L., Dawson, V.L., Dawson, T.M., 2004. S-nitrosylation of parkin regulates ubiquitination and compromises parkin's protective function. Science 304, 1328–1331. https://doi.org/10.

1126/science.1093891.
Cocco, E., Marrosu, M.G., 2014. The current role of mitoxantrone in the treatment of multiple sclerosis. Expert Rev. Neurother. 14, 607–616. https://doi.org/10.1586/14737175.2014.915742.
Colledge, M., Snyder, E.M., Crozier, R.A., Soderling, J.A., Jin, Y., Langeberg, L.K., Lu, H., Bear, M.F., Scott, J.D., 2003. Ubiquitination regulates PSD-95 degradation and AMPA receptor surface expression. Neuron 40, 595–607. https://doi.org/10.1016/S0896-6273(03)00687-1.
Collins, P., Mitxitorena, I., Carmody, R., 2016. The ubiquitination of NF-κB subunits in the control of transcription. Cells 5, 23. https://doi.org/10.3390/cells5020023.
D'Souza, M.S., 2015. Glutamatergic transmission in drug reward: implications for drug addiction. Front. Neurosci. 9, 404. https://doi.org/10.3389/fnins.2015.00404.
Das, J., Samadder, A., Das, S., Paul, A., Khuda-Bukhsh, A.R., 2016. Nanopharmaceutical approach for enhanced anti-cancer activity of betulinic acid in lung-cancer treatment via activation of PARP: interaction with DNA as a target anti-cancer potential of nano-betulinic acid in lung cancer. J. Pharmacopuncture 19, 37–44. https://doi.org/10.3831/KPI.2016.19.005.
Dawson, T.M., Dawson, V.L., 2010. The role of parkin in familial and sporadic Parkinson's disease. Mov. Disord. 25, S32–S39. https://doi.org/10.1002/mds.22798.
de Pril, R., Fischer, D.F., Roos, R.A.C., van Leeuwen, F.W., 2007. Ubiquitin-conjugating enzyme E2-25K increases aggregate formation and cell death in polyglutamine diseases. Mol. Cell. Neurosci. 34, 10–19. https://doi.org/10.1016/j.mcn.2006.09.006.
Deture, M.A., Dickson, D.W., 2019. The neuropathological diagnosis of Alzheimer's disease. Mol. Neurodegener. 14, 32. https://doi.org/10.1186/s13024-019-0333-5.
DiAntonio, A., Haghighi, A.P., Portman, S.L., Lee, J.D., Amaranto, A.M., Goodman, C.S., 2001. Ubiquitination-dependent mechanisms regulate synaptic growth and function. Nature 412, 449–452. https://doi.org/10.1038/35086595.
Dianzani, C., Bellavista, E., Liepe, J., Verderio, C., Martucci, M., Santoro, A., Chiocchetti, A., Gigliotti, C.L., Boggio, E., Ferrara, B., Riganti, L., Keller, C., Janek, K., Niewienda, A., Fenoglio, C., Sorosina, M., Cantello, R., Kloetzel, P.M., Stumpf, M.P.H., Paul, F., Ruprecht, K., Galimberti, D., Boneschi, F.M., Comi, C., Dianzani, U., Mishto, M., 2017. Extracellular proteasome-osteopontin circuit regulates cell migration with implications in multiple sclerosis. Sci. Rep. 7, 43718. https://doi.org/10.1038/srep43718.
Díaz-Hernández, M., Valera, A.G., Morán, M.A., Gómez-Ramos, P., Alvarez-Castelao, B., Castaño, J.G., Hernández, F., Lucas, J.J., 2006. Inhibition of 26S proteasome activity by huntingtin filaments but not inclusion bodies isolated from mouse and human brain. J. Neurochem. 98, 1585–1596. https://doi.org/10.1111/j.1471-4159.2006.03968.x.
Djakovic, S.N., Schwarz, L.A., Barylko, B., DeMartino, G.N., Patrick, G.N., 2009. Regulation of the proteasome by neuronal activity and calcium/calmodulin-dependent protein kinase II. J. Biol. Chem. 284, 26655–26665. https://doi.org/10.1074/jbc.M109.021956.
Douglas, P.M., Dillin, A., 2010. Protein homeostasis and aging in neurodegeneration. J. Cell Biol. 190, 719–729. https://doi.org/10.1083/jcb.201005144.
Duda, D.M., Borg, L.A., Scott, D.C., Hunt, H.W., Hammel, M., Schulman, B.A., 2008. Structural insights into NEDD8 activation of cullin-RING ligases: conformational control of conjugation. Cell 134, 995–1006. https://doi.org/10.1016/j.cell.2008.07.022.
Durcan, T.M., Kontogiannea, M., Bedard, N., Wing, S.S., Fon, E.A., 2012. Ataxin-3 deubiquitination is coupled to parkin ubiquitination via E2 ubiquitin-conjugating enzyme. J. Biol. Chem. 287, 531–541. https://doi.org/10.1074/jbc.M111.288449.
Durham-Lee, J.C., Mokkapati, V.U.L., Johnson, K.M., Nesic, O., 2011. Amiloride improves locomotor recovery after spinal cord injury. J. Neurotrauma 28, 1319–1326. https://doi.org/10.1089/neu.2011.1921.
Esteban, J.A., Shi, S.H., Wilson, C., Nuriya, M., Huganir, R.L., Malinow, R., 2003. PKA phosphorylation of AMPA receptor subunits controls synaptic trafficking underlying plasticity. Nat. Neurosci. 6, 136–143. https://doi.org/10.1038/nn997.
Feduccia, A.A., Chatterjee, S., Bartlett, S.E., 2012. Neuronal nicotinic acetylcholine receptors: neuroplastic changes underlying alcohol and nicotine addictions. Front. Mol. Neurosci. 5, 83. https://doi.org/10.3389/fnmol.2012.00083.
Fernandez-Estevez, M.A., Casarejos, M.J., Sendon, J.L., Caldentey, J.G., Ruiz, C., Gomez, A., Perucho, J., De Yebenes, J.G., Mena, M.A., 2014. Trehalose reverses cell malfunction in fibroblasts from normal and Huntington's disease patients caused by proteosome inhibition. PLoS One 9, e90202. https://doi.org/10.1371/journal.pone.0090202.
Ferrara, N.C., Jarome, T.J., Cullen, P.K., Orsi, S.A., Kwapis, J.L., Trask, S., Pullins, S.E., Helmstetter, F.J., 2019. GluR2 endocytosis-dependent protein degradation in the amygdala mediates memory updating. Sci. Rep. 9, 5180. https://doi.org/10.1038/s41598-019-41526-1.
Fiesel, F.C., Moussaud-Lamodière, E.L., Ando, M., Springer, W., 2014. A specific subset of E2 ubiquitin-conjugating enzymes regulate Parkin activation and mitophagy differently. J. Cell Sci. 127, 3488–3504. https://doi.org/10.1242/jcs.147520.
Fonseca, A.C., Oliveira, C.R., Pereira, C.F., Cardoso, S.M., 2014. Loss of proteostasis induced by amyloid beta peptide in brain endothelial cells. Biochim. Biophys. Acta Mol. Cell Res. 1843, 1150–1161. https://doi.org/10.1016/j.bbamcr.2014.02.016.
Freibaum, B.D., Taylor, J.P., 2017. The role of dipeptide repeats in C9ORF72-related ALS-FTD. Front. Mol. Neurosci. 10, 35. https://doi.org/10.3389/fnmol.2017.00035.
Friese, M.A., Craner, M.J., Etzensperger, R., Vergo, S., Wemmie, J.A., Welsh, M.J., Vincent, A., Fugger, L., 2007. Acid-sensing ion channel-1 contributes to axonal degeneration in autoimmune inflammation of the central nervous system. Nat. Med. 13, 1483–1489. https://doi.org/10.1038/nm1668.
Fu, A.K.Y., Hung, K.W., Fu, W.Y., Shen, C., Chen, Y., Xia, J., Lai, K.O., Ip, N.Y., 2011. APCCdh1 mediates EphA4-dependent downregulation of AMPA receptors in homeostatic plasticity. Nat. Neurosci. 14, 181–191. https://doi.org/10.1038/nn.2715.
Fu, H., Hardy, J., Duff, K.E., 2018. Selective vulnerability in neurodegenerative diseases. Nat. Neurosci. 21, 1350–1358. https://doi.org/10.1038/s41593-018-0221-2.
Fuchsberger, T., Lloret, A., Viña, J., 2017. New functions of APC/C ubiquitin ligase in the nervous system and its role in Alzheimer's disease. Int. J. Mol. Sci. 18https://doi.org/10.3390/ijms18051057. pii: E1057.
Gadhave, K., Bolshette, N., Ahire, A., Pardeshi, R., Thakur, K., Trandafir, C., Istrate, A., Ahmed, S., Lahkar, M., Muresanu, D.F., Balea, M., 2016. The ubiquitin proteasomal system: a potential target for the management of Alzheimer's disease. J. Cell. Mol. Med. 20, 1392–1407. https://doi.org/10.1111/jcmm.12817.
Geetha, T., Zheng, C., McGregor, W.C., Douglas White, B., Diaz-Meco, M.T., Moscat, J., Babu, J.R., 2012. TRAF6 and p62 inhibit amyloid β-induced neuronal death through p75 neurotrophin receptor. Neurochem. Int. 61, 1289–1293. https://doi.org/10.1016/j.neuint.2012.09.005.
Geisler, S., Vollmer, S., Golombek, S., Kahle, P.J., 2014. The ubiquitin-conjugating enzymes UBE2N, UBE2L3 and UBE2D2/3 are essential for Parkin-dependent mitophagy. J. Cell Sci. 127, 3280–3293. https://doi.org/10.1242/jcs.146035.
Gerakis, Y., Hetz, C., 2018. Emerging roles of ER stress in the etiology and pathogenesis of Alzheimer's disease. FEBS J. 285, 995–1011. https://doi.org/10.1111/febs.14332.
Gersch, M., Gladkova, C., Schubert, A.F., Michel, M.A., Maslen, S., Komander, D., 2017. Mechanism and regulation of the Lys6-selective deubiquitinase USP30. Nat. Struct. Mol. Biol. 24, 920–930. https://doi.org/10.1038/nsmb.3475.
Gilbert, T.M., Zürcher, N.R., Catanese, M.C., Tseng, C.E.J., Di Biase, M.A., Lyall, A.E., Hightower, B.G., Parmar, A.J., Bhanot, A., Wu, C.J., Hibert, M.L., Kim, M., Mahmood, U., Stufflebeam, S.M., Schroeder, F.A., Wang, C., Roffman, J.L., Holt, D.J., Greve, D.N., Pasternak, O., Kubicki, M., Wey, H.Y., Hooker, J.M., 2019. Neuroepigenetic signatures of age and sex in the living human brain. Nat. Commun. 10, 2945. https://doi.org/10.1038/s41467-019-11031-0.
Giordana, M.T., Richiardi, P., Trevisan, E., Boghi, A., Palmucci, L., 2002. Abnormal ubiquitination of axons in normally myelinated white matter in multiple sclerosis brain. Neuropathol. Appl. Neurobiol. 28, 35–41. https://doi.org/10.1046/j.1365-2990.2002.00372.x.
Gizyńska, M., Witkowska, J., Karpowicz, P., Rostankowski, R., Chocron, E.S., Pickering, A.M., Osmulski, P., Gaczynska, M., Jankowska, E., 2019. Proline- and arginine-rich peptides as flexible allosteric modulators of human proteasome activity. J. Med. Chem. 62, 359–370. https://doi.org/10.1021/acs.jmedchem.8b01025.
Gõmez-Vallejo, V., Ugarte, A., García-Barroso, C., Cuadrado-Tejedor, M., Szczupak, B., Dopeso-Reyes, I.G., Lanciego, J.L., García-Osta, A., Llop, J., Oyarzabal, J., Franco, R., 2016. Pharmacokinetic investigation of sildenafil using positron emission tomography and determination of its effect on cerebrospinal fluid cGMP levels. J. Neurochem. 136, 403–415. https://doi.org/10.1111/jnc.13454.
Gong, B., Cao, Z., Zheng, P., Vitolo, O.V., Liu, S., Staniszewski, A., Moolman, D., Zhang, H., Shelanski, M., Arancio, O., 2006. Ubiquitin hydrolase Uch-L1 rescues β-Amyloid-Induced decreases in synaptic function and contextual memory. Cell 126, 775–788. https://doi.org/10.1016/j.cell.2006.06.046.
Gong, B., Chen, F., Pan, Y., Arrieta-Cruz, I., Yoshida, Y., Haroutunian, V., Pasinetti, G.M., 2010. SCFFbx2-E3-ligase-mediated degradation of BACE1 attenuates Alzheimer's disease amyloidosis and improves synaptic function. Aging Cell 9, 1018–1031. https://doi.org/10.1111/j.1474-9726.2010.00632.x.
González-Lizárraga, F., Socías, S.B., Ávila, C.L., Torres-Bugeau, C.M., Barbosa, L.R.S., Binolfi, A., Sepúlveda-Díaz, J.E., Del-Bel, E., Fernandez, C.O., Papy-Garcia, D., Itri, R., Raisman-Vozari, R., Chehín, R.N., 2017. Repurposing doxycycline for synucleinopathies: remodelling of α-synuclein oligomers towards non-toxic parallel beta-sheet structured species. Sci. Rep. 7, 41755. https://doi.org/10.1038/srep41755.
Guglielmotto, M., Monteleone, D., Vasciaveo, V., Repetto, I.E., Manassero, G., Tabaton, M., Tamagno, E., 2017. The decrease of Uch-L1 activity is a common mechanism responsible for Aβ 42 accumulation in Alzheimer's and vascular disease. Front. Aging Neurosci. 9, 320. https://doi.org/10.3389/fnagi.2017.00320.
Guntupalli, S., Jang, S.E., Zhu, T., Huganir, R.L., Widagdo, J., Anggono, V., 2017. GluA1 subunit ubiquitination mediates amyloid-β-induced loss of surface α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors. J. Biol. Chem. 292, 8186–8194. https://doi.org/10.1074/jbc.M116.774554.
Guo, Y.X., He, L.Y., Zhang, M., Wang, F., Liu, F., Peng, W.X., 2016. 1,25-Dihydroxyvitamin D3 regulates expression of LRP1 and RAGE in vitro and in vivo, enhancing Aβ1-40 brain-to-blood efflux and peripheral uptake transport. Neuroscience 8, 20. https://doi.org/10.1016/j.neuroscience.2016.01.041.
Haddad, D.M., Vilain, S., Vos, M., Esposito, G., Matta, S., Kalscheuer, V.M., Craessaerts, K., Leyssen, M., Nascimento, R.M.P., Vianna-Morgante, A.M., DeStrooper, B., VanEsch, H., Morais, V.A., Verstreken, P., 2013. Mutations in the intellectual disability gene Ube2a cause neuronal dysfunction and impair parkin-dependent mitophagy. Mol. Cell 50, 831–843. https://doi.org/10.1016/j.molcel.2013.04.012.
Hans, F., Fiesel, F.C., Strong, J.C., Jackel, S., Rasse, T.M., Geisler, S., Springer, W., Schulz, J.B., Voigt, A., Kahle, P.J., 2014. UBE2E ubiquitin-conjugating enzymes and ubiquitin isopeptidase y regulate TDP-43 protein ubiquitination. J. Biol. Chem. 289, 19164–19179. https://doi.org/10.1074/jbc.M114.561704.
Harding, R.J., Tong, Y.F., 2018. Proteostasis in Huntington's disease: disease mechanisms and therapeutic opportunities. Acta Pharmacol. Sin. 39, 754–769. https://doi.org/10.1038/aps.2018.11.
Hawkins, B.T., Abbruscato, T.J., Egleton, R.D., Brown, R.C., Huber, J.D., Campos, C.R., Davis, T.P., 2004. Nicotine increases in vivo blood-brain barrier permeability and alters cerebral microvascular tight junction protein distribution. Brain Res. 1027, 48–58. https://doi.org/10.1016/j.brainres.2004.08.043.
He, Z., Wang, F., Ma, J., Sen, S., Zhang, J., Gwack, Y., Zhou, Y., Sun, Z., 2016. Ubiquitination of RORγt at lysine 446 limits Th17 differentiation by controlling coactivator recruitment. J. Immunol. 197, 1148–1158. https://doi.org/10.4049/jimmunol.1600548.
Hebron, M.L., Lonskaya, I., Sharpe, K., Weerasinghe, P.P.K., Algarzae, N.K., Shekoyan,

A.R., Moussa, C.E.H., 2013. Parkin ubiquitinates tar-DNA binding protein-43 (TDP-43) and promotes its cytosolic accumulation via interaction with histone deacetylase 6 (HDAC6). J. Biol. Chem. 288, 4103–4115. https://doi.org/10.1074/jbc.M112.419945.

Heink, S., Ludwig, D., Kloetzel, P.M., Krüger, E., 2005. IFN-γ-induced immune adaptation of the proteasome system is an accelerated and transient response. Proc. Natl. Acad. Sci. U. S. A. 102, 9241–9246. https://doi.org/10.1073/pnas.0501711102.

Hershko, A., 2005. The ubiquitin system for protein degradation and some of its roles in the control of the cell division cycle. Cell Death Differ. 44, 5932–5943. https://doi.org/10.1038/sj.cdd.4401702.

Hjerpe, R., Aillet, F., Lopitz-Otsoa, F., Lang, V., England, P., Rodriguez, M.S., 2009. Efficient protection and isolation of ubiquitylated proteins using tandem ubiquitin-binding entities. EMBO Rep. 10, 1250–1258. https://doi.org/10.1038/embor.2009.192.

Hodul, M., Dahlberg, C.L., Juo, P., 2017. Function of the deubiquitinating enzyme USP46 in the nervous system and its regulation by WD40-Repeat proteins. Front. Synaptic Neurosci. 9, 16. https://doi.org/10.3389/fnsyn.2017.00016.

Holmberg, C.I., Staniszewski, K.E., Mensah, K.N., Matouschek, A., Morimoto, R.I., 2004. Inefficient degradation of truncated polyglutamine proteins by the proteasome. EMBO J. 23, 4307–4318. https://doi.org/10.1038/sj.emboj.7600426.

Hong, L., Huang, H.C., Jiang, Z.F., 2014. Relationship between amyloid-beta and the ubiquitin-proteasome system in Alzheimer's disease. Neurol. Res. 36, 276–282. https://doi.org/10.1179/1743132813Y.0000000288.

Honke, N., Shaabani, N., Zhang, D.E., Hardt, C., Lang, K.S., 2016. Multiple functions of USP18. Cell Death Dis. 7, e2444. https://doi.org/10.1038/cddis.2016.326.

Howard, R.A., Sharma, P., Hajjar, C., Caldwell, K.A., Caldwell, G.A., du Breuil, R., Moore, R., Boyd, L., 2007. Ubiquitin conjugating enzymes participate in polyglutamine protein aggregation. BMC Cell Biol. 8, 32. https://doi.org/10.1186/1471-2121-8-32.

Hsieh, M.C., Ho, Y.C., Lai, C.Y., Chou, D., Chen, G.Den, Lin, T.Bin, Peng, H.Y., 2018. Spinal TNF-α impedes Fbxo45-dependent Munc13-1 ubiquitination to mediate neuropathic allodynia in rats. Cell Death Dis. 9, 811. https://doi.org/10.1038/s41419-018-0859-4.

Huang, L., Ho, P., Chen, C.H., 2007. Activation and inhibition of the proteasome by betulinic acid and its derivatives. FEBS Lett. 581, 4955–4959. https://doi.org/10.1016/j.febslet.2007.09.031.

Hung, A.Y., Sung, C.C., Brito, I.L., Sheng, M., 2010. Degradation of postsynaptic scaffold GKAP and regulation of dendritic spine morphology by the TRIM3 ubiquitin ligase in rat hippocampal neurons. PLoS One 5, e9842. https://doi.org/10.1371/journal.pone.0009842.

Hyrskyluoto, A., Bruelle, C., Lundh, S.H., Do, H.T., Kivinen, J., Rappou, E., Reijonen, S., Waltimo, T., Petersén, Å., Lindholm, D., Korhonen, L., 2014. Ubiquitin-specific protease-14 reduces cellular aggregates and protects against mutant huntingtin-induced cell degeneration: involvement of the proteasome and ER stress-activated kinase IRE1α Hum. Mol. Genet. 23, 5928–5939. https://doi.org/10.1093/hmg/ddu317.

Iwata, A., Nagashima, Y., Matsumoto, L., Suzuki, T., Yamanaka, T., Date, H., Deoka, K., Nukina, N., Tsuji, S., 2009. Intranuclear degradation of polyglutamine aggregates by the ubiquitin-proteasome system. J. Biol. Chem. 284, 9796–9803. https://doi.org/10.1074/jbc.M809739200.

Jacoupy, M., Hamon-Keromen, E., Ordureau, A., Erpapazoglou, Z., Coge, F., Corvol, J.-C., Nosjean, O., Mannoury la Cour, C., Millan, M.J., Boutin, J.A., Harper, J.W., Brice, A., Guedin, D., Gautier, C.A., Corti, O., 2019. The PINK1 kinase-driven ubiquitin ligase Parkin promotes mitochondrial protein import through the presequence pathway in living cells. Sci. Rep. 9, 11829. https://doi.org/10.1038/s41598-019-47352-9.

Jansen, A.H.P., Reits, E.A.J., Hol, E.M., 2014. The ubiquitin proteasome system in glia and its role in neurodegenerative diseases. Front. Mol. Neurosci. 7, 73. https://doi.org/10.3389/fnmol.2014.00073.

Jarome, T.J., Werner, C.T., Kwapis, J.L., Helmstetter, F.J., 2011. Activity dependent protein degradation is critical for the formation and stability of fear memory in the amygdala. PLoS One 6, e24349. https://doi.org/10.1371/journal.pone.0024349.

Jazwa, A., Rojo, A.I., Innamorato, N.G., Hesse, M., Fernández-Ruiz, J., Cuadrado, A., 2011. Pharmacological targeting of the transcription factor NRf2 at the basal ganglia provides disease modifying therapy for experimental Parkinsonism. Antioxid. Redox Signal. 14, 2347–2360. https://doi.org/10.1089/ars.2010.3731.

Jedynak, J., Hearing, M., Ingebretson, A., Ebner, S.R., Kelly, M., Fischer, R.A., Kourrich, S., Thomas, M.J., 2016. Cocaine and amphetamine induce overlapping but distinct patterns of AMPAR plasticity in nucleus accumbens medium spiny neurons. Neuropsychopharmacology 41, 464–476. https://doi.org/10.1038/npp.2015.168.

Ji, C.H., Kwon, Y.T., 2017. Crosstalk and interplay between the ubiquitin-proteasome system and autophagy. Mol. Cells 40, 441–449. https://doi.org/10.14348/molcells.2017.0115.

Jing, X., Sui, W.H., Wang, S., Xu, X.F., Yuan, R.R., Chen, X.R., Ma, H.X., Zhu, Y.X., Sun, J.K., Yi, F., Chen, Z.Y., Wang, Y., 2017. HDAC7 ubiquitination by the E3 Ligase CBX4 is involved in contextual fear conditioning memory formation. J. Neurosci. 37, 3848–3863. https://doi.org/10.1523/JNEUROSCI.2773-16.2017.

Johnston-Carey, H.K., Pomatto, L.C.D., Davies, K.J.A., 2016. The Immunoproteasome in oxidative stress, aging, and disease. Crit. Rev. Biochem. Mol. Biol. 51, 268–281. https://doi.org/10.3109/10409238.2016.1172554.

Joshi, V., Amanullah, A., Upadhyay, A., Mishra, R., Kumar, A., Mishra, A., 2016. A decade of boon or burden: what has the chip ever done for cellular protein quality control mechanism implicated in neurodegeneration and aging? Front. Mol. Neurosci. 9, 93. https://doi.org/10.3389/fnmol.2016.00093.

Juenemann, K., Schipper-Krom, S., Wiemhoefer, A., Kloss, A., Sanz, A.S., Reits, E.A.J., 2013. Expanded polyglutamine-containing N-terminal huntingtin fragments are entirely degraded by mammalian proteasomes. J. Biol. Chem. 288, 27068–27084. https://doi.org/10.1074/jbc.M113.486076.

Juenemann, K., Wiemhoefer, A., Reits, E.A., 2015. Detection of ubiquitinated huntingtin species in intracellular aggregates. Front. Mol. Neurosci. 8, 1. https://doi.org/10.3389/fnmol.2015.00001.

Juenemann, K., Jansen, A.H.P., Van Riel, L., Merkx, R., Mulder, M.P.C., An, H., Statsyuk, A., Kirstein, J., Ovaa, H., Reits, E.A., 2018. Dynamic recruitment of ubiquitin to mutant huntingtin inclusion bodies. Sci. Rep. 8, 1405. https://doi.org/10.1038/s41598-018-19538-0.

Kan, M.H., Yang, T., Fu, H.Q., Fan, L., Wu, Y., Terrando, N., Wang, T.L., 2016. Pyrrolidine dithiocarbamate prevents neuroinflammation and cognitive dysfunction after endotoxemia in rats. Front. Aging Neurosci. 8, 175. https://doi.org/10.3389/fnagi.2016.00175.

Kanarek, N., Ben-Neriah, Y., 2012. Regulation of NF-κB by ubiquitination and degradation of the IκBs. Immunol. Rev. 246, 77–94. https://doi.org/10.1111/j.1600-065X.2012.01098.x.

Kane, L.A., Lazarou, M., Fogel, A.I., Li, Y., Yamano, K., Sarraf, S.A., Banerjee, S., Youle, R.J., 2014. PINK1 phosphorylates ubiquitin to activate parkin E3 ubiquitin ligase activity. J. Cell Biol. 205, 143–153. https://doi.org/10.1083/jcb.201402104.

Kaneko, M., 2016. Physiological roles of ubiquitin ligases related to the endoplasmic reticulum. Yakugaku Zasshi 136, 805–809. https://doi.org/10.1248/yakushi.15-00292-2.

Kato, A., Rouach, N., Nicoll, R.A., Bredt, D.S., 2005. Activity-dependent NMDA receptor degradation mediated by retrotranslocation and ubiquitination. Proc. Natl. Acad. Sci. U. S. A. 102, 5600–5605. https://doi.org/10.1073/pnas.0501769102.

Keller, J.N., Hanni, K.B., Markesbery, W.R., 2000a. Impaired proteasome function in Alzheimer's disease. J. Neurochem. 75, 436–439. https://doi.org/10.1046/j.1471-4159.2000.0750436.x.

Keller, J.N., Hanni, K.B., Markesbery, W.R., 2000b. Possible involvement of proteasome inhibition in aging: implications for oxidative stress. Mech. Ageing Dev. 113, 61–70. https://doi.org/10.1016/S0047-6374(99)00101-3.

Keller, J.N., Huang, F.F., Markesbery, W.R., 2000c. Decreased levels of proteasome activity and proteasome expression in aging spinal cord. Neuroscience 98, 149–156. https://doi.org/10.1016/S0306-4522(00)00067-1.

Khan, N.A., Khan, I., Abid, S.M.A., Zaib, S., Ibrar, A., Andleeb, H., Hameed, S., Iqbal, J., 2017. Quinolinic carboxylic acid derivatives as potential multi-target compounds for neurodegeneration: monoamine oxidase and cholinesterase inhibition. Med. Chem. (Los Angeles) 14, 74–85. https://doi.org/10.2174/1573406413666170525125231.

Kim, C.H., Leblanc, P., Kim, K.S., 2016. 4-amino-7-chloroquinoline derivatives for treating Parkinsons disease: implications for drug discovery. Expert Opin. Drug Discov. 11, 337–341. https://doi.org/10.1517/17460441.2016.1154529.

Kiprowska, M.J., Stepanova, A., Todaro, D.R., Galkin, A., Haas, A., Wilson, S.M., Figueiredo-Pereira, M.E., 2017. Neurotoxic mechanisms by which the USP14 inhibitor IU1 depletes ubiquitinated proteins and Tau in rat cerebral cortical neurons: relevance to Alzheimer's disease. Biochim. Biophys. Acta Mol. Basis Dis. 1863, 1157–1170. https://doi.org/10.1016/j.bbadis.2017.03.017.

Ko, H.S., Bailey, R., Smith, W.W., Liu, Z., Shin, J.H., Lee, Y.Il, Zhang, Y.J., Jiang, H., Ross, C.A., Moore, D.J., Patterson, C., Petrucelli, L., Dawson, T.M., Dawson, V.L., 2009. CHIP regulates leucine-rich repeat kinase-2 ubiquitination, degradation, and toxicity. Proc. Natl. Acad. Sci. U. S. A. 106, 2897–2902. https://doi.org/10.1073/pnas.0810123106.

Kwon, Y.T., Ciechanover, A., 2017. The ubiquitin code in the ubiquitin-proteasome system and autophagy. Trends Biochem. Sci. 42, 873–886. https://doi.org/10.1016/j.tibs.2017.09.002.

Lai, A.C., Crews, C.M., 2017. Induced protein degradation: an emerging drug discovery paradigm. Nat. Rev. Drug Discov. 16, 101–114. https://doi.org/10.1038/nrd.2016.211.

Launay, N., Ruiz, M., Fourcade, S., Schlüter, A., Guilera, C., Ferrer, I., Knecht, E., Pujol, A., 2013. Oxidative stress regulates the ubiquitin-proteasome system and immunoproteasome functioning in a mouse model of X-adrenoleukodystrophy. Brain 136, 891–904. https://doi.org/10.1093/brain/aws370.

Lebraud, H., Wright, D.J., Johnson, C.N., Heightman, T.D., 2016. Protein degradation by in-cell self-assembly of proteolysis targeting chimeras. ACS Cent. Sci. 2, 927–934. https://doi.org/10.1021/acscentsci.6b00280.

Lee, S.J., Langhans, S.A., 2012. Anaphase-promoting complex/cyclosome protein Cdc27 is a target for curcumin-induced cell cycle arrest and apoptosis. BMC Cancer 12, 44. https://doi.org/10.1186/1471-2407-12-44.

Lee, B.H., Lee, M.J., Park, S., Oh, D.C., Elsasser, S., Chen, P.C., Gartner, C., Dimova, N., Hanna, J., Gygi, S.P., Wilson, S.M., King, R.W., Finley, D., 2010. Enhancement of proteasome activity by a small-molecule inhibitor of USP14. Nature 467, 179–184. https://doi.org/10.1038/nature09299.

Lee, J.K., Chang, N., Yoon, Y., Yang, H., Cho, H., Kim, E., Shin, Y., Kang, W., Oh, Y.T., Mun, G.I., Joo, K.M., Nam, D.H., Lee, J., 2016. USP1 targeting impedes GBM growth by inhibiting stem cell maintenance and radioresistance. Neuro Oncol. 18, 37–47. https://doi.org/10.1093/neuonc/nov091.

Lee, Y.C., Huang, W.C., Lin, J.H., Kao, T.J., Lin, Hui Ching, Lee, K.H., Lin, Hsin Chuan, Shen, C.K.J., Chang, W.C., Huang, C.C., 2018. Znf179 E3 ligase-mediated TDP-43 polyubiquitination is involved in TDP-43- ubiquitinated inclusions (UBI) (+)-related neurodegenerative pathology. J. Biomed. Sci. 25, 76. https://doi.org/10.1186/s12929-018-0479-4.

Leestemaker, Y., de Jong, A., Witting, K.F., Penning, R., Schuurman, K., Rodenko, B., Zaal, E.A., van de Kooij, B., Laufer, S., Heck, A.J.R., Borst, J., Scheper, W., Berkers, C.R., Ovaa, H., 2017. Proteasome activation by small molecules. Cell Chem. Biol. 24, 725–736.e7. https://doi.org/10.1016/j.chembiol.2017.05.010.

Lefaki, M., Papaevgeniou, N., Chondrogianni, N., 2017. Redox regulation of proteasome function. Redox Biol. 13, 452–458. https://doi.org/10.1016/j.redox.2017.07.005.

Li, F., Gong, Q., Dong, H., Shi, J., 2012a. Resveratrol, a neuroprotective supplement for Alzheimer's disease. Curr. Pharm. Des. 18, 27–33. https://doi.org/10.2174/138161212798919075.

Li, J.M., Yang, Y., Zhu, P., Zheng, F., Gong, F.L., Mei, Y.W., 2012b. Mitoxantrone exerts both cytotoxic and immunoregulatory effects on activated microglial cells. Immunopharmacol. Immunotoxicol. 34, 36–41. https://doi.org/10.3109/08923973.2011.572890.

Lin, A.W., Man, H.Y., 2013. Ubiquitination of neurotransmitter receptors and post-synaptic scaffolding proteins. Neural Plast. 2013, 432057. https://doi.org/10.1155/2013/432057.

Liu, S., Chen, Z.J., 2011. Expanding role of ubiquitination in NF-κB signaling. Cell Res. 21, 6–21. https://doi.org/10.1038/cr.2010.170.

Liu, Q.Y., Lei, J.X., Sikorska, M., Liu, R., 2008. A novel brain-enriched E3 ubiquitin ligase RNF182 is up regulated in the brains of Alzheimer's patients and targets ATP6V0C for degradation. Mol. Neurodegener. 3, 4. https://doi.org/10.1186/1750-1326-3-4.

Liu, S., Yu, C., Yang, F., Paganini-Hill, A., Fisher, M.J., 2012. Phosphodiesterase inhibitor modulation of brain microvascular endothelial cell barrier properties. J. Neurol. Sci. 320, 45–51. https://doi.org/10.1016/j.jns.2012.06.005.

Liu, Y., Hettinger, C.L., Zhang, D., Rezvani, K., Wang, X., Wang, H., 2014. Sulforaphane enhances proteasomal and autophagic activities in mice and is a potential therapeutic reagent for Huntington's disease. J. Neurochem. 129, 539–547. https://doi.org/10.1111/jnc.12647.

Liu, H., Li, W., Rose, M.E., Hickey, R.W., Chen, J., Uechi, G.T., Balasubramani, M., Day, B.W., Patel, K.V., Graham, S.H., 2015. The point mutation UCH-L1 C152A protects primary neurons against cyclopentenone prostaglandin-induced cytotoxicity: implications for post-ischemic neuronal injury. Cell Death Dis. 6, e1966. https://doi.org/10.1038/cddis.2015.323.

Liu, X., Hebron, M., Shi, W., Lonskaya, I., Moussa, C.E.H., 2019. Ubiquitin specific protease-13 independently regulates parkin ubiquitination and alpha-synuclein clearance in alpha-synucleinopathies. Hum. Mol. Genet. 28, 548–560. https://doi.org/10.1093/hmg/ddy365.

Lokireddy, S., Kukushkin, N.V., Goldberg, A.L., 2015. cAMP-induced phosphorylation of 26S proteasomes on Rpn6/PSMD11 enhances their activity and the degradation of misfolded proteins. Proc. Natl. Acad. Sci. U. S. A. 112, E7176–E7185. https://doi.org/10.1073/pnas.1522332112.

Lonskaya, I., Desforges, N.M., Hebron, M.L., Moussa, C.E.H., 2013. Ubiquitination increases parkin activity to promote autophagic a-synuclein clearance. PLoS One 8, e83914. https://doi.org/10.1371/journal.pone.0083914.

Lu, B., Al-Ramahi, I., Valencia, A., Wang, Q., Berenshteyn, F., Yang, H., Gallego-Flores, T., Ichcho, S., Lacoste, A., Hild, M., Difiglia, M., Botas, J., Palacino, J., 2013. Identification of NUB1 as a suppressor of mutant Huntingtin toxicity via enhanced protein clearance. Nat. Neurosci. 16, 562–570. https://doi.org/10.1038/nn.3367.

Lu, K., Psakhye, I., Jentsch, S., 2014. A new class of ubiquitin-Atg8 receptors involved in selective autophagy and polyQ protein clearance. Autophagy 10 (12), 2381–2382. https://doi.org/10.4161/15548627.2014.981919.

Lu, K., den Brave, F., Jentsch, S., 2017. Receptor oligomerization guides pathway choice between proteasomal and autophagic degradation. Nat. Cell Biol. 19 (6), 732–739. https://doi.org/10.1038/ncb3531.

Lu, M., Liu, T., Jiao, Q., Ji, J., Tao, M., Liu, Y., You, Q., Jiang, Z., 2018. Discovery of a Keap1-dependent peptide PROTAC to knockdown Tau by ubiquitination-proteasome degradation pathway. Eur. J. Med. Chem. 146, 251–259. https://doi.org/10.1016/j.ejmech.2018.01.063.

Lüscher, C., Malenka, R.C., 2011. Drug-evoked synaptic plasticity in addiction: from molecular changes to circuit remodeling. Neuron 69, 650–663. https://doi.org/10.1016/j.neuron.2011.01.017.

Lussier, M.P., Herring, B.E., Nasu-Nishimura, Y., Neutzner, A., Karbowski, M., Youle, R.J., Nicoll, R.A., Roche, K.W., 2012. Ubiquitin ligase RNF167 regulates AMPA receptor-mediated synaptic transmission. Proc. Natl. Acad. Sci. U. S. A. 109, 19426–19431. https://doi.org/10.1073/pnas.1217477109.

Mabb, A.M., Ehlers, M.D., 2010. Ubiquitination in postsynaptic function and plasticity. Annu. Rev. Cell Dev. Biol. 26, 179–210. https://doi.org/10.1146/annurev-cellbio-100109-104129.

Mabb, A.M., Je, H.S., Wall, M.J., Robinson, C.G., Larsen, R.S., Qiang, Y., Corrêa, S.A.L., Ehlers, M.D., 2014. Triad3A regulates synaptic strength by ubiquitination of arc. Neuron 82, 1299–1316. https://doi.org/10.1016/j.neuron.2014.05.016.

Mahad, D.H., Trapp, B.D., Lassmann, H., 2015. Pathological mechanisms in progressive multiple sclerosis. Lancet Neurol. 14, 183–193. https://doi.org/10.1016/S1474-4422(14)70256-X.

Maheshwari, M., Shekhar, S., Singh, B.K. uma., Jamal, I., Vatsa, N., Kumar, V., Sharma, A., Jana, N.R. anja., 2014. Deficiency of Ube3a in Huntington's disease mice brain increases aggregate load and accelerates disease pathology. Hum. Mol. Genet. 23, 6235–6245. https://doi.org/10.1093/hmg/ddu343.

Maiti, P., Manna, J., Veleri, S., Frautschy, S., 2014. Molecular chaperone dysfunction in neurodegenerative diseases and effects of curcumin. Biomed Res. Int. 2014, 495091. https://doi.org/10.1155/2014/495091.

Malhotra, S., Morcillo-Suárez, C., Nurtdinov, R., Rio, J., Sarro, E., Moreno, M., Castilló, J., Navarro, A., Montalban, X., Comabella, M., 2013. Roles of the ubiquitin peptidase USP18 in multiple sclerosis and the response to interferon-β treatment. Eur. J. Neurol. 20, 1390–1397. https://doi.org/10.1111/ene.12193.

Malm, T.M., Iivonen, H., Goldsteins, G., Keksa-Goldsteine, V., Ahtoniemi, T., Kanninen, K., Salminen, A., Auriola, S., Van Groen, T., Tanila, H., Koistinaho, J., 2007. Pyrrolidine dithiocarbamate activates Akt and improves spatial learning in APP/PS1 mice without affecting β-amyloid burden. J. Neurosci. 27, 3712–3721. https://doi.org/10.1523/JNEUROSCI.0059-07.2007.

Mandel, I., Paperna, T., Miller, A., 2012. Aberrant expression of the apoptosis-related proteins BAK and MCL1 in T cells in multiple sclerosis. J. Neuroimmunol. 244, 51–56. https://doi.org/10.1016/j.jneuroim.2011.12.026.

Mao, J., Xia, Q., Liu, C., Ying, Z., Wang, H., Wang, G., 2017. A critical role of Hrd1 in the regulation of optineurin degradation and aggresome formation. Hum. Mol. Genet. 26, 1877–1889. https://doi.org/10.1093/hmg/ddx096.

Massaly, N., Dahan, L., Baudonnat, M., Hovnanian, C., Rekik, K., Solinas, M., David, V., Pech, S., Zajac, J.M., Roullet, P., Mouledous, L., Frances, B., 2013. Involvement of protein degradation by the ubiquitin proteasome system in opiate addictive behaviors. Neuropsychopharmacology 38, 596–604. https://doi.org/10.1038/npp.2012.217.

Massaly, N., Francès, B., Moulédous, L., 2015. Roles of the ubiquitin proteasome system in the effects of drugs of abuse. Front. Mol. Neurosci. 7, 99. https://doi.org/10.3389/fnmol.2014.00099.

Mealer, R.G., Murray, A.J., Shahani, N., Subramaniam, S., Snyder, S.H., 2014. Rhes, a striatal-selective protein implicated in huntington disease, binds Beclin-1 and activates autophagy. J. Biol. Chem. 289, 3547–3554. https://doi.org/10.1074/jbc.M113.536912.

Medina, J.H., Viola, H., 2018. ERK1/2: a key cellular component for the formation, retrieval, reconsolidation and persistence of memory. Front. Mol. Neurosci. 11, 361. https://doi.org/10.3389/fnmol.2018.00361.

Mhyre, T.R., Boyd, J.T., Hamill, R.W., Maguire-Zeiss, K.A., 2012. Parkinson's disease. Subcell. Biochem. 65, 389–455. https://doi.org/10.1007/978-94-007-5416-4_16.

Miller, V.M., Nelson, R.F., Gouvion, C.M., Williams, A., Rodriguez-Lebron, E., Harper, S.Q., Davidson, B.L., Rebagliati, M.R., Paulson, H.L., 2005. CHIP suppresses polyglutamine aggregation and toxicity in vitro and in vivo. J. Neurosci. 25, 9152–9161. https://doi.org/10.1523/JNEUROSCI.3001-05.2005.

Min, J.W., Lü, L., Freeling, J.L., Martin, D.S., Wang, H., 2017. USP14 inhibitor attenuates cerebral ischemia/reperfusion-induced neuronal injury in mice. J. Neurochem. 140, 826–833. https://doi.org/10.1111/jnc.13941.

Minagar, A., Ma, W., Zhang, X., Wang, X., Zhang, K., Steven Alexander, J., Gonzalez-Toledo, E., Albitar, M., 2012. Plasma ubiquitin-proteasome system profile in patients with multiple sclerosis: correlation with clinical features, neuroimaging, and treatment with interferon-beta-1b. Neurol. Res. 34, 611–618. https://doi.org/10.1179/1743132812Y.0000000055.

Mishra, A., Maheshwari, M., Chhangani, D., Fujimori-Tonou, N., Endo, F., Joshi, A.P., Jana, N.R., Yamanaka, K., 2013. E6-AP association promotes SOD1 aggresomes degradation and suppresses toxicity. Neurobiol. Aging 34, 1310.e11–1310.e23. https://doi.org/10.1016/j.neurobiolaging.2012.08.016.

Mitra, S., Finkbeiner, S., 2008. The ubiquitin-proteasome pathway in Huntington's disease. Sci. World J. 8, 421–433. https://doi.org/10.1100/tsw.2008.60.

Miyauchi, Y., Yogosawa, S., Honda, R., Nishida, T., Yasuda, H., 2002. Sumoylation of Mdm2 by protein inhibitor of activated STAT (PIAS) and RanBP2 enzymes. J. Biol. Chem. 277, 50131–50136. https://doi.org/10.1074/jbc.M208319200.

Miyazaki, K., Fujita, T., Ozaki, T., Kato, C., Kurose, Y., Sakamoto, M., Kato, S., Goto, T., Itoyama, Y., Aoki, M., Nakagawara, A., 2004. NEDL1, a novel ubiquitin-protein isopeptide ligase for dishevelled-1, targets mutant superoxide dismutase-1. J. Biol. Chem. 279, 11327–11335. https://doi.org/10.1074/jbc.M312389200.

Morgan, A.R., Turic, D., Jehu, L., Hamilton, G., Hollingworth, P., Moskvina, V., Jones, L., Lovestone, S., Brayne, C., Rubinsztein, D.C., Lawlor, B., Gill, M., O'Donovan, M.C., Owen, M.J., Williams, J., 2007. Association studies of 23 positional/functional candidate genes on chromosome 10 in late-onset Alzheimer's disease. Am. J. Med. Genet. B Neuropsychiatr. Genet. 144, 762–770. https://doi.org/10.1002/ajmg.b.30509.

Mukaetova-Ladinska, E.B., Hurt, J., Jakes, R., Xuereb, J., Honer, W.G., Wischik, C.M., 2000. α-Synuclein inclusions in Alzheimer and Lewy body diseases. J. Neuropathol. Exp. Neurol. 59, 408–417. https://doi.org/10.1093/jnen/59.5.408.

Mulherkar, S.A., Sharma, J., Jana, N.R., 2009. The ubiquitin ligase E6-AP promotes degradation of α-synuclein. J. Neurochem. 110, 1955–1964. https://doi.org/10.1111/j.1471-4159.2009.06293.x.

Murata, H., Sakaguchi, M., Kataoka, K., Huh, N.H., 2013. SARM1 and TRAF6 bind to and stabilize PINK1 on depolarized mitochondria. Mol. Biol. Cell 24, 2772–2784. https://doi.org/10.1091/mbc.E13-01-0016.

Musiek, E.S., Holtzman, D.M., 2016. Mechanisms linking circadian clocks, sleep, and neurodegeneration. Science (80-.) 354, 1004–1008. https://doi.org/10.1126/science.aah4968.

Myeku, N., Duff, K.E., 2018. Targeting the 26S proteasome to protect against proteotoxic diseases. Trends Mol. Med. 24, 18–29. https://doi.org/10.1016/j.molmed.2017.11.006.

Myeku, N., Clelland, C.L., Emrani, S., Kukushkin, N.V., Yu, W.H., Goldberg, A.L., Duff, K.E., 2016. Tau-driven 26S proteasome impairment and cognitive dysfunction can be prevented early in disease by activating cAMP-PKA signaling. Nat. Med. 22, 46–53. https://doi.org/10.1038/nm.4011.

Nagamine, S., Kabuta, T., Furuta, A., Yamamoto, K., Takahashi, A., Wada, K., 2010. Deficiency of ubiquitin carboxy-terminal hydrolase-L1 (UCH-L1) leads to vulnerability to lipid peroxidation. Neurochem. Int. 57, 102–110. https://doi.org/10.1016/j.neuint.2010.04.015.

Nagashima, Y., Kowa, H., Tsuji, S., Iwata, A., 2011. FAT10 protein binds to polyglutamine proteins and modulates their solubility. J. Biol. Chem. 286, 29594–29600. https://doi.org/10.1074/jbc.M111.261032.

Naik, E., Webster, J.D., DeVoss, J., Liu, J., Suriben, R., Dixit, V.M., 2014. Regulation of proximal T cell receptor signaling and tolerance induction by deubiquitinase Usp9X. J. Exp. Med. 211, 1947–1955. https://doi.org/10.1084/jem.20140860.

Nakashima, H., Nguyen, T., Goins, W.F., Chiocca, E.A., 2015. Interferon-stimulated gene 15 (ISG15) and ISG15-linked proteins can associate with members of the selective autophagic process, histone deacetylase 6 (HDAC6) and SQSTM1/p62. J. Biol. Chem. 290, 1485–1495. https://doi.org/10.1074/jbc.M114.593871.

Nalbandian, A., Llewellyn, K.J., Nguyen, C., Yazdi, P.G., Kimonis, V.E., 2015. Rapamycin and chloroquine: the in vitro and in vivo effects of autophagy-modifying drugs show promising results in Valosin containing protein multisystem proteinopathy. PLoS One 10, e0122888. https://doi.org/10.1371/journal.pone.0122888.

Nath, S.R., Lieberman, A.P., 2017. The ubiquitination, disaggregation and proteasomal

degradation machineries in polyglutamine disease. Front. Mol. Neurosci. 10, 78. https://doi.org/10.3389/fnmol.2017.00078.
Navabi, S.P., Sarkaki, A., Mansouri, E., Badavi, M., Ghadiri, A., Farbood, Y., 2018. The effects of betulinic acid on neurobehavioral activity, electrophysiology and histological changes in an animal model of the Alzheimer's disease. Behav. Brain Res. 337, 99–106. https://doi.org/10.1016/j.bbr.2017.10.002.
Nemes, Z., Devreese, B., Steinert, P.M., Van Beeumen, J., Fésüs, L., 2004. Cross-linking of ubiquitin, HSP27, parkin, and α-synuclein by γ-glutamyl-ε-lysine bonds in Alzheimer's neurofibrillary tangles. FASEB J. 18, 1135–1137. https://doi.org/10.1096/fj.04-1493fje.
Nihira, N.T., Ogura, K., Shimizu, K., North, B.J., Zhang, J., Gao, D., Inuzuka, H., Wei, W., 2017. Acetylation-dependent regulation of MDM2 E3 ligase activity dictates its oncogenic function. Sci. Signal. 10, eaai8026. https://doi.org/10.1126/scisignal.aai8026.
Ochaba, J., Monteys, A.M., O'Rourke, J.G., Reidling, J.C., Steffan, J.S., Davidson, B.L., Thompson, L.M., 2016. PIAS1 regulates mutant huntingtin accumulation and Huntington's disease-associated phenotypes in vivo. Neuron 90, 507–520. https://doi.org/10.1016/j.neuron.2016.03.016.
Orlicky, S., Tang, X., Neduva, V., Elowe, N., Brown, E.D., Sicheri, F., Tyers, M., 2010. An allosteric inhibitor of substrate recognition by the SCF Cdc4 ubiquitin ligase. Nat. Biotechnol. 28, 733–737. https://doi.org/10.1038/nbt.1646.
Paiva, S.L., Da Silva, S.R., De Araujo, E.D., Gunning, P.T., 2018. Regulating the master regulator: controlling ubiquitination by thinking outside the active site. J. Med. Chem. 61, 405–421. https://doi.org/10.1021/acs.jmedchem.6b01346.
Pankiv, S., Clausen, T.H., Lamark, T., Brech, A., Bruun, J.A., Outzen, H., Øvervatn, A., Bjørkøy, G., Johansen, T., 2007. p62/SQSTM1 binds directly to Atg8/LC3 to facilitate degradation of ubiquitinated protein aggregates by autophagy*[S]. J. Biol. Chem. 282, 24131–24145. https://doi.org/10.1074/jbc.M702824200.
Patten, S.A., Aggad, D., Martinez, J., Tremblay, E., Petrillo, J., Armstrong, G.A., La Fontaine, A., Maios, C., Liao, M., Ciura, S., Wen, X.Y., Rafuse, V., Ichida, J., Zinman, L., Julien, J.P., Kabashi, E., Robitaille, R., Korngut, L., Parker, J.A., Drapeau, P., 2017. Neuroleptics as therapeutic compounds stabilizing neuromuscular transmission in amyotrophic lateral sclerosis. JCI Insight 2https://doi.org/10.1172/jci.insight.97152. pii: 97152.
Peterson, L.K., Fujinami, R.S., 2007. Inflammation, demyelination, neurodegeneration and neuroprotection in the pathogenesis of multiple sclerosis. J. Neuroimmunol. 184, 37–44. https://doi.org/10.1016/j.jneuroim.2006.11.015.
Popa-Wagner, A., Mitran, S., Sivanesan, S., Chang, E., Buga, A.M., 2013. ROS and brain diseases: the good, the bad, and the ugly. Oxid. Med. Cell. Longev. 2013, 963520. https://doi.org/10.1155/2013/963520.
Proctor, C.J., Gray, D.A., 2010. GSK3 and p53 — is there a link in Alzheimer's disease? Mol. Neurodegener. 5, 7. https://doi.org/10.1186/1750-1326-5-7.
Qin, C., Hu, Y., Zhou, B., Fernandez-Salas, E., Yang, C.Y., Liu, L., McEachern, D., Przybranowski, S., Wang, M., Stuckey, J., Meagher, J., Bai, L., Chen, Z., Lin, M., Yang, J., Ziazadeh, D.N., Xu, F., Hu, J., Xiang, W., Huang, L., Li, S., Wen, B., Sun, D., Wang, S., 2018. Discovery of QCA570 as an exceptionally potent and efficacious proteolysis targeting chimera (PROTAC) degrader of the bromodomain and extra-terminal (BET) proteins capable of inducing complete and durable tumor regression. J. Med. Chem. 61, 6685–6704. https://doi.org/10.1021/acs.jmedchem.8b00506.
Quik, M., Bordia, T., Zhang, D., Perez, X.A., 2015. Nicotine and nicotinic receptor drugs: potential for parkinson's disease and drug-induced movement disorders. Int. Rev. Neurobiol. 124, 247–271. https://doi.org/10.1016/bs.irn.2015.07.005.
Ragagnin, A.M.G., Shadfar, S., Vidal, M., Jamali, M.S., Atkin, J.D., 2019. Motor neuron susceptibility in ALS/FTD. Front. Neurosci. 13, 532. https://doi.org/10.3389/fnins.2019.00532.
Rajakumar, D., Alexander, M., Oommen, A., 2013. Oxidative stress, NF-κB and the ubiquitin proteasomal pathway in the pathology of calpainopathy. Neurochem. Res. 38, 2009–2018. https://doi.org/10.1007/s11064-013-1107-z.
Reddy, P.H., Oliver, D.M., 2019. Amyloid Beta and phosphorylated tau-induced defective autophagy and mitophagy in Alzheimer's disease. Cells 8, 488. https://doi.org/10.3390/cells8050488.
Regland, B., Lehmann, W., Abedini, I., Blennow, K., Jonsson, M., Karlsson, I., Sjögren, M., Wallin, A., Xilinas, M., Gottfries, C.G., 2001. Treatment of Alzheimer's disease with clioquinol. Dement. Geriatr. Cogn. Disord. 12, 408–414. https://doi.org/10.1159/000051288.
Ren, Z.Y., Liu, M.M., Xue, Y.X., Ding, Z.B., Xue, L.F., Zhai, S.Di, Lu, L., 2013. A critical role for protein degradation in the nucleus accumbens core in cocaine reward memory. Neuropsychopharmacology 38, 778–790. https://doi.org/10.1038/npp.2012.243.
Rial, D., Castro, A.A., Machado, N., Garção, P., Gonçalves, F.Q., Silva, H.B., Tomé, Â.R., Köfalvi, A., Corti, O., Raisman-Vozari, R., Cunha, R.A., Prediger, R.D., 2014. Behavioral phenotyping of Parkin-deficient mice: looking for early preclinical features of Parkinson's disease. PLoS One 9, e114216. https://doi.org/10.1371/journal.pone.0114216.
Richet, E., Pooler, A.M., Rodriguez, T., Novoselov, S.S., Schmidtke, G., Groettrup, M., Hanger, D.P., Cheetham, M.E., van der Spuy, J., 2012. NUB1 modulation of GSK3β reduces tau aggregation. Hum. Mol. Genet. 21, 5254–5267. https://doi.org/10.1093/hmg/dds376.
Ristic, G., Tsou, W.L., Todi, S.V., 2014. An optimal ubiquitin-proteasome pathway in the nervous system: the role of deubiquitinating enzymes. Front. Mol. Neurosci. 7, 72. https://doi.org/10.3389/fnmol.2014.00072.
Rodrigues, E.M., Scudder, S.L., Goo, M.S., Patrick, G.N., 2016. Aβ-induced synaptic alterations require the E3 ubiquitin ligase Nedd4-1. J. Neurosci. 36, 1590–1595. https://doi.org/10.1523/JNEUROSCI.2964-15.2016.
Rose, I., 2005. Ubiquitin at fox chase. Cell Death Differ. 102, 11575–11577. https://doi.org/10.1038/sj.cdd.4401710.
Rosen, K.M., Moussa, C.E.H., Lee, H.K., Kumar, P., Kitada, T., Qin, G., Fu, Q., Querfurth, H.W., 2010. Parkin reverses intracellular β-amyloid accumulation and its negative effects on proteasome function. J. Neurosci. Res. 88, 167–178. https://doi.org/10.1002/jnr.22178.
Rott, R., Szargel, R., Shani, V., Hamza, H., Savyon, M., Elghani, F.A., Bandopadhyay, R., Engelender, S., 2017. SUMOylation and ubiquitination reciprocally regulate α-synuclein degradation and pathological aggregation. Proc. Natl. Acad. Sci. U. S. A. 114, 13176–13181. https://doi.org/10.1073/pnas.1704351114.
Saito, R., Kaneko, M., Kitamura, Y., Takata, K., Kawada, K., Okuma, Y., Nomura, Y., 2014. Effects of oxidative stress on the solubility of HRD1, a ubiquitin ligase implicated in Alzheimer's disease. PLoS One 9, e94576. https://doi.org/10.1371/journal.pone.0094576.
Sari, Y., 2011. Huntington's disease: from mutant huntingtin protein to neurotrophic factor therapy. Int. J. Biomed. Sci. 7, 89–100.
Schipper-Krom, S., Juenemann, K., Reits, E.A.J., 2012. The ubiquitin-proteasome system in Huntington's disease: are proteasomes impaired, initiators of disease, or coming to the rescue? Biochem. Res. Int. 2012, 837015. https://doi.org/10.1155/2012/837015.
Schludi, M.H., May, S., Grässer, F.A., Rentzsch, K., Kremmer, E., Küpper, C., Klopstock, T., Ceballos-Baumann, A., Danek, A., Diehl-Schmid, J., Fassbender, K., Förstl, H., Kornhuber, J., Otto, M., Dieterich, M., Feuerecker, R., Giese, A., Klünemann, H., Kurz, A., Levin, J., Lorenzl, S., Meyer, T., Nübling, G., Roeber, S., 2015. Distribution of dipeptide repeat proteins in cellular models and C9orf72 mutation cases suggests link to transcriptional silencing. Acta Neuropathol. 130, 537–555. https://doi.org/10.1007/s00401-015-1450-z.
Schneekloth, J.S., Fonseca, F.N., Koldobskiy, M., Mandal, A., Deshaies, R., Sakamoto, K., Crews, C.M., 2004. Chemical genetic control of protein levels: selective in vivo targeted degradation. J. Am. Chem. Soc. 126, 3748–3754. https://doi.org/10.1021/ja039025z.
Schou, J., Prockop, L.D., Dahlström, G., Rohde, C., 1977. Penetration of delta-9-tetrahydrocannabinol and 11-OH-delta-9-tetrahydrocannabinol through the Blood-brain Barrier. Acta Pharmacol. Toxicol. (Copenh) 41, 33–38. https://doi.org/10.1111/j.1600-0773.1977.tb02120.x.
Schreiber, J., Végh, M.J., Dawitz, J., Kroon, T., Loos, M., Labonté, D., Li, K.W., Van Nierop, P., Van Diepen, M.T., De Zeeuw, C.I., Kneussel, M., Meredith, R.M., Smit, A.B., Van Kesteren, R.E., 2015. Ubiquitin ligase TRIM3 controls hippocampal plasticity and learning by regulating synaptic γ-actin levels. J. Cell Biol. 211, 569–586. https://doi.org/10.1083/jcb.201506048.
Schweiger, S., Matthes, F., Posey, K., Kickstein, E., Weber, S., Hettich, M.M., Pfurtscheller, S., Ehninger, D., Schneider, R., Krauß, S., 2017. Resveratrol induces dephosphorylation of Tau by interfering with the MID1-PP2A complex. Sci. Rep. 7, 13753. https://doi.org/10.1038/s41598-017-12974-4.
Sellebjerg, F., Krakauer, M., Khademi, M., Olsson, T., Sørensen, P.S., 2012. FOXP3, CBLB and ITCH gene expression and cytotoxic T lymphocyte antigen 4 expression on CD4 + CD25 high T cells in multiple sclerosis. Clin. Exp. Immunol. 170, 149–155. https://doi.org/10.1111/j.1365-2249.2012.04654.x.
Semmler, S., Gagné, M., Garg, P., Pickles, S.R., Baudouin, C., Hamon-Keromen, E., Destroismaisons, L., Khalfallah, Y., Chaineau, M., Caron, E., Bayne, A.N., Trempe, J.-F., Cashman, N.R., Star, A.T., Haqqani, A.S., Durcan, T.M., Meiering, E.M., Robertson, J., Grandvaux, N., Plotkin, S.S., McBride, H.M., Velde, C.V., 2019. The E3 ubiquitin ligase TRAF6 is a novel interacting protein of amyotrophic lateral sclerosis-linked misfolded SOD1. bioRxiv, 780460. https://doi.org/10.1101/780460.
Serna, S., Xolalpa, W., Lang, V., Aillet, F., England, P., Reichardt, N., Rodriguez, M.S., 2016. Efficient monitoring of protein ubiquitylation levels using TUBEs-based microarrays. FEBS Lett. 590, 2748–2756. https://doi.org/10.1002/1873-3468.12289.
Shin, J.H., Ko, H.S., Kang, H., Lee, Yunjong, Lee, Yun Il, Pletinkova, O., Troconso, J.C., Dawson, V.L., Dawson, T.M., 2011. PARIS (ZNF746) repression of PGC-1α contributes to neurodegeneration in Parkinson's disease. Cell 144, 689–702. https://doi.org/10.1016/j.cell.2011.02.010.
Silva, M.C., Ferguson, F.M., Cai, Q., Donovan, K.A., Nandi, G., Patnaik, D., Zhang, T., Huang, H.T., Lucente, D.E., Dickerson, B.C., Mitchison, T.J., Fischer, E.S., Gray, N.S., Haggarty, S.J., 2019. Targeted degradation of aberrant tau in frontotemporal dementia patient-derived neuronal cell models. eLife 8, e45457. https://doi.org/10.7554/eLife.45457.
Singh, B.K., Vatsa, N., Kumar, V., Shekhar, S., Sharma, A., Jana, N.R., 2017. Ube3a deficiency inhibits amyloid plaque formation in APPswe/PS1δE9 mouse model of Alzheimer's disease. Hum. Mol. Genet. 26, 4042–4054. https://doi.org/10.1093/hmg/ddx295.
Sivandzade, F., Prasad, S., Bhalerao, A., Cucullo, L., 2019. NRF2 and NF-κB interplay in cerebrovascular and neurodegenerative disorders: molecular mechanisms and possible therapeutic approaches. Redox Biol. 21, 101059. https://doi.org/10.1016/j.redox.2018.11.017.
Song, T., Liang, S., Liu, J., Zhang, T., Yin, Y., Geng, C., Gao, S., Feng, Y., Xu, H., Guo, D., Roberts, A., Gu, Y., Cang, Y., 2018. CRL4 antagonizes SCFFbxo7-mediated turnover of cereblon and BK channel to regulate learning and memory. PLoS Genet. 14, e1007165. https://doi.org/10.1371/journal.pgen.1007165.
Sugiura, A., Nagashima, S., Tokuyama, T., Amo, T., Matsuki, Y., Ishido, S., Kudo, Y., McBride, H.M., Fukuda, T., Matsushita, N., Inatome, R., Yanagi, S., 2013. MITOL regulates endoplasmic reticulum-mitochondria contacts via mitofusin2. Mol. Cell 51, 20–34. https://doi.org/10.1016/j.molcel.2013.04.023.
Sultana, R., Boyd-Kimball, D., Poon, H.F., Cai, J., Pierce, W.M., Klein, J.B., Merchant, M., Markesbery, W.R., Butterfield, D.A., 2006. Redox proteomics identification of oxidized proteins in Alzheimer's disease hippocampus and cerebellum: an approach to understand pathological and biochemical alterations in AD. Neurobiol. Aging 27, 1564–1576. https://doi.org/10.1016/j.neurobiolaging.2005.09.021.
Sun, J., Zhu, G., Liu, Y., Standley, S., Ji, A., Tunuguntla, R., Wang, Y., Claus, C., Luo, Y.,

Baudry, M., Bi, X., 2015. UBE3A regulates synaptic plasticity and learning and memory by controlling SK2 channel endocytosis. Cell Rep. 12, 449–461. https://doi.org/10.1016/j.celrep.2015.06.023.
Sweeney, P., Park, H., Baumann, M., Dunlop, J., Frydman, J., Kopito, R., McCampbell, A., Leblanc, G., Venkateswaran, A., Nurmi, A., Hodgson, R., 2017. Protein misfolding in neurodegenerative diseases: implications and strategies. Transl. Neurodegener. 6, 6. https://doi.org/10.1186/s40035-017-0077-5.
Szargel, R., Rott, R., Engelender, S., 2008. Synphilin-1 isoforms in Parkinson's disease: regulation by phosphorylation and ubiquitylation. Cell. Mol. Life Sci. 65, 80–88. https://doi.org/10.1007/s00018-007-7343-0.
Tashiro, Y., Urushitani, M., Inoue, H., Koike, M., Uchiyama, Y., Komatsu, M., Tanaka, K., Yamazaki, M., Abe, M., Misawa, H., Sakimura, K., Ito, H., Takahashia, R., 2012. Motor neuron-specific disruption of proteasomes, but not autophagy, replicates amyotrophic lateral sclerosis. J. Biol. Chem. 287, 42984–42994. https://doi.org/10.1074/jbc.M112.417600.
Tetzlaff, J.E., Putcha, P., Outeiro, T.F., Ivanov, A., Berezovska, O., Hyman, B.T., McLean, P.J., 2008. CHIP targets toxic α-synuclein oligomers for degradation. J. Biol. Chem. 283, 17962–17968. https://doi.org/10.1074/jbc.M802283200.
Thompson, L.M., Aiken, C.T., Kaltenbach, L.S., Agrawal, N., Illes, K., Khoshnan, A., Martinez-Vincente, M., Arrasate, M., O'Rourke, J.G., Khashwji, H., Lukacsovich, T., Zhu, Y.Z., Lau, A.L., Massey, A., Hayden, M.R., Zeitlin, S.O., Finkbeiner, S., Green, K.N., LaFerla, F.M., Bates, G., Huang, L., Patterson, P.H., Lo, D.C., Cuervo, A.M., Marsh, J.L., Steffan, J.S., 2009. IKK phosphorylates Huntingtin and targets it for degradation by the proteasome and lysosome. J. Cell Biol. 187, 1083–1099. https://doi.org/10.1083/jcb.200909067.
Tramutola, A., Di Domenico, F., Barone, E., Perluigi, M., Butterfield, D.A., 2016. It is all about (U)biquitin: role of altered ubiquitin-proteasome system and UCHL1 in alzheimer disease. Oxid. Med. Cell. Longev. 2016, 2756068. https://doi.org/10.1155/2016/2756068.
Tramutola, A., Triani, F., Di Domenico, F., Barone, E., Cai, J., Klein, J.B., Perluigi, M., Butterfield, D.A., 2018. Poly-ubiquitin profile in Alzheimer disease brain. Neurobiol. Dis. 118, 129–141. https://doi.org/10.1016/j.nbd.2018.07.006.
Tsai, Y.M., Chien, C.F., Lin, L.C., Tsai, T.H., 2011. Curcumin and its nano-formulation: the kinetics of tissue distribution and blood-brain barrier penetration. Int. J. Pharm. 416, 331–338. https://doi.org/10.1016/j.ijpharm.2011.06.030.
Tsang, A.H.K., Lee, Y.I.L., Ko, H.S., Savitt, J.M., Pletnikova, O., Troncoso, J.C., Dawson, V.L., Dawson, T.M., Chung, K.K.K., 2009. S-Nitrosylation of XIAP compromises neuronal survival in Parkinson's disease. Proc. Natl. Acad. Sci. U. S. A. 106, 4900–4905. https://doi.org/10.1073/pnas.0810595106.
Twohig, D., Nielsen, H.M., 2019. α-Synuclein in the pathophysiology of Alzheimer's disease. Mol. Neurodegener. 14, 23. https://doi.org/10.1186/s13024-019-0320-x.
Uchida, T., Tamaki, Y., Ayaki, T., Shodai, A., Kaji, S., Morimura, T., Banno, Y., Nishitsuji, K., Sakashita, N., Maki, T., Yamashita, H., Ito, H., Takahashi, R., Urushitani, M., 2016. CUL2-mediated clearance of misfolded TDP-43 is paradoxically affected by VHL in oligodendrocytes in ALS. Sci. Rep. 6, 19118. https://doi.org/10.1038/srep19118.
Umeda, T., Tomiyama, T., Sakama, N., Tanaka, S., Lambert, M.P., Klein, W.L., Mori, H., 2011. Intraneuronal amyloid β oligomers cause cell death via endoplasmic reticulum stress, endosomal/lysosomal leakage, and mitochondrial dysfunction in vivo. J. Neurosci. Res. 89, 1031–1042. https://doi.org/10.1002/jnr.22640.
United Nations, 2019. World Population Prospects 2019]. United Nations. Dep. Econ. Soc. Aff. World Popul. Prospect. 2019. Report ST/. . https://population.un.org/wpp/Publications/Files/W.
Upadhya, S.C., Smith, T.K., Hegde, A.N., 2004. Ubiquitin-proteasome-mediated CREB repressor degradation during induction of long-term facilitation. J. Neurochem. 91, 210–219. https://doi.org/10.1111/j.1471-4159.2004.02707.x.
Uranga, R.M., Katz, S., Salvador, G.A., 2013. Enhanced phosphatidylinositol 3-kinase (PI3K)/Akt signaling has pleiotropic targets in hippocampal neurons exposed to iron-induced oxidative stress. J. Biol. Chem. 288, 19773–19784. https://doi.org/10.1074/jbc.M113.457622.
Vaden, J.H., Bhattacharyya, B.J., Chen, P.C., Watson, J.A., Marshall, A.G., Phillips, S.E., Wilson, J.A., King, G.D., Miller, R.J., Wilson, S.M., 2015. Ubiquitin-specific protease 14 regulates c-Jun N-terminal kinase signaling at the neuromuscular junction. Mol. Neurodegener. 10, 3. https://doi.org/10.1186/1750-1326-10-3.
Vilchez, D., Saez, I., Dillin, A., 2014. The role of protein clearance mechanisms in organismal ageing and age-related diseases. Nat. Commun. 5, 5659. https://doi.org/10.1038/ncomms6659.
von Grabowiecki, Y., Abreu, P., Blanchard, O., Palamiuc, L., Benosman, S., Mériaux, S., Devignot, V., Gross, I., Mellitzer, G., Gonzalez de Aguilar, J.L., Gaiddon, C., 2016. Transcriptional activator TAp63 is upregulated in muscular atrophy during ALS and induces the pro-atrophic ubiquitin ligase Trim63. eLife 5, e10528. https://doi.org/10.7554/eLife.10528.
Walters, B.J., Hallengren, J.J., Theile, C.S., Ploegh, H.L., Wilson, S.M., Dobrunz, L.E., 2014. A catalytic independent function of the deubiquitinating enzyme USP14 regulates hippocampal synaptic short-term plasticity and vesicle number. J. Physiol. 592, 571–586. https://doi.org/10.1113/jphysiol.2013.266015.
Wang, Y., Branicky, R., Stepanyan, Z., Carroll, M., Guimond, M.P., Hihi, A., Hayes, S., McBride, K., Hekimi, S., 2009. The anti-neurodegeneration drug clioquinol inhibits the aging-associated protein CLK-1. J. Biol. Chem. 284, 314–323. https://doi.org/10.1074/jbc.M807579200.
Wang, B., Zeng, L., Merillat, S.A., Fischer, S., Ochaba, J., Thompson, L.M., Barmada, S.J., Scaglione, K.M., Paulson, H.L., 2018a. The ubiquitin conjugating enzyme Ube2W regulates solubility of the Huntington's disease protein, huntingtin. Neurobiol. Dis. 109, 127–136. https://doi.org/10.1016/j.nbd.2017.10.002.
Wang, Q., Jiao, F., Zhang, P., Yan, J., Zhang, Z., He, F., Zhang, Q., Lv, Z., Peng, X., Cai, H., Tian, B., 2018b. CDK5-mediated phosphorylation-dependent ubiquitination and degradation of E3 ubiquitin ligases GP78 accelerates neuronal death in Parkinson's disease. Mol. Neurobiol. 55, 3709–3717. https://doi.org/10.1007/s12035-017-0579-2.
Weinreb, O., Mandel, S., Bar-Am, O., Yogev-Falach, M., Avramovich-Tirosh, Y., Amit, T., Youdim, M.B.H., 2009. Multifunctional neuroprotective derivatives of Rasagiline as anti-Alzheimer's disease drugs. Neurotherapeutics 6, 163–174. https://doi.org/10.1016/j.nurt.2008.10.030.
Wertz, I.E., Newton, K., Seshasayee, D., Kusam, S., Lam, C., Zhang, J., Popovych, N., Helgason, E., Schoeffler, A., Jeet, S., Ramamoorthi, N., Kategaya, L., Newman, R.J., Horikawa, K., Dugger, D., Sandoval, W., Mukund, S., Zindal, A., Martin, F., Quan, C., Tom, J., Fairbrother, W.J., Townsend, M., Warming, S., Devoss, J., Liu, J., Dueber, E., Caplazi, P., Lee, W.P., Goodnow, C.C., Balazs, M., Yu, K., Kolumam, G., Dixit, V.M., 2015. Phosphorylation and linear ubiquitin direct A20 inhibition of inflammation. Nature 528, 370–375. https://doi.org/10.1038/nature16165.
Widagdo, J., Chai, Y.J., Ridder, M.C., Chau, Y.Q., Johnson, R.C., Sah, P., Huganir, R.L., Anggono, V., 2015. Activity-dependent ubiquitination of GluA1 and GluA2 regulates AMPA receptor intracellular sorting and degradation. Cell Rep. 10, 783–795. https://doi.org/10.1016/j.celrep.2015.01.015.
Widagdo, J., Guntupalli, S., Jang, S.E., Anggono, V., 2017. Regulation of AMPA receptor trafficking by protein ubiquitination. Front. Mol. Neurosci. 10, 347. https://doi.org/10.3389/fnmol.2017.00347.
Williams, K.L., Topp, S., Yang, S., Smith, B., Fifita, J.A., Warraich, S.T., Zhang, K.Y., Farrawell, N., Vance, C., Hu, X., Chesi, A., Leblond, C.S., Lee, A., Rayner, S.L., Sundaramoorthy, V., Dobson-Stone, C., Molloy, M.P., Van Blitterswijk, M., Dickson, D.W., Petersen, R.C., Graff-Radford, N.R., Boeve, B.F., Murray, M.E., Pottier, C., Don, E., Winnick, C., McCann, E.P., Hogan, A., Daoud, H., Levert, A., Dion, P.A., Mitsui, J., Ishiura, H., Takahashi, Y., Goto, J., Kost, J., Gellera, C., Gkazi, A.S., Miller, J., Stockton, J., Brooks, W.S., Boundy, K., Polak, M., Muñoz-Blanco, J.L., Esteban-Pérez, J., Rábano, A., Hardiman, O., Morrison, K.E., Ticozzi, N., Silani, V., De Belleroche, J., Glass, J.D., Kwok, J.B.J., Guillemin, G.J., Chung, R.S., Tsuji, S., Brown, R.H., García-Redondo, A., Rademakers, R., Landers, J.E., Gitler, A.D., Rouleau, G.A., Cole, N.J., Yerbury, J.J., Atkin, J.D., Shaw, C.E., Nicholson, G.A., Blair, I.P., 2016. CCNF mutations in amyotrophic lateral sclerosis and frontotemporal dementia. Nat. Commun. 7, 11253. https://doi.org/10.1038/ncomms11253.
Winkle, C.C., Olsen, R.H.J., Kim, H., Moy, S.S., Song, J., Gupton, S.L., 2016. Trim9 deletion alters the morphogenesis of developing and adult-born hippocampal neurons and impairs spatial learning and memory. J. Neurosci. 36, 4940–4958. https://doi.org/10.1523/JNEUROSCI.3876-15.2016.
Wong, H.K., Bauer, P.O., Kurosawa, M., Goswami, A., Washizu, C., Machida, Y., Tosaki, A., Yamada, M., Knöpfel, T., Nakamura, T., Nukina, N., 2008. Blocking acid-sensing ion channel 1 alleviates Huntington's disease pathology via an ubiquitin-proteasome system-dependent mechanism. Hum. Mol. Genet. 17, 3223–3235. https://doi.org/10.1093/hmg/ddn218.
Xu, P., Duong, D.M., Seyfried, N.T., Cheng, D., Xie, Y., Robert, J., Rush, J., Hochstrasser, M., Finley, D., Peng, J., 2009. Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation. Cell 137, 133–145. https://doi.org/10.1016/j.cell.2009.01.041.
Xu, J., Zhang, X.Z., Zhang, Y.J., Li, Xiu Ming, Cai, Z.L., Li, Xiao Min, 2015. Silencing of SIAH1 in SH-SY5Y affects α-synuclein degradation pathway. Int. J. Clin. Exp. Pathol. 8, 12885–12892.
Xu, Y., Zhao, F., Qiu, Q., Chen, K., Wei, J., Kong, Q., Gao, B., Melo-Cardenas, J., Zhang, B., Zhang, Jinping, Song, J., Zhang, D.D., Zhang, Jianing, Fan, Y., Li, H., Fang, D., 2016. The ER membrane-anchored ubiquitin ligase Hrd1 is a positive regulator of T-cell immunity. Nat. Commun. 7, 12073. https://doi.org/10.1038/ncomms12073.
Yan, C., Xu, W., Huang, Y., Li, M., Shen, Y., You, H., Liang, X., 2016. HRD1-mediated IGF-1R ubiquitination contributes to renal protection of resveratrol in db/db mice. Mol. Endocrinol. 30, 600–613. https://doi.org/10.1210/me.2015-1277.
Yao, I., Takagi, H., Ageta, H., Kahyo, T., Sato, S., Hatanaka, K., Fukuda, Y., Chiba, T., Morone, N., Yuasa, S., Inokuchi, K., Ohtsuka, T., MacGregor, G.R., Tanaka, K., Setou, M., 2007. SCRAPPER-dependent ubiquitination of active zone protein RIM1 regulates synaptic vesicle release. Cell 130, 943–957. https://doi.org/10.1016/j.cell.2007.06.052.
Yao, I., Takao, K., Miyakawa, T., Ito, S., Setou, M., 2011. Synaptic E3 ligase SCRAPPER in contextual fear conditioning: extensive behavioral phenotyping of Scrapper heterozygote and overexpressing mutant mice. PLoS One 6, e17317. https://doi.org/10.1371/journal.pone.0017317.
Yoon, S.O., Park, D.J., Ryu, J.C., Ozer, H.G., Tep, C., Shin, Y.J., Lim, T.H., Pastorino, L., Kunwar, A.J., Walton, J.C., Nagahara, A.H., Lu, K.P., Nelson, R.J., Tuszynski, M.H., Huang, K., 2012. JNK3 perpetuates metabolic stress induced by aβ peptides. Neuron 75, 824–837. https://doi.org/10.1016/j.neuron.2012.06.024.
Yoshida, Y., Saeki, Y., Murakami, A., Kawawaki, J., Tsuchiya, H., Yoshihara, H., Shindo, M., Tanaka, K., 2015. A comprehensive method for detecting ubiquitinated substrates using TR-TUBE. Proc. Natl. Acad. Sci. U. S. A. 112, 4630–4635. https://doi.org/10.1073/pnas.1422313112.
Yue, W., Chen, Z., Liu, H., Yan, Chen, Chen, M., Feng, D., Yan, Chaojun, Wu, H., Du, L., Wang, Y., Liu, J., Huang, X., Xia, L., Liu, L., Wang, X., Jin, H., Wang, J., Song, Z., Hao, X., Chen, Q., 2014. A small natural molecule promotes mitochondrial fusion through inhibition of the deubiquitinase USP30. Cell Res. 24, 482–496. https://doi.org/10.1038/cr.2014.20.
Zhang, M., Deng, Y., Luo, Y., Zhang, S., Zou, H., Cai, F., Wada, K., Song, W., 2012. Control of BACE1 degradaton and APP processing by ubiquitin carboxyl-terminal hydrolase L1. J. Neurochem. 120, 1129–1138. https://doi.org/10.1111/j.1471-4159.2011.07644.x.
Zhang, M., Cai, F., Zhang, Shuting, Zhang, Si, Song, W., 2014. Overexpression of ubiquitin carboxyl-terminal hydrolase L1 (UCHL1) delays Alzheimer's progression in vivo. Sci. Rep. 4, 7298. https://doi.org/10.1038/srep07298.

Zhang, J., Zhang, R., Zhan, Z., Li, X., Zhou, F., Xing, A., Jiang, C., Chen, Y., An, L., 2017. Beneficial effects of sulforaphane treatment in Alzheimer's disease may be mediated through reduced HDAC1/3 and increased P75NTR expression. Front. Aging Neurosci. 9, 121. https://doi.org/10.3389/fnagi.2017.00121.

Zhao, T., Hong, Y., Li, S., Li, X.J., 2016. Compartment-dependent degradation of mutant huntingtin accounts for its preferential accumulation in neuronal processes. J. Neurosci. 36, 8317–8328. https://doi.org/10.1523/JNEUROSCI.0806-16.2016.

Zheng, Z., Keifer, J., 2009. PKA has a critical role in synaptic delivery of GluR1- and GluR4-containing AMPARs during initial stages of acquisition of in vitro classical conditioning. J. Neurophysiol. 101, 2539–2549. https://doi.org/10.1152/jn.91282.2008.

Zheng, Q., Huang, T., Zhang, L., Zhou, Y., Luo, H., Xu, H., Wang, X., 2016. Dysregulation of ubiquitin-proteasome system in neurodegenerative diseases. Front. Aging Neurosci. 8, 303. https://doi.org/10.3389/fnagi.2016.00303.

Zhong, L., Tan, Y., Zhou, A., Yu, Q., Zhou, J., 2005. Ring finger ubiquitin-protein isopeptide ligase Nrdp1/FLRF regulates parkin stability and activity. J. Biol. Chem. 280, 9425–9430. https://doi.org/10.1074/jbc.M408955200.

Zhou, L., Wang, H., Ren, H., Hu, Q., Ying, Z., Wang, G., 2015. Bcl-2 decreases the affinity of SQSTM1/p62 to poly-ubiquitin chains and suppresses the aggregation of misfolded protein in neurodegenerative disease. Mol. Neurobiol. 52, 1180–1189. https://doi.org/10.1007/s12035-014-8908-1.

Zhou, Q., Chen, B., Wang, X., Wu, L., Yang, Y., Cheng, X., Hu, Z., Cai, X., Yang, J., Sun, X., Lu, W., Yan, H., Chen, J., Ye, J., Shen, J., Cao, P., 2016. Sulforaphane protects against rotenone-induced neurotoxicity in vivo: involvement of the mTOR, Nrf2, and autophagy pathways. Sci. Rep. 6, 32206. https://doi.org/10.1038/srep32206.

Zhou, B., Hu, J., Xu, F., Chen, Z., Bai, L., Fernandez-Salas, E., Lin, M., Liu, L., Yang, C.Y., Zhao, Y., McEachern, D., Przybranowski, S., Wen, B., Sun, D., Wang, S., 2018. Discovery of a small-molecule degrader of Bromodomain and extra-terminal (BET) proteins with picomolar cellular potencies and capable of achieving tumor regression. J. Med. Chem. 61, 462–481. https://doi.org/10.1021/acs.jmedchem.6b01816.

Zhu, W., Xie, W., Pan, T., Xu, P., Fridkin, M., Zheng, H., Jankovic, J., Youdim, M.B.H., Le, W., 2007. Prevention and restoration of lactacystin-induced nigrostriatal dopamine neuron degeneration by novel brain-permeable iron chelators. FASEB J. 21, 3835–3844. https://doi.org/10.1096/fj.07-8386com.

Zucchelli, S., Marcuzzi, F., Codrich, M., Agostoni, E., Vilotti, S., Biagioli, M., Pinto, M., Carnemolla, A., Santoro, C., Gustincich, S., Persichetti, F., 2011. Tumor necrosis factor receptor-associated factor 6 (TRAF6) associates with Huntingtin protein and promotes its atypical ubiquitination to enhance aggregate formation. J. Biol. Chem. 286, 25108–25117. https://doi.org/10.1074/jbc.M110.187591.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/342927571

Women Empowerment in India: A Critical Analysis

Article · June 2020

CITATIONS
0

READS
7

2 authors:



Seema Singh
Delhi Technological University
49 PUBLICATIONS **64** CITATIONS
SEE PROFILE



Antra Singh
Delhi Technological University
4 PUBLICATIONS **4** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

University & SME's collaboration View project

Enhancing Employability Skill of Girls and Young Women View project

All content following this page was uploaded by Seema Singh on 14 July 2020.

The user has requested enhancement of the downloaded file.

ISSN:2320-0693
Vol-19-Issue-44-June-2020

Women Empowerment in India: A Critical Analysis

Prof. Seema Singh* and Dr. Antra Singh,**

** Professor of Economics and ** ICSSR Post Doctoral Fellow,*

Delhi Technological University, Delhi- 110 042, INDIA

Abstract

Being a traditional patriarchal society, women have been given a secondary status which is reflected in the economic, social and political spheres. However, women equality and empowerment has always remained a priority area and has been taken utmost care by stake holders. The paper critically investigates the Indian status among other countries and tries to find out preparedness to achieve Sustainable development Goal -5 of the United Nations. The paper develops argument on the basis of secondary sources as review of existing literature published in journal, books, reports of various, NGOs, Government and international organisations and websites. The paper critically examines women empowerment in India, various models and dimensions. The paper discusses constitutional safe guards as well as plans and programmes by the government and their implementation, indicators of women empowerment. However, the country ranks low while comparing with other countries. There is need of reassessing and modifying programmes to achieve SDG-5 by 2030.

Keywords: Women Empowerment; Gender Equality; Government; Social; Rights

This is a preliminary study by the University Women Association of Delhi for the project entitled, "Enhancing Employability Skills for Empowerment of Girls and Young Women" funded by Graduate Women International, Geneva.

1. Introduction

Traditionally being a patriarchal society, women have secondary role to play in every household in India. However, issues related to welfare of women have always been a priority among policy makers since independence, though the path has witnessed repositioning with time and as per the requirement from the Fifth Five Year Plan (1974-78) onwards, there has been a remarkable shift in the approach to women's issues from welfare to development. India has also endorsed various global efforts as the the Mexico Plan of Action (1975), the Nairobi Forward Looking Strategies (1985), Convention on Elimination of All Forms of Discrimination Against Women

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

(CEDAW) in 1993, the Beijing Declaration as well as the Platform for Action (1995) and the Outcome Document adopted by the UNGA Session on Gender Equality and Development and Peace for the 21st century. In this background, the paper discusses status of women empowerment in India and its status at the international spectrum with regard to achievement of SDG Goal 5 of the United Nations by 2030. The plan of the study is as follows. The next section deals with the review of literature and discusses the concept of women empowerment with supportive models alongwith dimensions. Section 3 provides the research objectives and methodology of the study. Section 4 analyses various legislations and government programmes for women empowerment, variegated initiatives taken for women empowerment by different N.G.O.s and corporate sector, international comparison regarding status of women empowerment followed by Section 5 which provides suggestions to meet the challenges and path forward to move ahead.

2. Review of Literature

Empowerment refers to a situation where the powerless gain greater control over resources and ideologies. It has been associated with terms such as autonomy, power, status and agency. The Indian constitution has very clearly given an equal level playing ground to women and has directed authorities to frame rules and regulations to safe guard the right. However, feminist scholars during 1970s as a way to challenge patriarchy, as a radical approach concerned with transforming power relations in favour of women's rights and gender equality It in 1980s (Batliwala, 1993, 2007) and as an individual process of self-transformation during 1990s (Batliwala, 1993; Kabeer, 1994; Rowlands, 1997; Sen, 1997).They highlight the complex reciprocal relationship between women's self-understanding (Kabeer, 1994) and capacity for self-expression (Sen, 1997), as well as women's access to and control over material resources.

The decade of nineties may be considered as an impertinent phase for women empowerment throughout the world as most of the world conferences highlighted towards seizing the opportunities. It accentuated issues of reproductive and sexual rights, violence against women and gender inequality. Even among the eight the Millennium Development Goals (MDG) of UN developed in 2000 was targeting to be achieved by 2015, the Goal three was to promote gender equality and empower women. Women empowerment is the defined as the change in the context of a women's life, which enables her increased capacity for leading a fulfilling human life. It gets reflected both in external qualities (health, mobility, education and awareness, status in the

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

family, participation in decision making) and internal qualities (self awareness and self confidence) Human Development in South Asia (2000) (Mathew, 2003). United Nations Development Fund for Women (UNIFEM) has been re-branded as the UN Entity for Gender Equality and the empowerment of Women.

2.1. Models of Women Empowerment

Three Dimensional Model

Women empowerment is considered as a process that takes place over time, making women agents who formulate choices, control resources and make strategic life choices (Lee-Rife, 2010). She emphasised on the strategic life choices (choice of livelihood, marital decision and child bearing decision) which are considered as first-order choices and how these impact second-order choice (inculcating values in children, daily household decision making and managing well being of the family) that are less consequential.

Figure.1 Three Dimensional Model by Kabeer



(Source: Kabeer, 1999)

The researcher emphasised having access to resources (material, human and social in form) is an obligatory pre-condition of empowerment as resources increase the ability to exercise choice (Kabeer, 1999). Kabeer emphasised that merely having access to resources is not sufficient women must also have the ability to identify and utilise the resources (Malhotra, Schuler and Boender, 2002). As per Kabeer, empowerment is both an end and a means to end for example one status is attaining education another status is acquiring gainful employment and a further step to it could be control over decision making related to self or family is the outcome of

Copyright © 2020Author

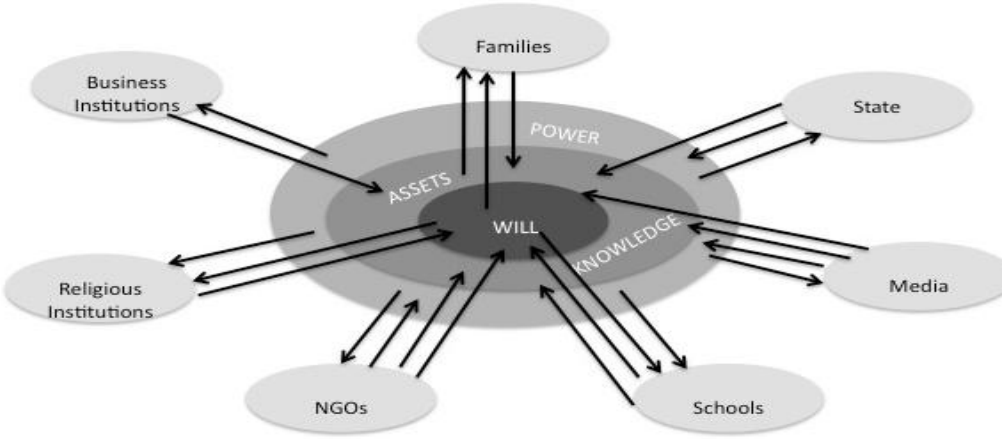
ISSN:2320-0693
Vol-19-Issue-44-June-2020

empowerment. Thus attaining education can be an end or it can also be a means to end but the final outcome of empowerment is having the capabilities of achievement.

The Commission on Women and Development's Conceptual Framework

The Commission on Women and Development, an International Aid Agency, the Government of Belgium designed an empowerment framework and methodology influenced by Kabeer, Longwez, Rowlands and Leon and the DAWN women's movement. The commission was of the view that empowerment occurred at two levels i.e. the individual level and the collective level. But the earlier model primarily highlighted the individual aspects of empowerment and identified the four core concepts of empowerment being assets, knowledge, will and capacity. This model splits resources into assets (material resources that give one greater economic power) and knowledge (provides people with techniques, training, literacy and critical analysis skills). Will is defined as psychological strength or power within to make one's own choices followed by capacity which is described as having the opportunities to use resources (assets, knowledge and will), make decisions and take on responsibility.

Figure 2. Conceptual Framework by the Commission on Women and Development



(Source: Commission of Women and Development, 2007)

The figure illustrates empowerment process as a catalyst for social change depicting will, assets and knowledge in concentric circles surrounded by power (power with others and power within). It also considers the institutional context comprising of family, government, media, educational institutions, NGOs, religious institutions and business institutions. The arrows pointing in and outward illustrates the potential to affect the components and not all arrows are same in weight.

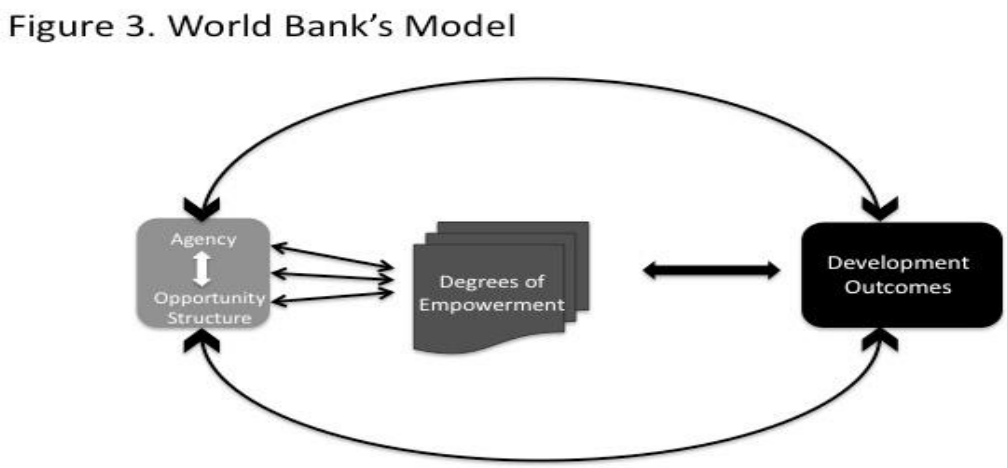
Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

World Bank's Agency and Opportunity Structure Model – Ruth Alsop and Nina Heinsohn

The capacity to make effective choices is dependent on two major factors contributing to empowerment and they are agency and opportunity structure. In this model assets are the indicators of agency and are characterised as psychological, informational, organisational, material, social, financial or human in nature. The other pre-condition in this model is the opportunity structure defined as the informal and formal institutions in which the individuals and the groups consider and includes laws, regulatory frameworks and norms governing behaviour, arguing that the opportunity structure determines who has access to assets and whether they can use their assets. The author hypothesizes that agency and opportunity structure are associated with the degrees (measured by the existence of choice, the use of choice and the achievement of choice) of empowerment of individuals and groups.

Figure 3. World Bank Model



(Source: Alsop and Heinsohn, 2005)

As discussed, researchers viewed women empowerment as a process with numerous inputs and factors that influence the process and finally, the outcome of these dimensions needs to be measured to quantify women empowerment. Numerous indicators are required to capture different dimensions of women's agency and empowerment like sexual and reproductive health, household decision-making, economic activities, participation in community decision making bodies and national politics and gender norms and women's motivations for their choices and

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

actions (Hanmer, 2015). These dimensions are mostly identical with minor alterations in sub divisions which depend on the focus of study and can be elaborated accordingly.

2.2. Dimensions of Women Empowerment

The Human Development Index was introduced for the first time by United Nations Development Program (UNDP) in 1990 to measure the socio economic progress of a nation. The 1995 report emphasised on empowering women as it was required for the overall human development. To study the statistics related to human development and the share of women, two indices were introduced namely Gender Development Index (GDI) that takes into account inequality in the achievements between women and men and Gender Empowerment Measure (GEM) which measures the extent to which a women has influence in decision-making in politics, in professional life and in organisations.

The Women Empowerment Matrix (Wieringa, 1994) was contrived to map out the general gender issues at national level and to outline the contours of dimensions of empowerment and to demonstrate the interlinkages of gender issues related to specific issues like education, health and women labour. The women empowerment matrix is a tool to aid the researcher in getting an overview of the various levels of women subordination presented along with various spheres in which women subordination is acted out. The matrix can be used for raising awareness, planning, policy analysis and others.

Table 1. Women Empowerment Matrix

Particulars	**Physical**	**Sociocultural**	**Religious**	**Political**	**Legal**	**Economic**
Individual						
Household						
Community						
State						
Region						
Global						

(Source: Wieringa, 1994)

The matrix provides a structure and holistic nature of specific issues to be dealt in case of women empowerment. Despite covering major issues still it lacks focus on class and ethinicity alongwith it is useful for qualitative usage and fails to provide historical depth.

The process of empowerment has five dimensions and they are Cognitive, Psychological, Economic, Political and Physical (Mokta, 2014). The different dimensions are explained below:

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

Table 2. Process of Empowerment and its dimensions

Dimensions	**Explanation**	**Parameters**
Cognitive	Women having an understanding of the conditions and causes of their subordination at the micro and macro levels. It involves making choices that may go against cultural expectations and norms	Raising self-esteem and self-confidence of women. • Elimination of discrimination and all forms of violence against women and girl child. •Building and strengthening partnership with civil society particularly women's organisations. • Enforcement of constitutional and legal provisions and safeguarding rights of women. • Building a positive image of women in the society and recognising their contributions in social, economic and political sphere. • Developing ability among women to think critically. • Fostering decision-making and collective action. • Enabling women to make informed choices. • Ensuring women's participation in all walks of life. • Providing information, knowledge, skills for self-employment
Psychological	Belief that women can act at personal and societal levels to improve their individual realities and the society in which they live	
Economic	Women have access to, and control over, productive resources, thus ensuring some degree of financial autonomy. However she notes that changes in the economic balance of power do not necessarily alter traditional gender roles or norms	
Political	Women have the capability to analyse, organise and mobilise for social change;	
Physical	Element of gaining control over one's body and sexuality and the ability to protect oneself against sexual violence to the empowerment process	

(Source: Mokta, 2014)

3. Proposed Model of Study and Methodology

The models discussed in the second section have dealt with factors which influences women empowerment. *Wieringa, 1993 and* Mokta 2014 have also discussed various dimension of Women empowerment. The model proposed herein discusses women empowerment on three facets and they are economic empowerment, social empowerment and political empowerment. These facets of women empowerment depends on legislative protections, affirmative policies for training, control over resources and various protective and promotional programmes of the government, NGOs and the corporate sector.

The study also aims to identify the gaps between present state of women empowerment in India and what is expected from stakeholders to achieve the Sustainable Development Goal – 5 on gender equality by 2030. Relevant secondary sources like scholarly articles, report by various ministries functional under Government of India, reports by corporate and non government organisations and international organisations have been used for review, analysis and discussion.

4. Analysis

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

From the Fifth Five Year Plan (1974-78) onwards there has been a marked shift in the approach to women's issues from welfare to development. In recent years, the empowerment of women has been recognized as the central issue in determining the status of women. The National Commission for Women was set up by an Act of Parliament in 1990 to safeguard the rights and legal entitlements of women. Several initiatives have been taken by the Government of India to achieve women empowerment. In course of discussing women empowerment, Tomar and Singh (2018) have given detailed list of constitutional safe guards. Relevant of them are given below:

4.1. Constitutional Articles, Government Plans and Programmes for Women Empowerment

Indian planners and policy makers have taken up the issue of women empowerment quite seriously. This section deals with the government's effort towards women empowerment.

4.1.1. Constitutional Articles

Women empowerment has been given required attention by the constitution committee headed by Baba Saheb Bhim Rao Ambedkar while framing the Indian Constitution. Some of the articles specifically dedicated to women have been discussed below:

4.1.1.1. Economic Empowerment

Provisions in the Indian Constitution

i. Equality of opportunity for all citizen in matters relating to employment or appointment to any office under the state (Article 16)

ii. The state to direct its policy towards securing for men and women equally the right for an adequate means of livelihood (Article 39(a)); and equal pay for equal work for both men and women [Article 39(d)]

iii. The state to make provision for securing just and humane conditions of work and for maternity relief (Article 42)

4.1.1.2. Political Empowerment

i. The 73rd Constitutional Amendment Acts passed in 1992 by the parliament of India ensures that not less than one-third (including the number of seats reserved for women belonging to Scheduled Caste and Scheduled Tribes) of the total number of seats to be filled by direct

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

election in every Panchayat to be reserved for women and such seats to be allotted by rotation to different constituencies in a Panchayat [Article 243D(3)]

ii. Not less than one-third of the total number of offices of the Chairpersons in the Panchayats at each level to be reserved for women (Article 243 D(4)]

iii. Not less than one -third (including the number of seats reserved for women belonging to the scheduled Caste and the Scheduled Tribes) of the total number of seats to be filled by direct election in every Municipality to be reserved for women and such seats to be allotted by rotation to different constituencies in a Municipality (Article 243T(3))

4.1.1.3. Social Empowerment

i. Equality before Law for women (Article 14)

ii. No discrimination against any citizen on grounds of sex beside religion, race, caste etc. (Article- 15)

iii. The state to make special provision in favour of women and children (Article 15(3))

iv. To promote justice, on the basis of equal opportunity and to provide free legal aid by suitable legislation or scheme or in any other way to ensure that opportunities for securing justice to all irrespective of economic inequality or other disability (Article 39A)

v. The state to promote with special care the educational and economic interests of the weaker sections of the people and to protect them from social injustice and all forms of exploitation (Article 46)

vi. To renounce practices derogatory to the dignity of women [Article 51(A)(e)]

4.1.2. Government Plans and Programmes

Since the inception of gender equality in the Constitution of India, it has been one of the priority areas and with special focus as SDG-5 government has been actively working to achieve gender equality through women empowerment and few initiatives have been discussed as below:

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

4.1.2.1. Economic Empowerment

i. The Ministry of Women and Child development has been administering 'Support to Training and Employment Programme for Women (STEP) Scheme' since 1986-87 as a 'Central Sector Scheme'. The scheme aims to provide employability skills and competency to women in the age group of 16 years and above to enable women to become self-employed/ entrepreneurs. The grant under the Scheme is given to an institution/ organisation including NGOs directly. The assistance will be available in any sector including but not limited to the Agriculture, Horticulture, Food Processing, Handlooms, Tailoring, Stitching, Embroidery, Zari etc, Handicrafts, Computer & IT enable services along with soft skills and skills for the work place such as spoken English, Gems & Jewellery, Travel & Tourism, Hospitality (STEP, 2009)

ii. The Government of India being concerned about the difficulties faced by working women, introduced a scheme in 1972-73 of grant-in- aid for construction of new/ expansion of existing buildings for providing hostel facilities to working women in cities, smaller towns and also in rural areas where employment opportunities for women exist. Based on an evaluation, the existing scheme has been revised to promote availability of safe and conveniently located accommodation for working women who need to live away from their families due to professional commitments (Scheme for Working Women Hostel, 1972-73). Rajiv Gandhi National Creche Scheme for the Children of Working Mothers was recast by government of India in 2006 and the aim of the programme was to provide day-care facilities for children (6 months to 6 years) of working mothers in the community, to improve nutrition and health status of children, to promote physical, cognitive, social and emotional development of children and further to educate and empower parents /caregivers for better childcare.

iii. The Ministry of Women and Child Development launched Mahila E-Haat a bilingual portal on 7th March, 2016. This is a unique direct online marketing platform leveraging technology for supporting women entrepreneurs/Self Help Groups/ Non Government Organisations for showcasing the products and services which are made, manufactured or undertaken by them. It is an initiative for meeting aspirations and needs of women. This was done keeping in mind that technology is a critical component for business efficiency and to make it available to the majority of Indian women entrepreneurs, Self Help Groups and Non Government Organisations. Mahila E-haat received the SKOCH GOLD Award on 09th Sep, 2016 and was also awarded 'SKOCH Order-of-Merit' Award, as it was adjudged as one of the "Top 100 Projects in India"

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

for the year 2016. To increase visibility major PSUs, IRCTC, Nationalised Banks like SBI have provided a link to Mahila E -haat on its website (https://wcd.nic.in/sites/default/files/Mahila%20E-haat%2024012017.pdf).

iv. The Finance Minister in his budget speech (2017-18) announced setting up of Mahila Shakti Kendra at village level in 14 lakh ICDS Anganwadi Centres with an allocation of 500 crore. The purpose of this scheme is to provide "one stop convergent support services for empowering rural women with opportunities for skill development, employment, digital literacy, health and nutrition". Ministry of Women and Child Development has proposed a sub-scheme namely Mahila Shakti Kendra under the Umbrella Scheme of Mission for Protection and Empowerment for Women, this scheme is envisaged to provide an interface for rural women to approach the government for availing their entitlements and for empowering them through training and capacity building. The scheme provides convergent support is being proposed for equal access to healthcare, quality education, career and vocational guidance, employment, health and safety, social security and digital literacy at Gram Panchayats level in selected districts/ blocks across the country (https://wcd.nic.in/sites/default/files/Mahila%20Shakti%20Kendra%20Scheme.pdf).National Mission for Empowerment of women Scheme is effectively merged with MSK wherein the National level and State level structures will continue and facilitate the implementation of MSK at the district, block and Gram panchayat level duly covering the Anganwadi Centres.

4.1.2.2. Social Empowerment

i. Beti Bachao Beti Padhao (BBBP) Scheme was launched in January 2015 in Panipat, Haryana by our Hon'ble Prime Minister to address the issue of decline in Child Sex ratio and related issues of empowerment of girls and women over a life cycle continuum. The scheme was initially launched in 100 districts in 2014-15, was expanded to 61 additional districts in 2015-16 and then it was expanded to all 640 districts of the country through a nation-wide mass media campaign, focused intervention and multi-sectoral action in select districts. Goal of the scheme is to celebrate the girl child and enable her education. (Beti Bachao Beti Padhao Scheme Guidelines, 2019) The campaign aims at ensuring that girls are born, nurtured and educated without discrimination to become empowered citizens of this country with equal rights. The government introduced many innovative activities under this scheme like Digital Guddi-Gudda Board, Udaan-Sapneya Di Duniya De Rubaru, My Aim My Target Campaign, LAkshya se

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

Rubaru, Pahal-Ek Kadam Nari Samman ki Aur, Ghar ki Pehchan Beti ke Naam, Bal Cabinet, Introduce Pink Cards and many more.Panic Button on Mobile Phones was an initiative by The Ministry of Women and Child Development since June 2014 had taken up the issue of installation of physical panic button on mobile phones.

ii. One Stop Centres (OSCs) are intended to support women affected by violence, in private and public spaces, within the family, community and at the workplace. Women facing physical, sexual, emotional, psychological and economic abuse, irrespective of age, class, caste, education status, marital status, race and culture will be facilitated with support and redressal. Aggrieved women facing any kind of violence due to attempted sexual harassment, sexual assault, domestic violence, trafficking, honour related crimes, acid attacks or witch-hunting who have reached out or been referred to the OSC will be provided with specialized services. Under this Scheme, in the first phase, one OSC envisaged to be established in each State/Union Territory to facilitate access to an integrated range of services including medical, legal, and psychological support. Further, 150 additional Centres are taken up in second phase during2016-17 in addition to 36 Centres in the first phase. The OSCs will be integrated with 181 and other existing helplines (One Stop Centre Scheme 2015). Women affected by violence and in need of redressal services could be referred to OSC through these help lines.

iii. The Scheme of Universalisation of Women Helpline is intended to provide 24 hours immediate and emergency response to women affected by violence through referral (linking with appropriate authority such as police, One Stop Centre, hospital) and information about women related government schemes programs across the country through a single uniform number. Women Helpline (WHL) will be integrated with One Stop Centre Scheme (OSC) under which one OSC shall be established in every State/UT to provide integrated support and assistance to women affected by violence, both in private and public spaces under one roof. Women affected by violence and in need of redressal services will be referred to OSC through WHL. Under the Scheme, the States/Union Territories will utilise or augment their existing women helplines through a dedicated single national number. Department of Telecommunication, Government of India has allocated short code 181 to all States/Union Territories (Universalisation of Women Helpline, 2015)

iv. UJJAWALA scheme has been effective from 2016 Ministry has formulated a Central Scheme "Comprehensive Scheme for Prevention of Trafficking for Rescue, Rehabilitation and Re-

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

Integration of Victims of Trafficking for Commercial Sexual Exploitation-Ujjawala". The new scheme has been conceived primarily for the purpose of preventing trafficking on the one hand and rescue and rehabilitation of victims on the other. The beneficiaries of the scheme are women and children who are vulnerable to trafficking for commercial sexual exploitation (Ujjawala, 2016).The implementing agencies can be the Social Welfare/Women and Child Welfare Department of State Government, Women's Development Corporations, Women's Development Centres, Urban Local Bodies, reputed Public/Private Trust or Voluntary Organizations.

v. Swadhar –A Scheme for Women in Difficult Circumstances was launched by the Department of Women and Child Development in 2001- 02. The scheme envisions a supportive institutional framework for women victims of difficult circumstances so that they could lead their life with dignity and conviction (Swadhar Greh, 2015). It envisages that shelter, food, clothing, and health as well as economic and social security are assured for such women. It also envisions that the special needs of these women are properly taken care of and under no circumstances they should be left unattended or abandoned which could lead to their exploitation and desolation.

vi. The Ministry of Women and Child Development, celebrates International Women's Day every year on 8th March as it is the nodal Ministry for the issues related to women. Taking this as an opportune moment to acknowledge Women's achievements, it has been decided by the Ministry in the year 2015, to confer Nari Shakti Puruskars to eminent women and institutions in recognition of their service towards the cause of women empowerment. In the last decade, there has been concerted effort by the Government to recognise and encourage women as reflected through a National Policy for Empowerment of Women in 2001. The issues related to women has gained utmost importance and focussed attention. "Nari Shakti Puruskars" shall showcase the Government's commitment towards women with the aim of strengthening their legitimate place in the society (http://narishaktipuraskar.wcd.gov.in/aboutus). It will also provide an opportunity to the current generation to understand the contribution of women in building of society and the nation.

vii. Following the tragedy (heinous rape of para-medical student on a moving bus in Delhi) of December 2012, the Government has set up a dedicated fund – Nirbhaya Fund – which can be utilized for projects specifically designed to improve the safety and security of women. It is a

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

non-lapsable corpus fund, being administered by Department of Economic Affairs, Ministry of Finance.

As per the guidelines issued by Ministry of Finance dated 25.03.2015, the Ministry of Women and Child Development (MWCD) is the nodal Ministry to appraise/recommend proposals and schemes to be funded under Nirbhaya Fund. MWCD further has the responsibility to review and monitor the progress of sanctioned schemes in conjunction with the line Ministries/Departments. An Empowered Committee was set up on 26.10.2015 to appraise and recommend proposals to be funded under Nirbhaya Fund. Government also initiated a portal to facilitate victims/ complainants to report cyber crime complaints online. Complaints on cyber crimes against women and children that are reported on this portal are dealt by law enforcement agencies/ police based on the information available in the complaints.

Women are susceptible to violence both in public and private spaces, including domestic violence, sexual assault, rape, stalking and others. A gender responsive police service requires specific training, increased presence of female personnel within the police force and community outreach to integrate gender issues into policies, protocols and operational procedures Government of India, 2016)In recent years, there has been enactment of various legislations in the Parliament which address issues related to Gender Based Violence like the Criminal Law Amendment Act, 2013, the Sexual Harassment of Women at Workplace (Prevention, Prohibition and Redressal) Act, 2013,the Protection of Women From Domestic Violence Act, 2005 and provided a platform to women facing violence to take the recourse to law. According to the latest National Crime Records Bureau data, during the year 2018, 3,78,277 incidences of crime against women (both under Indian Penal Code and other laws) were reported as against the 3,38,954 cases reported during 2016.

Table 3. Crime against women (IPC + SLL)

Particulars	**2016**	**2017**	**2018**
Total States	322949	345989	363817
Total Union Territories	16005	13860	14460
Total All India	338954	359849	378277

(Source: National Crime Records Bureau report of 2018, 2017 and 2016)

As observed by the Working Group on Women's Agency and Empowerment, these numbers have to be viewed keeping in mind that not all crimes against women are reported. The actual numbers may give even greater cause for concern. Role of Police is pivotal in safety and

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

security of citizens in general and women in particular. To increase the visibility of women in the police force, Home Ministry has carried forward the initiative to give 33% reservation to women in police force by implementing it in UTs and propagating in the States (12th Five Year Plan Report of the Working Group on Women's Agency and Empowerment, 2011). There has been an increasing emphasis on gender sensitivity of police force through training programmes, performance appraisal, women police stations to tackle crime against women.

Past research concluded that women empowerment can bring about demographic change lead to positive reproductive health outcomes (Mason, 1995). Empowering women economically is associated with higher levels of physical mobility, economic security, autonomous decision-making, exposure to violence, legal and political awareness and participation (Schuler and Hashemi, 1994). Women empowerment has also been linked with self-confidence, financial confidence, attitude towards gender norms, autonomous decision-making, household communication, partner relationships, participation in social groups and collective action (Mabsout, 2011; Hindin, 2000; Yount et al., 2019; Basu and Koolwal, 2005; Upadhyay et al. 2014).

4.2. NGOs Initiative for Women Empowerment

Women empowerments especially in rural areas have been primary focus for Non-Government Organisations (NGOs) beside the Government. The Ministry of Women and Child Development collaborates with many NGOs for the implementation of its various schemes and for conducting research studies. The Government of India implements many of its schemes with the active participation and support of the NGOs as they play an important role in accelerating the pace of implementation at the ground level. Table 4 lists out presence of performing NGOs in states (with 77.55% of NGOs working for women empowerment) and it states the highest presence in Delhi and the least in Goa, Mizoram and Tripura and one of the reasons could be based on the statistics of crime rate in these states.

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

Table 4. State-wise List of NGOs

State	**Performing NGOs**	**% age of NGOs working for Women Empowerment***
Andhra Pradesh	35	68.6
Arunachal Pradesh	4	100
Assam	38	81.6
Bihar	18	100
Chandigarh	7	57.1
Chattisgarh	7	100
Delhi	146	54.8
Goa	1	-
Gujarat	16	93.8
Haryana	17	58.8
Himachal Pradesh	8	100
Jammu and Kashmir	2	50
Jharkhand	11	90.9
Karnataka	63	74.7
Kerala	12	75
Madhya Pradesh	12	75
Maharashtra	82	90.2
Manipur	37	89.1
Meghalaya	2	-
Mizoram	1	100
Nagaland	2	100
Orissa	72	87.5
Punjab	5	60
Rajasthan	25	92
Tamil Nadu	53	87.7
Telangana	20	75
Uttar Pradesh	61	86.9
Uttarakhand	15	73.3
Tripura	1	-
West Bengal	37	73
Total	**811**	**77.5**

* the area of work included women development or/and empowerment

(Source: List of Performing and Non-Performing NGOs issued by Ministry of Women and Child Development, Govt. Of India)

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

Some works conducted by NGOs in this area are as follows:

i. Apna Ghar Ashram support destitute and disabled women get a safe and caring home. More than 50% of their residents have been rehabilitated. 458 men and 1,004 women currently live in the Ashrams spread across 17 major cities across India. The Ashram was also awarded the State Award by the State Government as the Best Services Organization.

ii. Started in 1973, Alpnalaya works for women empowerment in India by helping them to overcome the social, political, and economic barriers. They provide sewing classes and train women who otherwise have not been given the choice or opportunity to become financially independent. 209 women have economically benefited from Self Help groups organized by Apnalaya. In 2017, this NGO for women was awarded the Champion Level- GuideStar India Platinum for the highest levels of transparency and public accountability.

iii. Since its inception in 2005, Maitri has worked with over 45,000 individuals on the issues of social and health problems and public health concerns through education, community outreach, networking and legal advocacy. The organization has been addressing the issue of violence against women, running a victim support program and providing support to abandoned elderly widows in Vrindavan.

iv. Action Aid India since 1972 shelters the poor and the excluded. Gauravi is the first crisis center set up in India with the help of the Department of Health in 2014 to fight for Indian women's rights and saving them from domestic violence. Currently, there are 22 Gauravi centers across the country and over 40,000 cases have been registered in the centers.

v. Ashadeep Mission was initiated to develop a better quality of life (medical treatment and therapy) for people with mental disabilities. The Navachetana home run by Ashadeep rescues women with mental health issues and provides them with shelter, food, and medical care. They also help these women reunite with their families if possible after treatment. So far, 355 women have been rescued and rehabilitated under Ashadeep mission.

vi. Seven Sisters Development Assistance works to create a just and equitable society in Northeast India devoid of poverty. The organization has two programs, leadership programs for the educated youth of North East India and empowering poor rural women through livelihood activities via Self Help Groups (SHGs). Since it was established in 2011, SeSTA has mobilized nearly 8,000 women to form 700 Self Help Groups (SHGs).

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

vii. Foundation for Mother and Child Health works to bring good health and nutritional provisions to vulnerable mothers and children in poor communities. Their programs encourage preventive health, balanced nutrition, and child development practices in underprivileged communities. So far, the programs initiated by FMCH impact close to 927 pregnant and lactating mothers annually. Their services include regular growth monitoring, availability of accurate information and support from the detection of pregnancy till the child completes two years.

viii. Majlis Manch's legal center through its program 'Rahat', provides socio-legal support to victims of domestic violence and sexual abuse. The all-women team of lawyers and social workers meet with victims and explain the process and procedures involved in a court case. They are also told about all the legal, public, and private schemes that they have access to.

4.3. Corporate Social Responsibility (CSR) for Women Empowerment

As per the new Companies Bill passed by the lower house of Indian parliament in December 2012, Public Sector Units (PSUs) and private companies having a turnover of over 10 billion or a net worth of 5 billion or recorded net profit of 50 million and above are required to spend 2 per cent of their average net profits in the preceding three financial years towards CSR. A few of the CSR initiatives by companies for women empowerment is listed in table 5.

Table 5. Companies CSR Initiatives for Women Empowerment

Companies	**Focus Area**	**Prescribed CSR of Co. In FY 2017-18 (amount spent) in INR**	**Locations**	**Implementing Partner**
ITC	Capacity Building of Women (gainful employment opportunities)	290.47 Cr (14.76 Cr)	Andhra Pradesh, Telangana, Karnataka, Tamil Nadu, Bihar, Madhya Pradesh, Uttar Pradesh, Assam, Rajasthan, West Bengal, Delhi, Maharashtra, Odisha, Punjab, Uttarakhand, Himachal Pradesh	BAIF Development Research Foundation, Pratham Education Foundation, Ramakrishna Mission, Bandhan Konnagar, SEWA Bharat, Foundation for Ecological Security
IndusInd Bank Ltd.	Livelihood generation of rescued trafficked women (rehabilitation and skill development of women)	70.17 Cr (11.63 Cr)	Bihar (Gaya, Nalanda and Nawada)	Justice and Care
Godrej Consumer	Project Salon-i (trains young girls	18.83 Cr (8.86 Cr)	Gujarat, Maharashtra, Uttar Pradesh, Bihar, Chattisgarh,	Dhriiti, Ambuja Cement Foundation, DDJF, Don

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

Products Ltd	and women in beauty and hair care)		Assam, Nagaland, Manipur, Mizoram, Delhi	Bosco Tech Society, Father Agnel Ashram, Labournet, Saath, Pratham, NSHM, Save the Children India, Tara Livelihood Academy, Unnati
Inter Globe Aviation Limited	Project 6E Shakti (skill enhancement by providing agriculture and non-agriculture based training)	42.03 Cr (4.67 Cr)	Maharashtra, Assam, Meghalaya, Uttar Pradesh, Chattisgarh, Rajasthan	AFARM Grameen Sahara, Bandhan Konnagar, Divinity Serves
Hindustan Zinc Ltd	Sakhi (provide platforms for women to engage in full and effective participation in the public life of their villages)	189.65 Cr (3.57 Cr)	Rajasthan, Uttarakhand	Saheli Samiti & Manjari Foundation
Bayer Crop Science Limited	Discovering Hands-Training Center for Visually Impaired Women (training centre for visually impaired women to become medical tactile examiners in early breast Cancer Detection)	10.09 Cr (1.88 Cr)	Delhi	Indian Cancer Society, National Association for the Blind, Rajiv Gandhi Cancer & Research Centre, Disha Foundation
Sterlite Technologies Ltd	Vocational Training for Women (conducts vocational training on nursing, computer operations, data entry, tailoring, fashion designing and basic beauty culture)	3.63 Cr (1.72 Cr)	Maharashtra	Jeewan Jyoti Women Empowerment Institution
CEAT Ltd	Swayam (training underprivileged women in driving skills to enhance livelihoods across the transport sector)	10.71 Cr (1.45 Cr)	Maharashtra, Rajasthan, Delhi, Madhya Pradesh, Tamil Nadu	RPG Foundation
Sasken Technologies Ltd	m-bachat (platform for financial independence of women)	4.51 Cr (1.03 Cr)	Gujarat	SEWA
Apollo Tyres Ltd	Livelihood Enhancement Project for Women (supporting women farmers by training them on modern techniques and ways to increase agricultural productivity)	21.57 Cr (0.77 Cr)	Gujarat	Apollo Tyres Foundation

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

(Source: Compiled from the websites of the companies)

Big corporate houses consider women empowerment as a vital aspect of CSR and it has been observed that they usually carry these interventions in the surrounding areas of the main operational regions either by the in-house team or outsourced professional CSR team and it is sometimes carried out in collaboration with NGOs.

The 55th Round of the National Sample Survey (July 1999-June 2000), which was the first-ever nation-wide survey on informal sector non-agricultural enterprises, showed that the total number of informal workers in non-agricultural enterprises was 79.7 million. Out of these, 30 million were home based (Sudarshan et al, 2001) and a vast majority of these workers are women. Organisations like Dastkar has provided a platform to artisans (especially in rural areas) engaged in hand embroidery and other crafts to receive fair price (eliminating the middle men) for their products. It provides fair wage and fair rate for the products of the local artisans.

4.4. SDG Goal 5- Achieve Gender Equality: Status of India

Under SDG Goal 5 to achieve gender equality by 2030 there are several targets to be achieved aand they are as follows:

Target 1.End all forms of discrimination against all women and girls everywhere

Target 2. Eliminate all forms of violence against all women and girls in the public and private spheres, including trafficking and sexual and other types of exploitation

Target 3. Eliminate all harmful practices, such as child, early and forced marriage and female genital mutilation

Target 4.Recognize and value unpaid care and domestic work through the provision of public services, infrastructure and social protection policies and the promotion of shared responsibility within the household and the family as nationally appropriate

Target 5. Ensure women's full and effective participation and equal opportunities for leadership at all levels of decision-making in political, economic and public life

Target 6. Ensure universal access to sexual and reproductive health and reproductive rights as agreed in accordance with the Programme of Action of the International Conference on

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

Population and Development and the Beijing Platform for Action and the outcome documents of their review conferences

Target 7. Undertake reforms to give women equal rights to economic resources, as well as access to ownership and control over land and other forms of property, financial services, inheritance and natural resources, in accordance with national laws

Target 8. Enhance the use of enabling technology, in particular information and communications technology, to promote the empowerment of women

Target 9. Adopt and strengthen sound policies and enforceable legislation for the promotion of gender equality and the empowerment of all women and girls at all levels

The National Institution for Transforming India (NITI Aayog), has been assigned the responsibility of overseeing the implementation towards achievement of SDGs. A multi-disciplinary VNR Task Force was constituted by NITI Aayog to coordinate the review and process documentation.

India has achieved gender parity at primary education level and is on track to achieving parity at all education levels, the proportion of seats in the Lok Sabha held by women had only reached 11% and in Panchayati Raj Institutions by 46%. Though the sex ratio has improved marginally. Lowest was 927 women for 1000 male in 1991 which has marginally increased to 933 in 2001 and 940 in 2011) but, India is confronting the challenge of violence against women (domestic violence, sexual violence in public spaces, physical and mental exploitation at work place).

Thus, the Indian government has identified ending violence against women as a key national priority for the same it has targeted to end all forms of discrimination, all forms of violence and exploitation against girls/ women, eliminate all harmful practices like child, early and forced marriage, recognise and value unpaid care and domestic work through the provision of public services, social protection policies and promotion of shared responsibility within the household, enhance the usage of information and communications technology for empowerment of women.

4. Global Gender Gap Index: Status of India

4.4.1. World Economic Forum introduced the Global Gender Gap Index in 2006 as a framework for capturing the magnitude and scope of gender based disparities and tracking their progress

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

over time across four thematic dimensions and they are Economic Participation and Opportunity, Educational Attainment, Health and Survival and Political Empowerment.

Table 6. India's position in global Gender Gap

Year	Global Index	Economic Participation and Opportunity	Educational Attainment	Health and Survival	Political Empowerment
2020	112 (0.668)	149 (0.354)	112 (0.962)	150 (0.944)	18 (0.411)
2018	108 (0.665)	142 (.385)	114 (0.953)	147(0.940)	19 (0.382)
2016	87 (0.683)	136 (0.408)	113 (0.950)	142 (0.942)	9 (0.433)
2014	114 (0.6455)	134 (0.4096)	126 (0.8503)	141 (0.9366)	15 (0.3855)
2012	105 (0.6442)	123 (0.4588)	121 (0.8525)	134 (0.9312)	17 (0.3343)
2010	112 (0.6155)	128 (0.4025)	120 (0.8369)	132 (0.9312)	23 (0.2913)
2006	98 (0.601)	110 (0.397)	102 (0.819)	103 (0.962)	20 (0.227)

(Source: Global Gender Gap Report 2006, 2010, 2012, 2014, 2016, 2018 and 2020 by World Economic Forum)

The year 2016 witnessed an increase in the overall ranking which positioned India at 87 out of 144 countries and the reasons may be the Gross Domestic Product for the year which stood as US $ 2073.54 billions, population growth rate at 1.15%, population sex ratio (female to male) at 0.93 and human capital optimization at 57.73%. Another impertinent reason for the increase in the parity is associated with enrolment in primary education (1.01), secondary education (1.01) and tertiary education (0.94). Parity has also been maintained in healthy life expectancy (1.05) and sex ratio at birth (0.89).

India (112) maintains a stable ranking this year but its gap is directionally larger this year with a 33% gap yet to be bridged. The country records a decline in wage equality for similar work, succeeds in fully closing its tertiary education gender gap for the first time, and keeps primary and secondary education gaps closed for the third year running. However, it continues to rank third lowest in the world on Health and Survival, remaining the world's least-improved country on this sub index over the past decade. In fact, India actually widens the gender gap on this sub index this year. However, the condition of women in large fringes of India's society is precarious. Among the 153 countries studied, India is the only country where the economic gender gap is larger than the political gender gap.

4.4.2. South Asian Association for Regional Cooperation (SAARC) Nations is the regional intergovernmental organization and geopolitical union of states in South Asia. SAARC

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

comprises 3% of the world's area, 21% of the world's population and 4.21% (US$3.67 trillion) of the global economy, as of 2019. Table 7 points out that Maldives has gained first rank in achieving parity in educational attainment as compared to Pakistan that ranks 143 where the gap is 0.823 reason being the societal pressure in Pakistan which does not allow girls to attain education. India is improving in lowering the gender gap in providing education and penetrating in rural India by providing suitable infrastructure. Sri Lanka ranks first in maintaining gender parity in health and survival as compared to India which is also working towards achieving parity.

Table 7. Ranking of SAARC Nations in Global Gender Gap

Member States	**Global Index**	**Economic Participation and Opportunity**	**Educational Attainment**	**Health and Survival**	**Political Empowerment**
Afghanistan	NA	NA	NA	NA	NA
Bangladesh	50 (0.726)	141 (0.438)	120 (0.951)	119 (0.969)	7 (0.545)
Bhutan	131 (0.635)	130 (0.544)	116 (0.954)	144 (0.960)	132 (0.082)
India	112 (0.668)	149 (0.354)	112 (0.962)	150 (0.944)	18 (0.411)
Maldives	123 (0.646)	131 (0.518)	1 (1)	147 (0.953)	115 (0.111)
Nepal	101 (0.680)	101 (0.632)	133 (0.895)	131 (0.966)	59 (0.227)
Pakistan	151 (0.564)	150 (0.327)	143 (0.823)	149 (0.944)	93 (0.159)
Sri Lanka	102 (0.680)	126 (0.558)	88(0.988)	1 (0.980)	73 (0.193)

(Source: Global Gender Gap Report 2020 by World Economic Forum)

4.4.3. BRICS is the acronym for the economic bloc of countries consisting of Brazil, Russia, India, China and South Africa. Economists believe these four nations will become dominant suppliers of manufactured goods, services, and raw material by 2050 due to low labour and production costs. India is struggling to achieve gender parity among its counterparts which hold better position in all aspects. India ranks the lowest in comparison with other BRICS nations and needs to take effort to outshine its counterparts.

Table 8. Position of BRICS nation in Global Gender Gap

Member States	**Global Index**	**Economic Participation and Opportunity**	**Educational Attainment**	**Health and Survival**	**Political Empowerment**
Brazil	92 (0.691)	89 (0.653)	35 (1)	1 (0.980)	104 (0.133)
Russia (Russian Federation)	81 (0.706)	32 (0.749)	1(1)	1 (0.980)	122 (0.095)
India	112 (0.668)	149 (0.354)	112 (0.962)	150 (0.944)	18 (0.411)
China	106 (0.676)	91 (0.651)	100 (0.973)	153 (0.926)	95 (0.154)

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

South Africa	17 (0.780)	92 (0.649)	67 (0.994)	1 (0.980)	10 (0.497)

(Source: Global Gender Gap Report 2020 by World Economic Forum)

5. Conclusion and Recommendation

Women Empowerment is not merely a slogan but a prerequisite for an optimal capacity for overall development of a family, society, nation and a sustainable world. The paper has covered about affirmative legislation, plans and programmes initiated by the Government, NGOs and Corporate sector through their Corporate Social Responsibility over a very large spectrum starting from the Indian constitution to development in the recent years. It has been reflected by improvement in achieving gender parity in the category of educational attainment, showing improvement in political participation but showing a declining trend in economic participation and opportunity as well as in health and survival segment. In terms of Health and Survival, it is worse even among SAARC countries. As we have a decade towards achieving the SDG Goals, the Government needs to focus on these two segments more to achieve gender parity. There is urgent need of reassessing India's affirmative actions, giving fresh impetus on empowering girls and women not only through conventional education but also hand holding for skill acquisition and providing them with equitable economic opportunity with providing them with fitting medical and healthcare. The findings open up several avenues for policy discussion and intervention towards women empowerment.

References

Alsop, R. and Heinsohn, N. (2005) Measuring empowerment in practice: structuring analysis and framing indicators. World Bank

Basu, M. and Koolwal, G. B. (2005) Two concepts of empowerment: Some leads from DHS data on women's status and reproductive health. In S. Kishor (Ed.), A focus on gender: Collected papers on gender using DHS data, pg. 15-53

Batliwala, S. (1994) The meaning of women's empowerment: new concepts from action. In Population Policies Reconsidered: Health, Empowerment, and Rights, Sen G, Germain A, Chen LC (eds). Harvard Center for Population and Development Studies: Boston.

Batliwala, S. (2007) Taking the power out of empowerment: an experiential account. Development inPractice, 17(4/5), pg. 557–65.

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

Beti Bachao Beti Padhao Scheme (2019) Implementation Guidelines for State Government and UT Administrations, Governement of India pg. 1-103

Commission of Women and Development (2007) The Women Empowerment Approach: A Methodological Guide.

Government of India (2016) Sanction for Implementation of Mahila Police Volunteers during financial year 2016-17, pg. 1-17

Hindin, M (2000) Women's autonomy, women's status and fertility-related behavior in Zimbabwe. Population Research Policy Review, Vol.19, pg. 255–282

Lucia Hanmer & Jeni Klugman (2016) Exploring Women's Agency and Empowerment in Developing Countries: Where do we stand?, Feminist Economics, 22(1), pg. 237-263

Kabeer N. (1994) Reversed Realities: Gender Hierarchies in Development Thought. Verso: London.

Kabeer N. (1999) Resources, agency, achievements: reflections on the measurement of women'sempowerment. Development and Change 30(3), pg. 435–64.

Kabeer, N. (2008) Paid work, women's empowerment and gender justice, Pathways Working Paper3, Brighton: Pathways of Women's Empowerment.

Kabeer, N. (2011) Contextualising the economic pathways of women's empowerment, PathwaysPolicy Paper, Brighton: Pathways of Women's Empowerment.

Lee-Rife, SM. (2010) Women's empowerment and reproductive experiences over the lifecourse. Soc Sci Med

Mabsout, R (2011) Capability and Health Functioning in Ethiopian Households, Social Indicators Research, Vol 101, Issue 3, pg. 359-389

Malhotra A, Schuler SR, and Boender C. (2002) Measuring women's empowerment as variable in international development. World Bank Gender and DevelopmentGroup Background paper.

Mason, KO. (1995) International Union for the Scientific Study of Population. Gender and demographic change: What do we know?

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

Mathew, G. (2003) Keynote address in the workshop on "A Decade ofWomen's Empowerment through Local Governance" organized jointly byInstitute of Social Sciences and South Asia Partnership, Canada sponsored byInternational Development Research Centre

Ministry of Women and Child Development (2011) 12th Five Year Plan Report of the Working Group on Women's Agency and Empowerment, Government of India, pg. 1-118

One Stop Centre Scheme (2015) Government of India Guidelines pg. 1-12

Rowlands. (1997) Questioning Empowerment: Working with Women in Honduras. OxfamPublishing: Oxford.

Scheme for Working Women Hostel, 1972-73 Guidelines issued by Government of India, pg 1-28

Schuler, S.R. and Hashemi, S.M. (1994) Credit programs, women's empowerment andcontraceptive use in rural Bangladesh. Stud Fam Plann, 25(2), pg. 65-76.

Sen, G. (1997) Empowerment as an approach to poverty, Working Paper Series 97.07, backgroundpaper for the UNDP Human Development Report, New York: UNDP.

STEP (2009) Guidelines issued by Governement of India, pg 1-30

Sudarshan, R. M., Banerjee, M., Bhandari, L., Venkatraman, S. (2001), Outsourcing of Manufacturing to Households: Subcontracted Home Based Work in India, UNICEF Innocenti Research Centre (unpublished paper).

Swadhar Greh (2015) Scheme Guidelines issued by Government of India, pg 1-32

Tomar N. and Singh S. (2018), "How sensitive is the Indian Government Towards Women Empowerment in India" International Journal of Management, IT and Engineering, 08 (07), pp 66-83.

Ujjawala (2016) Scheme Guidelines, Government of India, pg 1-31it is not detail

Universalisation of Women Helpline (2015) Implementation Guidelines for State and Union Territories Administration, pg 1- 68

Copyright © 2020Author

ISSN:2320-0693
Vol-19-Issue-44-June-2020

Upadhyay, U., Gipson, J., Withers, M., Lewis, S., Ciaraldi, E., Fraser, A., Huchko. M. and Prata, N (2014) Women's empowerment and fertility: A review of Literature, Social Science and Medicine, Vol. 115, pg. 111-120

Wieringa, S. (1994) Women's Interests and Empowerment: Gender Planning Reconsidered, Development and Change, 25(4), pg 829-848

Yount, K., Cheong, Y., Maxwell, L., Heckert, J., Martinez, E. And Seymour, G (2019) Measurement properties of the project level Women Empowerment in Agriculture Index, World Development, Vol. 124, pg. 1-19

Copyright © 2020Author

View publication stats